Open Access Library Journal
2026, Volume 13, e14824
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Machine Learning Methods in Competitive Swimming Analysis: Paradigm Evolution, Architectural Framework, and Future Prospects

Sujing Su, Houwei Zhu*

College of Physical Education and Health Sciences, Zhejiang Normal University, Jinhua, China
Email: *zhuhouwei@zjnu.edu.cn

How to cite this paper: Su, S.J. and Zhu, H.W. (2026) Machine Learning Methods in Competitive Swimming Analysis: Paradigm Evolution, Architectural Framework, and Future Prospects. *Open Access Library Journal*, **13**: e14824.
https://doi.org/10.4236/oalib.1114824

Received: December 30, 2025
Accepted: January 11, 2026
Published: January 14, 2026

Abstract

In competitive swimming, victory is decided by hundredths of a second, yet traditional biomechanical analysis often struggles with data complexity and delayed feedback. This review addresses the current research fragmentation by proposing an innovative four-dimensional framework (Data, Feature, Model, and Application layers) to systematically categorize the machine learning (ML) landscape in swimming analysis. We deconstruct the racing process into macro-phases (start, swim-through, turns, and finish) and micro-stroke cycles, highlighting how deep learning architectures like Transformers are replacing manual annotation with automated action segmentation. Furthermore, the study identifies key performance predictors—such as Intra-cyclic Velocity Fluctuations (IVV), Stroke Index (SI), and Countermovement Jump (CMJ) impulse—as critical inputs for robust feature engineering. By integrating Explainable AI (XAI), this study bridges the gap between complex "black box" models and actionable coaching insights. Ultimately, we provide a methodological roadmap for building "digital twins" of athletes, shifting swimming science toward a synergy of biomechanical depth and computational precision.

Subject Areas

Sports Science

Keywords

Swimming Performance, Machine Learning, Biomechanics, Feature Engineering, Explainable AI

1. Introduction

Under the eternal theme of "Citius, Altius, Fortius" (Higher, Faster, Stronger) in competitive sports, swimming stands as a racing discipline where victories are decided by hundredths of a second. Consequently, the micro-composition and optