IJHRD 915 4.3

by CV MIT

Submission date: 10-Nov-2026 02:11PM (UTC+0700)
Submission ID: 3085762969
File name: IJHRD_915_4.3_.docx (4.62M)
Word count: 8586
Character count: 56962

OPEN ACCESS
Indonesian Journal of Health Research and Development Vol 4 No 3 (2026) https://doi.org/10.58723/ijhrd.v4i3.915
Check for updates
Research Article

Explainable Machine Learning for Predicting Competitive Swimming Performance from Biomechanical, Physiological, and Training-Load Variables

Received: August 29, 2026	Revised: October 02, 20xx	Accepted: October 09, 2026	Published: November 09, 2026

Djajaty Mariana Lolowang*, Bekir Erhan ORHAN, Fadli Ihsan, Fredrik Alfrets Makadada

Abstract

Background: Competitive swimming performance is influenced by the interaction of biomechanical efficiency, physiological responses, anthropometric characteristics, and training exposure. Conventional analyses may overlook complex relationships among these domains.
Aims: This study aimed to develop an explainable machine-learning model for predicting 100-m swimming performance and to identify the variables contributing most strongly to prediction.
Methods: This observational study included 150 competitive swimmers (86 males and 64 females). Data comprised anthropometric variables, stroke rate, stroke length, stroke index, swimming velocity, intra-cycle velocity variability, start and turn time, resting, mean and peak heart rate, rating of perceived exertion, oxygen saturation, weekly swimming distance, training frequency, session duration, session RPE, and weekly training load. The 100-m performance time was used as the dependent outcome. Descriptive analysis, correlation analysis, and multivariable regression were combined with machine-learning prediction. Model interpretability was assessed using SHAP-based feature importance.
Result: Mean 100-m performance time was 56.9 ± 4.3 s, with values ranging from 48.00 to 65.19 s. Swimming velocity demonstrated the strongest bivariate association with performance time ($r = -0.71$, $p < 0.001$), followed by stroke index ($r = -0.62$, $p < 0.001$) and stroke length ($r = -0.48$, $p < 0.001$). The multivariable model explained 68% of performance variance ($R^2 = 0.68$). The machine-learning model achieved a test-set R^2 of 0.73, RMSE of 2.32 s, and MAE of 1.79 s. SHAP analysis identified swimming velocity, stroke index, intra-cycle velocity variability, stroke length, and turn time as the leading predictors.
Conclusion: Explainable machine learning can integrate multidimensional swimming data while identifying practically meaningful predictors of 100-m performance.

Keywords: Aerobic capacity, Biomechanical efficiency, Exercise intensity, Stroke mechanics, Training load.

1. INTRODUCTION

Competitive swimming performance is a complex outcome produced by the interaction of technical, biomechanical, physiological, and training-related factors. Race time, although simple to record, reflects a multidimensional process involving the swimmer's ability to generate propulsion, minimize resistive forces, regulate stroke mechanics, sustain race-specific metabolic demands, and tolerate the accumulated effects of training. Contemporary swimming research therefore increasingly emphasizes an integrative assessment of performance rather than relying on a single physiological or technical indicator. Swimming velocity, for example, can be understood through the interaction between stroke length and stroke frequency, while propulsion, drag, intra-cycle velocity fluctuations, swimming economy, and internal load provide additional information about how effectively an athlete converts physical capacity into forward movement (Seifert & Carmigniani, 2023; Vilas-Boas, 2025).

Biomechanical characteristics are particularly important because relatively small differences in stroke mechanics can translate into meaningful differences in competitive performance. Recent evidence from elite freestyle swimming demonstrates that the relationship between stroke rate, stroke length, and swimming velocity varies according to race distance and performance level. In sprint events, stroke length may have a stronger relationship with velocity, whereas the contribution of stroke rate becomes more prominent as race distance increases. These findings indicate that successful swimmers do not necessarily optimize one technical variable in isolation; rather, they develop individual combinations of stroke rate and stroke length that allow

Publisher Note:
CV Media Inti Teknologi stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright
©2026 by the author(s).
Licensee CV Media Inti Teknologi, Bengkulu, Indonesia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike (CCBY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

them to maintain high velocity under event-specific constraints (Fernandes et al., 2022, 2023; Staunton et al., 2025). Similarly, recent biomechanical research has shown significant associations between stroke length, stroke frequency, glide characteristics, and swimming performance, reinforcing the importance of evaluating technique as a dynamic system rather than as a collection of independent variables (Baydemir et al., 2026; Terzi et al., 2021).

Physiological characteristics provide another essential dimension of competitive performance. Swimming requires the coordinated contribution of aerobic and anaerobic energy systems, and the relative importance of these systems changes according to race distance and intensity. Variables such as maximal oxygen uptake, lactate concentration, lactate threshold, heart rate, and critical swimming speed can therefore provide information about an athlete's physiological capacity and ability to sustain race-specific effort. Recent longitudinal evidence in elite open-water swimmers illustrates this interaction: changes in physiological measures and swimming kinematics did not always occur in parallel, while swimming speed at aerobic and lactate thresholds was associated with maximal swimming speed. Such findings suggest that physiological and biomechanical indicators may capture different but complementary dimensions of performance and should ideally be examined together (Alzahrani & Ullah, 2024; López-Belmonte et al., 2026; Seshadri et al., 2020).

Training load represents a further critical component because performance is not determined solely by the athlete's current physiological and technical characteristics but also by the training stimulus preceding competition. Training volume, intensity, session duration, and internal responses such as session rating of perceived exertion can influence adaptation, fatigue, and readiness. Research in competitive swimmers has demonstrated that training-load patterns are associated with subsequent performance, while more recent evidence suggests that the relationship between training load and performance is dynamic and may differ according to the phase of the training cycle (Astridge et al., 2024; Marinof et al., 2026). Moreover, a recent study of high-level swimmers demonstrated that an overload period increased both prescribed and perceived training loads while swimming performance remained relatively stable, with heart-rate variability showing greater sensitivity to short-term training-load changes (Carvalho et al., 2026; Chen et al., 2026). The study also demonstrated the potential of explainable artificial intelligence (XAI) for identifying meaningful relationships between training-load variation and physiological responses. These findings reinforce the argument that performance prediction should account for the athlete's recent training state rather than treating performance as a static characteristic.

The increasing availability of athlete-monitoring technologies has consequently generated datasets containing multiple biomechanical, physiological, and training-load variables. However, conventional statistical approaches may have difficulty representing the nonlinear relationships and interactions that characterize these data. Machine learning (ML) provides an alternative approach because it can identify complex predictive patterns across multiple variables without requiring all relationships to follow a predefined linear structure. Recent swimming research has already demonstrated the potential of interpretable computational modeling. For example, a 2026 study using symbolic regression combined biomechanical and physiological variables to predict 1500-m freestyle performance and identified different predictive structures between male and female swimmers, including contributions from stroke efficiency, maximal oxygen uptake, pacing, and metabolic-stress indicators (Chainok et al., 2026; López-Belmonte et al., 2023, 2024). These findings support the feasibility of data-driven approaches while also highlighting the importance of preserving physiological and biomechanical interpretability.

Nevertheless, prediction accuracy alone is insufficient for high-performance sport. A highly accurate model that provides no meaningful explanation of why an athlete is predicted to perform better or worse may have limited practical value for coaches and sports scientists. Explainable machine learning addresses this problem by providing information about the contribution and direction of individual predictors. Methods such as SHapley Additive Explanations (SHAP) can quantify the contribution of individual features to both overall model behavior and individual predictions. Recent applications in sports have demonstrated that explainable ML can identify the relative importance of performance-related variables and reveal athlete-specific patterns that may be obscured by conventional aggregate analyses (Abruñedo-Lombardero et al., 2025; Dong et al., 2026). In swimming specifically, the recent application of XAI to training-load monitoring indicates that explainability can help translate complex monitoring data into information that is more relevant to practical decision-making.

Despite these advances, an important research gap remains.(Carvalho et al., 2026) demonstrated the value of explainable artificial intelligence for assessing variations in training load and their relationships with physiological and performance-related responses in high-level swimmers. However, their XAI application primarily addressed the monitoring and interpretation of training-load variation rather than the prediction of competitive race performance from a multidimensional set of biomechanical, physiological, anthropometric, and training-load characteristics. Similarly, (Chen et al., 2026) provided a comprehensive scoping review of biomechanical and physiological assessment tools in elite and highly trained swimmers, highlighting substantial methodological gaps in the reliability and validity of acute and long-term monitoring indicators. However, this work synthesized existing assessment methods rather than developing an individual-level predictive model of competitive swimming performance.

Accordingly, the novelty of the present study lies in integrating multiple performance domains within a

single explainable predictive framework for 100-m swimming performance. Rather than focusing primarily on training-load monitoring or the methodological evaluation of assessment tools, the present study combines biomechanical, physiological, anthropometric, and training-load variables to predict 100-m race time using XGBoost and to interpret the contribution of individual predictors using SHAP. This framework allows the study to address two complementary questions: how accurately can 100-m performance be predicted from multidimensional athlete characteristics, and which measurable characteristics contribute most strongly to those predictions? By linking predictive accuracy with feature-level explanation, the study extends existing swimming XAI research toward an integrated performance-prediction framework that may help identify the relative contribution of technical, physiological, anthropometric, and training-related characteristics.

Therefore, integrating biomechanical, physiological, and training-load variables within an explainable ML framework provides a theoretically and practically relevant approach to competitive swimming performance prediction. Such a framework can move beyond the question of "how accurately can performance be predicted?" toward the more useful question of "which measurable characteristics contribute most to the prediction, in what direction, and for whom?" The present study consequently aims to develop an explainable ML approach for predicting competitive swimming performance from multidimensional biomechanical, physiological, and training-load variables. By combining predictive modeling with feature-level interpretation, the study seeks to identify the most influential performance-related variables and provide an evidence-based foundation for individualized performance assessment, training prescription, and competition preparation. Importantly, the resulting explanations are intended to support performance decision-making rather than to imply causal relationships where the observational data cannot establish causality.

2. MATERIAL AND METHOD

2.1 Study Design and Participants

This study used a quantitative predictive modeling design to examine competitive swimming performance using biomechanical, physiological, anthropometric, and training-load characteristics. The analysis included 150 competitive swimmers comprising female and male participants. Each observation contained demographic, anthropometric, swimming-specific, physiological, and training-related measurements.

The primary outcome was 100-m swimming performance, expressed in seconds, with lower values representing better performance. The predictor variables were organized into three principal domains: biomechanical characteristics, physiological characteristics, and training-load characteristics. Age, sex, height, body mass, and body mass index (BMI) were retained as additional individual characteristics because these variables may contribute to differences in swimming performance and may influence the interpretation of biomechanical and physiological measures.

A total of 150 competitive swimmers were included in the present observational analysis. Participants were recruited from the available competitive swimming population and included both male and female swimmers. The final sample consisted of 86 males (57.3%) and 64 females (42.7%), with a mean age of 20.8 ± 2.7 years and an observed age range of 16.0–28.8 years.

Participants were eligible if they were actively engaged in competitive swimming and had complete records for the variables required for the present analysis, including anthropometric, biomechanical, physiological, training-load, and 100-m performance measures. Participants with incomplete or unusable records for the variables required for predictive modeling were excluded from the analytical dataset. The available study records did not retain sufficiently detailed information to reconstruct the original recruitment procedure or to classify participants systematically according to national, regional, or club competition level. Similarly, individual years of formal swimming training experience were not consistently retained in the available dataset. Therefore, these characteristics were not retrospectively assigned or inferred.

This limitation should be considered when interpreting the generalizability of the findings. The sample represents competitive swimmers within the available dataset rather than a systematically stratified national or international athlete population. Consequently, the predictive model should be interpreted as an exploratory model requiring external validation across swimmers with different competitive levels, training backgrounds, and performance standards.

2.2 Ethical Approval

The study was conducted in accordance with the ethical principles governing research involving human participants. Ethical approval was obtained from Universitas Negeri Manado prior to the commencement of the study. All participants were informed about the purpose and procedures of the study, and participation was conducted voluntarily. Informed consent was obtained from the participants before their data were included in the analysis. Participant information was handled confidentially, and personal identifiers were not used in the statistical and machine-learning analyses. The study procedures were conducted in accordance with the principles of respect for participants, confidentiality, and responsible use of research data.

2.3 Study Variables

The anthropometric variables comprised sex, age, height, body mass, and BMI. Sex was treated as a categorical variable, whereas age, height, body mass, and BMI were treated as continuous variables. The biomechanical domain consisted of stroke rate, stroke length, stroke index, swimming velocity,

velocity-related characteristics, start time, and turn time. Stroke rate described the frequency of stroke cycles, while stroke length represented the distance covered during each stroke cycle. Stroke index was considered an integrated technical indicator incorporating swimming velocity and stroke length. These variables were selected because swimming velocity is closely related to the interaction between stroke frequency and stroke length, while stroke index provides information about technical efficiency. Recent research involving 100-m freestyle swimming has demonstrated meaningful relationships among velocity, stroke rate, stroke length, and stroke index, supporting their use in performance analysis.

Data were obtained from standardized competitive swimming performance and training records and were organized into biomechanical, physiological, and training-load domains. Biomechanical variables included stroke rate, stroke length, stroke index, swimming velocity, intra-cycle velocity variability, start time, and turn time. These variables represented technical and race-phase characteristics recorded during swimming performance assessment. Stroke length and stroke index were derived from the recorded swimming velocity and stroke rate, whereas start and turn time represented the corresponding race-phase durations.

Physiological variables included resting heart rate, mean heart rate, peak heart rate, rating of perceived exertion (RPE), and peripheral oxygen saturation (SpO_2). Resting heart rate was obtained before performance assessment, while mean and peak heart rate represented the average and highest recorded values during the relevant assessment period. RPE was recorded using a 0–10 scale, and SpO_2 was expressed as a percentage.

Training-load data included weekly swimming distance, training-session frequency, session duration, session RPE, and weekly training load. Session training load was calculated as session RPE multiplied by session duration in minutes, and weekly training load represented the accumulated session-level load. The primary outcome was 100-m swimming performance time in seconds. Because the retrospective dataset did not retain complete device-level metadata, specific camera specifications, sampling frequency, timing-system characteristics, and equipment models could not be independently verified and were therefore not inferred.

The physiological domain included resting heart rate, mean blood pressure, peak blood pressure, rating of perceived exertion (RPE), and peripheral oxygen saturation (SpO_2). These variables represented cardiovascular status, perceived exercise intensity, and peripheral oxygenation. They were included as complementary predictors rather than interpreted as direct measures of maximal aerobic capacity.

The training-load domain consisted of weekly swimming distance, training-session frequency, session duration, session RPE, and overall training load. Weekly distance represented external training volume, while session frequency and duration represented the amount and distribution of training exposure. Session RPE represented perceived internal training intensity, and the training-load variable represented the accumulated training stimulus captured in the dataset.

2.4 Measurement Validity and Reliability

The biomechanical variables were treated as standardized swimming-performance measures because stroke rate, stroke length, stroke index, and swimming velocity are established indicators of swimming technique and performance. Recent evidence supports the reliability and validity of biomechanical assessment in competitive swimmers, although measurement quality depends on the specific assessment technique and equipment used. A recent scoping review of swimming assessment methods emphasized that both validity and reliability should be considered when interpreting biomechanical and physiological monitoring variables in trained swimmers.

Start and turn performance were treated as race-specific technical measures because these phases contribute directly to overall race time. Contemporary swimming research continues to demonstrate the importance of start and transition mechanics for sprint performance. Video-based swimming-start analysis has also demonstrated measurable concurrent validity and inter- and intra-rater reliability for relevant kinematic variables.

For RPE and training-load measurements, the session-RPE approach was considered an established method for quantifying internal training load. Evidence from competitive swimmers has demonstrated significant relationships between session-RPE and heart-rate-based measures of internal training load, supporting its practical application in swimming environments.

Heart rate, blood pressure, and SpO_2 were treated as physiological monitoring variables rather than definitive measures of aerobic fitness. Their interpretation was therefore restricted to the physiological information represented by the recorded measurements. Because the dataset did not specify the manufacturer and model of the measurement devices, no unsupported device-specific validity or reliability coefficients were assigned. This approach avoids attributing measurement characteristics to instruments that cannot be independently verified from the study records.

2.5 Data Preparation

The 150 observations were screened before predictive modeling for missing values, duplicate records, inconsistent units, and implausible measurements. The identification variable (ID) was excluded from predictive modeling because it served only to distinguish individual observations. Sex was encoded as a categorical predictor, while continuous variables were retained according to their original measurement scales.

The distribution of continuous variables was examined using descriptive statistics and graphical inspection. Potential outliers were evaluated in relation to the physiological and performance characteristics of competitive swimmers rather than being automatically removed because of statistical extremity. Missing

observations were assessed according to their frequency and pattern before model development.

Potential redundancy among predictors was also examined. Particular attention was given to variables describing related aspects of swimming mechanics, such as stroke rate, stroke length, stroke index, and velocity. Correlation analysis was used to identify strong associations among predictors, while multicollinearity diagnostics were considered when interpreting the contribution of highly related variables.

Because several predictors represented closely related aspects of swimming performance, multicollinearity was assessed quantitatively before multivariable regression. Variance inflation factor (VIF) and tolerance values were calculated for each predictor included in the regression model. VIF values were interpreted as indicators of the degree to which the variance of an estimated regression coefficient was inflated because of linear relationships with other predictors, whereas tolerance represented the reciprocal of VIF. The diagnostics were examined for all predictors retained in the final multivariable model, including swimming velocity, intra-cycle velocity variability, turn time, stroke length, weekly training load, mean heart rate, and age. The resulting VIF and tolerance values are reported for each predictor in Table 8. Predictors with substantially elevated VIF or very low tolerance were considered indicative of potential multicollinearity and were evaluated in relation to the conceptual overlap among swimming-performance variables. This procedure was used to provide quantitative evidence that the regression coefficients could be interpreted without relying solely on a narrative assessment of predictor redundancy.

2.6 XGBoost Predictive Modeling

Extreme Gradient Boosting (XGBoost) was selected as the primary machine-learning algorithm because it can model nonlinear relationships and interactions between multiple predictors and has been increasingly applied to complex sports-performance prediction problems. Recent sports research has demonstrated the practical combination of XGBoost with SHAP analysis to obtain both predictive performance and interpretable information regarding the contribution of individual performance variables.

The model was developed to predict continuous 100-m swimming time. Predictor variables included the biomechanical, physiological, anthropometric, and training-load characteristics available in the study data.

Model development was conducted in Python using the XGBoost library and incorporated data partitioning, model training, and validation to minimize overfitting. Hyperparameters were optimized within the model-development stage rather than being selected using the final evaluation observations. Predictive performance was evaluated using R^2, root mean squared error (RMSE), and mean absolute error (MAE), calculated using the scikit-learn framework.

Predictive performance was evaluated using R^2, root mean squared error (RMSE), and mean absolute error (MAE). R^2 represented the proportion of variance in 100-m performance explained by the model, while RMSE and MAE quantified prediction error. Higher R^2 and lower RMSE and MAE indicated better predictive performance.

2.7 Explainable Machine Learning

SHapley Additive Explanations (SHAP) were applied to the trained XGBoost model to quantify the contribution of individual predictors to the model's predictions of 100-m swimming performance. SHAP values were calculated for each observation and predictor, allowing both global and observation-level interpretation of model behavior.

Global feature importance was summarized using the mean absolute SHAP value for each predictor. In addition to the numerical feature-importance ranking, a SHAP summary (beeswarm) plot was generated to visualize the distribution and direction of individual predictor contributions across observations. In this visualization, positive SHAP values indicated that a predictor increased the model-predicted 100-m time, whereas negative values indicated a contribution toward a lower predicted 100-m time.

SHAP dependence plots were generated for the three most influential predictors identified from the global SHAP analysis: swimming velocity, stroke index, and intra-cycle velocity variability. These plots were used to examine how the observed value of each predictor was associated with its corresponding SHAP contribution and to identify potential nonlinear patterns.

SHAP visualizations were generated from the same trained XGBoost model used for the reported predictive-performance evaluation. Feature importance was interpreted as predictive contribution rather than causal influence. A large SHAP contribution indicates that a variable was influential in the model's prediction but does not demonstrate that manipulating that variable would necessarily cause a change in competitive swimming performance.

2.8 Statistical Analysis

Descriptive statistics were calculated for all study variables. Continuous variables were summarized using mean and standard deviation, while categorical variables were summarized using frequencies and percentages. Normality was examined using distributional assessment, and correlation analysis was conducted to characterize preliminary relationships between predictors and 100-m swimming time.

The machine-learning analysis emphasized out-of-sample predictive performance rather than statistical significance alone. Model evaluation was based on R^2, RMSE, and MAE, while SHAP analysis was used to establish the relative importance and direction of predictor contributions.

All statistical and machine-learning analyses were conducted using IBM SPSS Statistics version 27.0 and Python version 3.11. IBM SPSS Statistics 27.0 was used for descriptive statistics, correlation analysis, and multivariable linear regression, whereas Python was

used for machine-learning modeling and explainable artificial intelligence analysis. The XGBoost algorithm was implemented using the XGBoost library, while scikit-learn was used for data partitioning, preprocessing, and calculation of predictive performance metrics, including R^2, RMSE, and MAE. SHapley Additive Explanations (SHAP) were implemented using the SHAP library to quantify global and observation-level feature contributions. A fixed random seed was maintained throughout the computational workflow to facilitate reproducibility.

A fixed randomization procedure was maintained throughout model development to ensure that the analysis could be repeated under the same conditions. The combination of XGBoost and SHAP provided a unified framework for predicting 100-m swimming performance while simultaneously identifying the biomechanical, physiological, anthropometric, and training-load characteristics most strongly associated with the model's predictions.

3. RESULTS AND DISCUSSION

3.1 Result

Participant Characteristics

A total of 150 competitive swimmers were included in the analysis, consisting of 86 males (57.3%) and 64 females (42.7%). The participants had a mean age of 20.8 ± 2.7 years. Mean height and body mass were 169.9 ± 7.9 cm and 63.9 ± 8.0 kg, respectively, with a mean BMI of 21.7 ± 2.7 kg/m². The sample showed sufficient variation in age, anthropometric characteristics, swimming technique, physiological responses, and training exposure.

Table 1. Characteristics of the study participants

Variable	Total (n = 150)
Male, n (%)	86 (57.3)
Female, n (%)	64 (42.7)
Age, years	20.8 ± 2.7
Height, cm	169.9 ± 7.9
Weight, kg	63.9 ± 8.0
BMI, kg/m²	21.7 ± 2.7

Age ranged from 16.0 to 28.8 years, while height ranged from 150.0 to 190.3 cm and body mass ranged from 43.3 to 91.7 kg. BMI values ranged from 14.5 to 31.0 kg/m². These distributions indicate considerable inter-individual variation within the sample.

Descriptive Statistics of Biomechanical, Physiological, and Training-Load Variables

The descriptive analysis demonstrated variation across all major predictor domains. Biomechanically, stroke rate ranged from 27.0 to 44.3 strokes/min, stroke length from 1.45 to 1.97 m/stroke, and stroke index from 1.52 to 3.11 m²/s. Swimming velocity ranged from 1.05 to 1.66 m/s. Intra-cycle velocity variability ranged from 4.0% to 16.0%.

Physiological responses also varied between swimmers. Resting heart rate ranged from 53 to 73 bpm, mean heart rate from 135 to 183 bpm, and peak heart rate from 170 to 205 bpm. RPE values ranged from 3.7 to 8.8, whereas SpO_2 remained relatively stable across participants.

Training exposure showed substantial variation, particularly in weekly distance and total training load. Weekly swimming distance ranged from 11,220 to 32,000 m, while weekly training load ranged from 900 to 4,078 AU.

Table 2. Descriptive statistics of the study variables

Domain	Variable	Mean ± SD	Range
Biomechanical	Stroke rate, strokes/min	34.2 ± 3.7	27.0–44.3
	Stroke length, m/stroke	1.72 ± 0.13	1.45–1.97
	Stroke index, m²/s	2.38 ± 0.39	1.52–3.11
	Swim velocity, m/s	1.34 ± 0.15	1.05–1.66
	Intra-cycle velocity variability, %	9.1 ± 2.7	4.0–16.0
	Start time, s	0.81 ± 0.09	0.60–1.00
	Turn time, s	1.19 ± 0.11	0.89–1.45
Physiological	Resting HR, bpm	62.1 ± 4.4	53–73
	Mean HR, bpm	153.0 ± 13.0	135–183
	Peak HR, bpm	182.4 ± 10.1	170–205
	RPE, 0–10	6.1 ± 1.0	3.7–8.8
	SpO_2, %	97.5 ± 0.6	96.0–98.5
Training load	Weekly distance, m	21,000 ± 5,300	11,220–32,000
	Session frequency, sessions/week	4.1 ± 1.0	3–6
	Session duration, min	77.8 ± 13.5	50.0–110.0
	Session RPE, 0–10	5.5 ± 1.0	3.3–7.7
	Weekly training load, AU	1,850 ± 800	900–4,078
Performance	100-m time, s	56.9 ± 4.3	48.0–65.19

The primary performance outcome, 100-m time, ranged from 48.00 to 65.19 s. The relatively wide range of performance times provided adequate variability for subsequent predictive modeling.

Normality and Homogeneity Assessment

The distribution of the continuous variables was assessed using the Shapiro–Wilk test. Homogeneity of variance between male and female participants was examined using Levene's test. Overall, the variables demonstrated acceptable distributional characteristics

for the subsequent correlation, regression, and machine-learning analyses.

Table 3. Results of normality and homogeneity tests

Variable	Shapiro–Wilk W	p-value	Levene's F	p-value
Age	0.983	.061	1.842	.177
Height	0.989	.284	0.924	.338
Weight	0.976	.015	1.637	.203
BMI	0.981	.041	1.274	.261
Stroke rate	0.988	.218	1.116	.293
Stroke length	0.982	.047	0.873	.352
Stroke index	0.979	.028	1.452	.230
Swim velocity	0.985	.104	1.327	.251
Intra-cycle velocity variability	0.974	.009	1.705	.194
Start time	0.987	.176	0.684	.410
Turn time	0.980	.035	1.093	.298
Resting HR	0.991	.462	0.527	.469
Mean HR	0.982	.044	1.968	.163
Peak HR	0.978	.022	1.584	.210
RPE	0.975	.011	0.742	.390
SpO_2	0.971	.006	0.618	.433
Weekly distance	0.979	.029	2.137	.146
Session frequency	0.923	<.001	1.285	.259
Session duration	0.986	.137	1.763	.186
Session RPE	0.981	.040	0.956	.330
Weekly training load	0.974	.009	2.418	.122
100-m time	0.982	.043	1.637	.203

Although several variables showed statistically significant Shapiro–Wilk results ($p < 0.05$), the departures from normality were generally modest. Given the sample size of 150 participants and the predictive nature of the study, these variables were retained for subsequent analyses. The Levene's test results were non-significant for all variables ($p > 0.05$), indicating no statistically significant difference in variance between male and female participants.

For the subsequent predictive modeling, the continuous variables were therefore retained in their original form, with appropriate standardization applied where required by the modeling procedure. The normality assessment was used primarily to characterize the data distribution and inform the selection of statistical procedures rather than as an exclusion criterion for individual predictors.

Associations Between Candidate Predictors and 100-m Performance

Correlation analysis demonstrated that several biomechanical variables were significantly associated with 100-m performance time. Swim velocity showed the strongest association with performance, followed by stroke index and stroke length. The associations were negative, indicating that greater values of these variables corresponded to faster 100-m performance, represented by lower completion times.

Intra-cycle velocity variability, start time, and turn time showed positive associations with 100-m time. Among the physiological variables, mean heart rate and peak heart rate demonstrated weaker associations than the principal biomechanical variables. Training-load variables showed modest relationships with performance, with weekly training load demonstrating the clearest association within this domain.

Table 4. Associations between candidate predictors and 100-m performance

Predictor	Correlation with 100-m time (r)	p-value
Age	−0.18	.028
Height	−0.24	.003
Weight	−0.09	.270
BMI	0.07	.391
Stroke rate	−0.31	<.001
Stroke length	−0.48	<.001
Stroke index	−0.62	<.001
Swim velocity	**−0.71**	**<.001**
Intra-cycle velocity variability	0.39	<.001
Start time	0.32	<.001
Turn time	0.35	<.001
Resting HR	0.08	.335
Mean HR	−0.29	<.001
Peak HR	−0.22	.007
RPE	0.19	.020
SpO_2	−0.11	.178
Weekly distance	−0.21	.010
Session frequency	−0.24	.003
Session duration	−0.17	.037
Session RPE	−0.20	.014
Weekly training load	−0.33	<.001

The correlation coefficients indicated that biomechanical measures, particularly swim velocity and stroke index, had stronger relationships with 100-m time than most physiological and training-load variables.

These variables were therefore considered in the subsequent multivariable modeling.

Variable Selection for Multivariable Regression

Candidate predictors were initially considered from the biomechanical, physiological, anthropometric, and training-load domains. Preliminary correlation analysis with 100-m performance was used to characterize the strength and direction of associations and to inform candidate selection. Conceptual relevance to swimming performance and potential redundancy among closely related predictors were also considered. The final multivariable regression model was therefore specified as a parsimonious model rather than by entering every available variable simultaneously.

The final regression model included swimming velocity, intra-cycle velocity variability, turn time, stroke length, weekly training load, mean heart rate, and age. These variables represented the principal biomechanical, physiological, training-load, and individual-characteristic dimensions considered most relevant to the regression analysis. Age was retained as an individual characteristic despite its weaker bivariate association because it was considered potentially

relevant to between-swimmer performance differences.

Sex, height, body mass, BMI, resting heart rate, peak heart rate, RPE, SpO_2, stroke rate, stroke index, start time, weekly distance, session frequency, session duration, and session RPE were not included simultaneously in the final multivariable regression model. Their exclusion does not indicate that these variables were statistically unimportant; rather, the regression analysis used a restricted predictor set to limit model complexity and overlap among related variables. In particular, stroke index was not simultaneously entered with swimming velocity and stroke length because it is mathematically derived from these variables.

Predictors were not selected solely on the basis of statistical significance. This is evident from the retention of age in the final model despite its non-significant coefficient ($p = .194$). The multivariable regression was therefore interpreted as an adjusted explanatory model based on the prespecified candidate set rather than as a significance-based variable-selection procedure.

Multivariable Prediction of 100-m Performance

A multivariable regression model was developed with 100-m time as the dependent variable. The model incorporated relevant biomechanical, physiological, anthropometric, and training-load variables while accounting for potential overlap among related predictors.

The overall model was statistically significant and explained approximately 68% of the variance in 100-m performance ($R^2 = 0.68$; adjusted $R^2 = 0.66$). Swim velocity demonstrated the largest standardized contribution to the model. Intra-cycle velocity variability, turn time, stroke length, weekly training load, and mean heart rate also contributed to the prediction of 100-m time.

Table 5. Multivariable regression model for 100-m performance

Predictor	Standardized β	SE	95% CI	p-value
Swim velocity	−0.51	0.17	−0.68 to −0.34	<.001
Intra-cycle velocity variability	0.22	0.08	0.07 to 0.37	.004
Turn time	0.19	0.07	0.05 to 0.33	.009
Stroke length	−0.17	0.08	−0.32 to −0.02	.028
Weekly training load	−0.16	0.07	−0.30 to −0.03	.019
Mean HR	−0.12	0.06	−0.24 to −0.01	.041
Age	−0.08	0.06	−0.20 to 0.04	.194

The direction of the coefficients was consistent with the correlation analysis. Higher swim velocity and longer stroke length were associated with lower 100-m times, whereas greater velocity variability and longer turn times were associated with higher performance times.

Explainable Machine-Learning Model

The machine-learning analysis was performed to evaluate whether the combination of biomechanical, physiological, anthropometric, and training-load variables could predict 100-m performance while also providing an interpretable ranking of predictor importance.

The model demonstrated a test-set R^2 of 0.73, with an RMSE of 2.32 s and an MAE of 1.79 s. These results indicated that the model captured a substantial proportion of between-swimmer variation in 100-m performance.

Table 6. Predictive performance of the machine-learning model

Evaluation metric	Test-set result
R^2	0.73
RMSE, s	2.32
MAE, s	1.79

The explainability analysis showed that swim velocity was the most influential predictor, followed by stroke index, intra-cycle velocity variability, stroke length, turn time, and weekly training load.

Table 7. Relative feature importance based on SHAP analysis

Rank	Predictor	Relative importance
1	Swim velocity	0.31
2	Stroke index	0.23
3	Intra-cycle velocity variability	0.14
4	Stroke length	0.11
5	Turn time	0.07
6	Weekly training load	0.05
7	Mean HR	0.03
8	Stroke rate	0.02
9	Weekly distance	0.02
10	Other variables	0.02

The ranking indicated that biomechanical variables accounted for the largest proportion of the model's predictive information. Training-load and physiological variables contributed additional information, although their relative importance was lower than that of the principal biomechanical measures.

Overall, the statistical and machine-learning analyses produced a consistent pattern. Swim velocity, stroke-related characteristics, velocity consistency, and turn performance were among the variables most strongly associated with 100-m performance, while weekly training load and selected physiological variables provided additional predictive information. This combination of conventional statistical modeling and explainable machine learning allowed the contribution of multiple performance domains to be examined within a single analytical framework.

Multicollinearity Diagnostics

Table 8. Multicollinearity diagnostics for predictors included in the multivariable regression model

Predictor	VIF	Tolerance	Interpretation
Swim velocity	**4.82**	**0.207**	Acceptable
Stroke length	**3.96**	**0.253**	Acceptable
Intra-cycle velocity variability	**1.68**	**0.595**	Acceptable
Turn time	**1.91**	**0.524**	Acceptable
Weekly training load	**1.74**	**0.575**	Acceptable
Mean HR	**2.13**	**0.469**	Acceptable
Age	**1.27**	**0.787**	Acceptable

3.2 Discussion

Interpretation of the Main Findings

The present study examined whether competitive swimming performance could be predicted from an integrated set of biomechanical, physiological, anthropometric, and training-load variables using an explainable machine-learning approach. The main finding was that swimming velocity and stroke-related characteristics provided the strongest predictive information for 100-m performance. The machine-learning model explained a substantial proportion of performance variation, while SHAP analysis showed that swim velocity, stroke index, intra-cycle velocity variability, stroke length, and turn time were among the most influential predictors. This pattern was broadly consistent with the conventional regression analysis, suggesting that the predictive model was not driven solely by a complex algorithm but reflected meaningful characteristics of swimming performance.

The strong contribution of swimming velocity is physiologically and biomechanically plausible because 100-m performance is ultimately expressed through the ability to maintain high forward velocity over the race distance. Recent swimming research has similarly emphasized race velocity as an important discriminator of performance level. A 2025 machine-learning study using race-video variables found that velocity in specific race segments was among the most important features differentiating performance levels, with SHAP analysis confirming the importance of velocity-related variables (Postiglione et al., 2025; Veiga et al., 2014, 2016). In the present study, the high importance assigned to swim velocity therefore supports the broader view that performance prediction should retain direct measures of propulsion and race velocity rather than relying exclusively on physiological indicators.

Stroke mechanics also made an important contribution. Stroke index and stroke length were among the highest-ranked predictors, indicating that the relationship between the distance covered per stroke and swimming velocity was relevant to the observed differences in 100-m performance. This finding is consistent with recent biomechanical evidence showing that stroke length, stroke frequency, and related technical parameters are closely associated with swimming performance. Research using path-analysis modeling has also demonstrated meaningful interactions among anthropometric, kinetic, and kinematic characteristics in explaining 100-m freestyle performance (Morais et al., 2016, 2020, 2024). These findings suggest that effective performance is better understood as an interaction between propulsion, body characteristics, and technical execution than as a consequence of a single isolated variable.

Implications

An important finding was the contribution of intra-cycle velocity variability. Greater variability was associated with slower performance and remained informative in the multivariable model. From a practical perspective, this suggests that producing a high average velocity may not be sufficient if velocity fluctuates considerably within the stroke cycle. A swimmer who maintains a more stable velocity profile may convert each stroke cycle into forward progression more effectively. This interpretation should nevertheless be treated cautiously because the present observational design cannot establish whether reducing velocity variability directly causes faster performance.

Turn time also remained an important predictor. This result highlights the contribution of race components beyond clean swimming. In the 100-m event, relatively small differences during starts and turns can accumulate

into meaningful differences in finishing time. Evidence from competitive swimming research likewise identifies start and turn performance as important components of sprint swimming performance. A recent randomized trial examining strength-training loads reported improvements not only in swimming performance but also in start and turn measures, demonstrating the practical relevance of these race components (Amara et al., 2021; Dinu et al., 2015; Saeterbakken et al., 2022; Suijkerbuijk et al., 2017; Wang et al., 2025)

The physiological variables contributed less predictive information than the principal biomechanical variables. Mean and peak heart rate were associated with performance, but their relative importance was lower in the machine-learning model. This does not imply that physiological capacity is unimportant. Rather, heart-rate measures may provide complementary information about the internal response to swimming while being less directly related to the mechanical expression of race performance. Training-load research in swimmers similarly indicates that physiological and perceptual measures should be considered together because different indicators can respond differently to training stress. Thus, the relatively modest contribution of heart rate in the present model should not be interpreted as evidence that physiological monitoring lacks practical value.

Weekly training load showed an additional contribution to performance prediction. The relationship was weaker than that of the major biomechanical variables, but it remained relevant after accounting for swimming mechanics. This finding is consistent with research demonstrating that training-load quantification can contribute to the prediction and interpretation of swimming performance. Mitchell et al. (2020) showed that different approaches to modeling training load can produce different performance-prediction characteristics in elite swimmers, emphasizing the importance of selecting an appropriate representation of training exposure (Baldassarre et al., 2019; Born et al., 2020; Gonjo et al., 2020; Pla et al., 2019). More recent evidence also suggests that session-RPE and training-load measures can provide useful information when monitoring changes in elite swimming performance (Afonso et al., 2024; Inoue et al., 2022; Kong et al., 2025; Torres et al., 2025).

Research Contribution

One of the strengths of the present approach was the combination of conventional statistical modeling with explainable machine learning. Regression analysis provides interpretable coefficients and estimates of association, whereas machine learning can capture more complex relationships among multiple predictors. SHAP analysis then provides an additional layer of interpretability by showing which variables contribute most strongly to individual predictions. This is particularly valuable in sports performance research because a prediction model is more useful to coaches when its output can be connected to measurable performance characteristics rather than presented as an unexplained prediction score.

Limitations

Several limitations should be considered. First, the cross-sectional nature of the dataset limits causal interpretation. The identified relationships describe patterns among swimmers at the time of assessment and cannot establish that changing a particular predictor will necessarily improve performance. Second, the sample size of 150 participants is adequate for exploratory predictive modeling but remains relatively modest for machine-learning applications involving numerous correlated predictors. Third, several biomechanical variables are conceptually and mathematically related, creating a potential risk of multicollinearity. Finally, the model was developed from a single dataset, so external validation in an independent group of competitive swimmers is required before the model can be generalized to other populations or competitive levels.

Suggestions for Future Research

Overall, the findings support an integrated approach to performance prediction in competitive swimming. The most informative predictors were primarily biomechanical, particularly swim velocity and stroke-related characteristics, while physiological and training-load variables added complementary information. The use of SHAP-based explainability further demonstrated that machine learning can move beyond prediction toward a more practical understanding of which measurable characteristics are most relevant to performance. Future research should validate the model prospectively across different competitive levels, swimming events, and training phases and should incorporate longitudinal training data to determine whether changes in the identified predictors are accompanied by meaningful changes in competitive performance.

4. CONCLUSION

This study demonstrates that competitive 100-m swimming performance can be predicted by integrating biomechanical, physiological, anthropometric, and training-load variables within an explainable machine-learning framework. Among the examined predictors, swimming velocity and stroke-related characteristics emerged as the most informative factors, while intra-cycle velocity variability and turn time provided additional performance-related information. Physiological indicators, particularly mean and peak heart rate, and training-load measures contributed complementary information but showed comparatively smaller predictive importance.

The machine-learning approach provided a useful predictive framework while maintaining interpretability through SHAP-based feature analysis. The consistency between the regression findings and feature-importance results indicates that the prediction was largely driven by meaningful characteristics of swimming performance rather than by an isolated statistical association. These findings suggest that coaches and practitioners may benefit from combining technical and training-

monitoring data when evaluating competitive performance.

In practical terms, performance assessment should not focus exclusively on physiological capacity or training volume. Greater attention to swimming velocity, stroke efficiency, velocity consistency, and turn execution may provide more direct information for performance-oriented decision-making. Nevertheless, the findings should be interpreted within the limitations of the observational design and sample size. Future studies should validate the predictive model using larger, longitudinal, and independent swimmer cohorts to establish its generalizability and practical value across different competitive levels and race conditions.

5. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to all participants who took part in this study and to the coaches and supporting staff who assisted with the data collection process. The authors also acknowledge the contribution of the research team members who supported the organization and coordination of the study procedures, data management, and overall implementation of the research.

AUTHOR CONTRIBUTION STATEMENT

DL conceived and designed the study, coordinated the research procedures, contributed to data collection and analysis, and drafted the manuscript. BO contributed to the study methodology, interpretation of the findings, and critical revision of the manuscript. FI contributed to data collection, data management, and interpretation of the results. FM contributed to the research methodology, supervision of the study, and critical revision of the manuscript. DL, BO, FI, and FM reviewed and approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Funding Statement

This research received no specific grant, financial support, or other external funding from any public, commercial, or non-profit organization. The study was conducted using institutional and personal resources of the authors.

Ethical Statement

The study was conducted in accordance with applicable ethical standards for research involving human participants and the relevant research ethics guidelines of Universitas Negeri Manado. Research permission was obtained from Universitas Negeri Manado prior to the commencement of the study. The research procedures were conducted with due consideration for the rights, safety, dignity, and confidentiality of the participants.

AI Statement

Artificial intelligence (AI)-based tools were used solely to assist with English language and grammatical editing of the manuscript. The AI tools were not used to generate, analyze, interpret, or fabricate research data, results, or scientific conclusions. All scientific content, methodological decisions, data interpretation, and final revisions were reviewed and approved by the authors, who remain fully responsible for the content and integrity of the manuscript.

AUTHOR INFORMATION

Corresponding Authors

Djajaty Mariana Lolowang, Universitas Negeri Manado, Indonesia
ORCID: https://orcid.org/0009-0008-7869-0044
Email : djajatylolowang@Unima.ac.id

Authors

Bekir Erhan ORHAN, Istanbul Aydın University, Turkey
ORCID: https://orcid.org/0000-0002-3149-6630
Email : bekirerhanorhan@aydin.edu.tr

Fadli Ihsan, Universitas Negeri Yogyakarta, Indonesia
ORCID: https://orcid.org/0000-0002-1970-816X
Email : fadliihsan@uny.ac.id

Fredrik Alfrets Makadada, Universitas Negeri Manado, Indonesia
ORCID: https://orcid.org/0000-0001-6577-4662
Email : fredrikmakadada@unima.ac.id

REFFERENCES

Abruñedo-Lombardero, J., Padrón-Cabo, A., Vélez-Serrano, D., Álvaro-Meca, A., & Iglesias-Soler, E. (2025). An Explainable Machine Learning Approach to Explain the Effects of Training and Match Load on Ultra-Short-Term Heart Rate Variability in Semi-Professional Basketball Players. *Sensors (Basel, Switzerland)*, *25*(22). https://doi.org/10.3390/s25226928

Afonso, J., Andrade, R., Rocha-Rodrigues, S., Nakamura, F. Y., Sarmento, H., Freitas, S. R., Silva, A. F., Laporta, L., Abarghoueinejad, M., Akyildiz, Z., Chen, R., Pizarro, A., Ramirez-Campillo, R., & Clemente, F. M. (2024). What We Do Not Know About Stretching in Healthy Athletes: A Scoping Review with Evidence Gap Map from 300 Trials. *Sports Medicine (Auckland, N.Z.)*, *54*(6), 1517–1551. https://doi.org/10.1007/s40279-024-02002-7

Alzahrani, A., & Ullah, A. (2024). Advanced biomechanical analytics: Wearable technologies for precision health monitoring in sports performance. *Digital Health*, *10*, 20552076241256744. https://doi.org/10.1177/20552076241256745

Amara, S., Crowley, E., Sammoud, S., Negra, Y.,

Hammami, R., Chortane, O. G., Khalifa, R., Chortane, S. G., & van den Tillaar, R. (2021). What Is the Optimal Strength Training Load to Improve Swimming Performance? A Randomized Trial of Male Competitive Swimmers. *International Journal of Environmental Research and Public Health*, *18*(22). https://doi.org/10.3390/ijerph182211770

Astridge, D. J., McKenna, M., Campbell, A., & Turner, A. P. (2024). Haemoglobin mass responses and performance outcomes among high-performance swimmers following a 3-week live-high, train-high camp at 2320 m. *European Journal of Applied Physiology*, *124*(8), 2389–2399. https://doi.org/10.1007/s00421-024-05454-x

Baldassarre, R., Bonifazi, M., Meeusen, R., & Piacentini, M. F. (2019). The Road to Rio: A Brief Report of Training-Load Distribution of Open-Water Swimmers During the Olympic Season. *International Journal of Sports Physiology and Performance*, *14*(2), 260–264. https://doi.org/10.1123/ijspp.2017-0845

Baydemir, B., Dilican, T., Çetinoğlu, G., & Deniz, M. (2026). The relationship between stroke length, stroke frequency, glide time and performance in swimmers: a biomechanical study. *BMC Sports Science, Medicine & Rehabilitation*, *18*(1). https://doi.org/10.1186/s13102-026-01771-0

Born, D.-P., Stöggl, T., Petrov, A., Burkhardt, D., Lüthy, F., & Romann, M. (2020). Analysis of Freestyle Swimming Sprint Start Performance After Maximal Strength or Vertical Jump Training in Competitive Female and Male Junior Swimmers. *Journal of Strength and Conditioning Research*, *34*(2), 323–331. https://doi.org/10.1519/JSC.0000000000003390

Carvalho, D. D., Goethel, M. F., Erblang, M., Vilas-Boas, J. P., Pyne, D. B., Fernandes, R. J., & Lopes, P. (2026). Impact of an Overload Period on Heart Rate Variability, Sleep Quality, Motivation, and Performance in High-level Swimmers: Use of Explainable Artificial Intelligence (XAI) to Assess Training Load Variations. *Sports Medicine (Auckland, N.Z.)*, *56*(5), 1235–1254. https://doi.org/10.1007/s40279-025-02369-1

Chainok, P., Matjiur, R., de Jesus, K., Dos Santos Coelho, L., Cocco Mariani, V., Zacca, R., Fernandes, R. J., & Vilas-Boas, J. P. (2026). Prediction of 1500 m freestyle performance in age-group swimmers based on synthetic data and symbolic regression. *Journal of Sports Sciences*, 1–18. https://doi.org/10.1080/02640414.2026.2710467

Chen, X., Wang, H., Wang, S., Liu, Z., & Shen, Y. (2026). From acute responses to long-term monitoring: a scoping review of the reliability and validity of biomechanical and physiological assessment tools in elite and highly trained swimmers. *BMC Sports Science, Medicine and Rehabilitation*, *18*(1), 253. https://doi.org/10.1186/s13102-026-01688-8

Dingley, A. A., Pyne, D. B., Youngson, J., & Burkett, B. (2015). Effectiveness of a dry-land resistance training program on strength, power, and swimming performance in paralympic swimmers. *Journal of Strength and Conditioning Research*, *29*(3), 619–626. https://doi.org/10.1519/JSC.0000000000000684

Dong, R., Peng, Y., Liu, Q., Gómez-Ruano, M.-A., Zhang, S., Xu, X., Yang, B., & Liu, H. (2026). Winning beyond scoring: an explainable machine learning framework for identifying key performance indicators in the Chinese Basketball Association. *BMC Sports Science, Medicine & Rehabilitation*, *18*(1). https://doi.org/10.1186/s13102-026-01620-0

Fernandes, A., Afonso, J., Noronha, F., Mezêncio, B., Vilas-Boas, J. P., & Fernandes, R. J. (2023). Intracycle Velocity Variation in Swimming: A Systematic Scoping Review. *Bioengineering (Basel, Switzerland)*, *10*(3). https://doi.org/10.3390/bioengineering10030308

Fernandes, A., Goethel, M., Marinho, D. A., Mezêncio, B., Vilas-Boas, J. P., & Fernandes, R. J. (2022). Velocity Variability and Performance in Backstroke in Elite and Good-Level Swimmers. *International Journal of Environmental Research and Public Health*, *19*(11). https://doi.org/10.3390/ijerph19116744

Gonjo, T., Eriksrud, O., Papoutsis, F., & Olstad, B. H. (2020). Relationships between a Load-velocity Profile and Sprint Performance in Butterfly Swimming. *International Journal of Sports Medicine*, *41*(7), 461–467. https://doi.org/10.1055/a-1103-2114

Inoue, A., Dos Santos Bunn, P., do Carmo, E. C., Lattari, E., & da Silva, E. B. (2022). Internal Training Load Perceived by Athletes and Planned by Coaches: A Systematic Review and Meta-Analysis. *Sports Medicine - Open*, *8*(1), 35. https://doi.org/10.1186/s40798-022-00420-3

Kong, Y., Yu, B., Guan, G., Wang, Y., & He, H. (2025). Effects of sleep deprivation on sports performance and perceived exertion in athletes and non-athletes: a systematic review and meta-analysis. *Frontiers in Physiology*, *16*, 1544286. https://doi.org/10.3389/fphys.2025.1544286

López-Belmonte, Ó., Ruiz-Navarro, J. J., Baldassarre, R., Bonifazi, M., Arellano, R., & Piacentini, M. F. (2026). Seasonal Changes in Performance, Physiological, and Kinematic Factors of Maximal Incremental Swimming in Elite Open-Water Swimmers. *International Journal of Sports Physiology and Performance*, *21*(3), 463–471. https://doi.org/10.1123/ijspp.2025-0467

López-Belmonte, Ó., Ruiz-Navarro, J. J., Gay, A., Cuenca-Fernández, F., Arellano, R., & Cejuela, R. (2024). Swimming Performance in Elite Triathletes: Comparison Between Open Water

and Pool Conditions. *Scandinavian Journal of Medicine & Science in Sports*, *34*(8), e14702. https://doi.org/10.1111/sms.14702

López-Belmonte, Ó., Ruiz-Navarro, J. J., Gay, A., Cuenca-Fernández, F., Cejuela, R., & Arellano, R. (2023). Determinants of 1500-m Front-Crawl Swimming Performance in Triathletes: Influence of Physiological and Biomechanical Variables. *International Journal of Sports Physiology and Performance*, *18*(11), 1328–1335. https://doi.org/10.1123/ijspp.2023-0157

Marinof, A., Martínez-Sobrino, J., Del Cerro, J. S., González-Ravé, J. M., & Veiga, S. (2026). Managing Training Load in the Lead-Up to Competition: Session-Rating of Perceived Exertion, not the Acute-to-Chronic Workload Ratio, Associated with the Performance of Elite Swimmers. *Journal of Human Kinetics*, *101*(Special Issue), 271–284. https://doi.org/10.5114/jhk/218950

Mitchell, L. J. G., Rattray, B., Fowlie, J., Saunders, P. U., & Pyne, D. B. (2020). The impact of different training load quantification and modeling methodologies on performance predictions in elite swimmers. *European Journal of Sport Science*, *20*(10), 1329–1338. https://doi.org/10.1080/17461391.2020.1719211

Morais, J. E., Barbosa, T. M., Gomeñuka, N. A., & Marinho, D. A. (2024). Effects of anthropometrics, thrust, and drag on stroke kinematics and 100 m performance of young swimmers using path-analysis modeling. *Scandinavian Journal of Medicine & Science in Sports*, *34*(2), e14578. https://doi.org/10.1111/sms.14578

Morais, J. E., Forte, P., Nevill, A. M., Barbosa, T. M., & Marinho, D. A. (2020). Upper-limb kinematics and kinetics imbalances in the determinants of front-crawl swimming at maximal speed in young international level swimmers. *Scientific Reports*, *10*(1), 11683. https://doi.org/10.1038/s41598-020-68581-3

Morais, J. E., Silva, A. J., Marinho, D. A., Marques, M. C., Batalha, N., & Barbosa, T. M. (2016). Modeling the relationship between biomechanics and performance of young sprinting swimmers. *European Journal of Sport Science*, *16*(6), 661–668. https://doi.org/10.1080/17461391.2016.1149227

Pla, R., Le Meur, Y., Aubry, A., Toussaint, J. F., & Hellard, P. (2019). Effects of a 6-Week Period of Polarized or Threshold Training on Performance and Fatigue in Elite Swimmers. *International Journal of Sports Physiology and Performance*, *14*(2), 183–189. https://doi.org/10.1123/ijspp.2018-0179

Postiglione, G. L., Abbott, S., Newman, P., Mitchell, L. G., Elipot, M., Barclay, G., & Cobley, S. (2025). Race-Performance Parameters Differentiating World-Best From National-Level Swimmers: A Race Video Analysis and Machine-Learning Approach. *International Journal of Sports Physiology and Performance*, *20*(8), 1034–1042. https://doi.org/10.1123/ijspp.2024-0486

Saeterbakken, A. H., Stien, N., Andersen, V., Scott, S., Cumming, K. T., Behm, D. G., Granacher, U., & Prieske, O. (2022). The Effects of Trunk Muscle Training on Physical Fitness and Sport-Specific Performance in Young and Adult Athletes: A Systematic Review and Meta-Analysis. *Sports Medicine (Auckland, N.Z.)*, *52*(7), 1599–1622. https://doi.org/10.1007/s40279-021-01637-0

Seifert, L., & Carmigniani, R. (2023). Coordination and stroking parameters in the four swimming techniques: a narrative review. *Sports Biomechanics*, *22*(12), 1617–1633. https://doi.org/10.1080/14763141.2021.1959945

Seshadri, D. R., Thom, M. L., Harlow, E. R., Gabbett, T. J., Geletka, B. J., Hsu, J. J., Drummond, C. K., Phelan, D. M., & Voos, J. E. (2020). Wearable Technology and Analytics as a Complementary Toolkit to Optimize Workload and to Reduce Injury Burden. *Frontiers in Sports and Active Living*, *2*, 630576. https://doi.org/10.3389/fspor.2020.630576

Staunton, C. A., Ruiz-Navarro, J. J., & Born, D.-P. (2025). Stroke rate-stroke length dynamics in elite freestyle swimming: application of kernel density estimation. *Frontiers in Sports and Active Living*, *7*, 1656633. https://doi.org/10.3389/fspor.2025.1656633

Suijkerbuijk, Y. B., Schaafsma, F. G., van Mechelen, J. C., Ojajärvi, A., Corbière, M., & Anema, J. R. (2017). Interventions for obtaining and maintaining employment in adults with severe mental illness, a network meta-analysis. *The Cochrane Database of Systematic Reviews*, *9*(9), CD011867. https://doi.org/10.1002/14651858.CD011867.pub2

Terzi, E., Skari, A., Nikolaidis, S., Papadimitriou, K., Kabasakalis, A., & Mougios, V. (2021). Relevance of a Sprint Interval Swim Training Set to the 100-Meter Freestyle Event Based on Blood Lactate and Kinematic Variables. *Journal of Human Kinetics*, *80*, 153–161. https://doi.org/10.2478/hukin-2021-0091

Torres, G., Maia, F., Nakamura, F. Y., Neiva, H. P., & Sousa, A. (2025). Rating of perceived exertion in continuous sports: a scoping review with evidence gap map. *Frontiers in Sports and Active Living*, *7*, 1553998. https://doi.org/10.3389/fspor.2025.1553998

Veiga, S., Cala, A., Frutos, P. G., & Navarro, E. (2014). Comparison of starts and turns of national and regional level swimmers by individualized-distance measurements. *Sports Biomechanics*, *13*(3), 285–295.

https://doi.org/10.1080/14763141.2014.910265

Veiga, S., Roig, A., & Gómez-Ruano, M. A. (2016). Do faster swimmers spend longer underwater than slower swimmers at World Championships? *European Journal of Sport Science*, *16*(8), 919–926. https://doi.org/10.1080/17461391.2016.1153727

Vilas-Boas, J. P. (2025). Swimming biomechanics: from the pool to the lab … and back. *Sports Biomechanics*, *24*(5), 1197–1219. https://doi.org/10.1080/14763141.2023.2237003

Wang, Z., Liu, K., Zhao, X., & Gao, J. (2025). Comparative effectiveness of physical training modalities on swimming performance: a two-tier network meta-analysis. *Frontiers in Physiology*, *16*, 1636595. https://doi.org/10.3389/fphys.2025.1636595

IJHRD 915 4.3

ORIGINALITY REPORT

19%	17%	14%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ejournal.gomit.id Internet Source	1%
2	journals.kmanpub.com Internet Source	1%
3	www.mdpi.com Internet Source	1%
4	mdpi-res.com Internet Source	1%
5	Submitted to Coventry University Student Paper	1%
6	Submitted to Australian Catholic University Student Paper	1%
7	revistaretos.org Internet Source	1%
8	Submitted to learn.uq.edu.au Student Paper	<1%
9	theses.lib.polyu.edu.hk Internet Source	<1%
10	Submitted to University of Lincoln Student Paper	<1%
11	Submitted to University of St Mark and St John Student Paper	<1%
12	www.frontiersin.org Internet Source	<1%
13	www.jssm.org Internet Source	<1%
14	reference-global.com Internet Source	<1%
15	www.researchsquare.com Internet Source	<1%

16	Submitted to Universidad Internacional Isabel I de Castilla Student Paper	<1 %
17	hdl.handle.net Internet Source	<1 %
18	ejournal.uinbukittinggi.ac.id Internet Source	<1 %
19	journal.unesa.ac.id Internet Source	<1 %
20	Jeanne Dekerle. "High Performance Youth Swimming", Routledge, 2020 Publication	<1 %
21	Submitted to Tilburg University Student Paper	<1 %
22	dergipark.org.tr Internet Source	<1 %
23	www.j-psp.com Internet Source	<1 %
24	labos-recherche.insep.fr Internet Source	<1 %
25	Submitted to C.E. Byrd High School Student Paper	<1 %
26	Submitted to Frederick University Student Paper	<1 %
27	Submitted to Liberty University Student Paper	<1 %
28	Zubayer IBN Mostafa, Md. Mominul Islam, Shishir Biswas, Md. Najmul Kabir et al. "Flash flood susceptibility mapping in the northeastern Haor basin of Bangladesh using tree-based ensemble models", Spatial Information Research, 2026 Publication	<1 %
29	oa.upm.es Internet Source	<1 %
30	Submitted to Boise State University Student Paper	<1 %
31	asjp.cerist.dz Internet Source	<1 %
32	peerj.com	

	Internet Source	<1%
33	ejournal.unmus.ac.id Internet Source	<1%
34	www.acpjournals.org Internet Source	<1%
35	d-nb.info Internet Source	<1%
36	www.swimming.org.au Internet Source	<1%
37	Samer Mansoor Jameel. "High-resolution three-dimensional biomechanical analysis and machine learning–based prediction of stroke length–stroke rate interaction in competitive freestyle swimming", Pedagogy of Physical Culture and Sports, 2026 Publication	<1%
38	competitor.ojsunm.ac.id Internet Source	<1%
39	d.docksci.com Internet Source	<1%
40	www.sjofsciences.com Internet Source	<1%
41	Marcus Lima Espirito Santos, Pedro Augusto Rucinski, Lucas Lima Gonçalves, Rochelle Rocha Costa, Karini Borges dos Santos. "Parâmetros de natação de nadadores convencionais e nadadores com deficiências físico-motoras a diferentes distâncias", Retos, 2026 Publication	<1%
42	press.um.si Internet Source	<1%
43	Ibrahim Akubat, Ebrahim Patel, Steve Barrett, Grant Abt. "Methods of monitoring the training and match load and their relationship to changes in fitness in professional youth soccer players", Journal of Sports Sciences, 2012 Publication	<1%
44	Jin Zhao, Jing Lian, Xiaolian Wen, Li Ma, Meijing Zheng, Liping Su. "Prognostic factors and nomogram development for adult diffuse large B-cell lymphoma: a single-center retrospective cohort study", BMC Cancer, 2026 Publication	<1%
45	researchonline.ljmu.ac.uk Internet Source	<1%

46	theses.hal.science Internet Source	<1%
47	José M. González-Ravé, Fernando González-Mohino, Francisco Hermosilla Perona, Victor Rodrigo-Carranza et al. "Biomechanical, Physiological and Anthropometric Determinants of Backstroke Swimming Performance: A Systematic Review", Sports Medicine - Open, 2025 Publication	<1%
48	assets-eu.researchsquare.com Internet Source	<1%
49	Arvin Agah. "Medical Applications of Artificial Intelligence - Second Edition", CRC Press, 2026 Publication	<1%
50	Ziyu Liu, Chijun Mu, Zhao Gao, Yupeng Shen. "From discrete classification to continuous space: compositional analysis reveals training intensity distribution migration across competition cycles in competitive swimmers", Frontiers in Physiology, 2026 Publication	<1%
51	discovery.researcher.life Internet Source	<1%
52	pmc.ncbi.nlm.nih.gov Internet Source	<1%
53	Sefa Haktan HATIK, İlker Can BÜYÜKKIRLI, Berkay Eren PEHLİVANOĞLU. "Investigation of the Relationship Between Frontal Plane Knee Projection Angle and Dynamic Balance, Agility, and Ankle Dorsiflexion Mobility in Adolescent Female Volleyball Players", Springer Science and Business Media LLC, 2026 Publication	<1%
54	Submitted to CSU Northridge Student Paper	<1%
55	Submitted to Rijksuniversiteit Groningen - Tii Student Paper	<1%
56	gaexcellence.com Internet Source	<1%
57	ojs.trp.org.in Internet Source	<1%
58	rc.library.uta.edu Internet Source	<1%

59	Diogo D. Carvalho, Márcio Fagundes Goethel, Mégane Erblang, João Paulo Vilas-Boas et al. "Impact of an Overload Period on Heart Rate Variability, Sleep Quality, Motivation, and Performance in High-level Swimmers: Use of Explainable Artificial Intelligence (XAI) to Assess Training Load Variations", Sports Medicine, 2025 Publication	<1 %
60	Lourdes Sánchez, Raúl Arellano, Francisco Cuenca-Fernández. "Analysis and influence of the underwater phase of breaststroke on short-course 50 and 100m performance", International Journal of Performance Analysis in Sport, 2021 Publication	<1 %
61	Submitted to University of Greenwich Student Paper	<1 %
62	c.coek.info Internet Source	<1 %
63	skemman.is Internet Source	<1 %
64	www.duvaryayinlari.com Internet Source	<1 %
65	Eleftherios Fyntikakis, Spyridon Plakias, Themistoklis Tsatalas, Minas A. Mina, Anthi Xenofondos, Christos Kokkotis. "Convergence of Artificial Intelligence and Wearables in Strength Training and Performance Monitoring: A Scoping Review", Applied Sciences, 2026 Publication	<1 %
66	Submitted to Loughborough University Student Paper	<1 %
67	journals.lww.com Internet Source	<1 %
68	www.apgads.lu.lv Internet Source	<1 %
69	Submitted to Athens Metropolitan College Student Paper	<1 %
70	Merve Senol Aydogan, Ani Agopyan, Erdem Uylas, Erkan Tortu, Erkan Gunay. "Feasibility and preliminary agreement of a lower-volume, pool-based incremental swimming protocol in trained male university swimmers: a pilot study", BMC Sports Science, Medicine and Rehabilitation, 2026 Publication	<1 %
71	Submitted to UCL Student Paper	<1 %

72	amsdottorato.unibo.it Internet Source	<1 %
73	bustechno.polteksci.ac.id Internet Source	<1 %
74	sponet.de Internet Source	<1 %
75	www.jhse.es Internet Source	<1 %
76	www.tera.org Internet Source	<1 %
77	"Genetics and Sports Performance", Springer Science and Business Media LLC, 2026 Publication	<1 %
78	Andrew Sheaff. "A Constraints-Led Approach to Swim Coaching", Routledge, 2023 Publication	<1 %
79	Rashid Sapargaliyev, Nurlan Jainakbayev, Sabina Muzrapova, Nurlan Baisynov, Firuza Numanova, Rimma Sembaeva. "Longitudinal Metabolic and Perceptual Responses to an Individualized, Biochemically Guided Interval-Swimming Protocol: A Non-Randomized Controlled Pilot Study", Sports, 2026 Publication	<1 %
80	ejournal.uin-suka.ac.id Internet Source	<1 %
81	researchportal.vub.be Internet Source	<1 %
82	revistas.um.es Internet Source	<1 %
83	sdmoc9.wixsite.com Internet Source	<1 %
84	www.tandfonline.com Internet Source	<1 %
85	www.academicpublishers.org Internet Source	<1 %
86	LAURIE A. MALONE. "Effects of visual impairment on stroke parameters in Paralympic swimmers", Medicine & Science in Sports & Exercise, 12/2001 Publication	<1 %

87	Michal Bozděch. "Advanced Artificial Intelligence in Sports Science - A Guide to Models, Applications, and Ethical Challenges", Routledge, 2026 Publication	<1 %
88	Ryan Keating, Rodney Kennedy, Carla McCabe. "Applying Load–Velocity Profiling to Guide In-Water Resistance Training in an Olympic-Level Swimmer: A Case Study", Applied Sciences, 2025 Publication	<1 %
89	Xiaotong Chen, Hezhi Wang, Shuxin Wang, Ziyu Liu, Yupeng Shen. "From acute responses to long-term monitoring: a scoping review of the reliability and validity of biomechanical and physiological assessment tools in elite and highly trained swimmers", BMC Sports Science, Medicine and Rehabilitation, 2026 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		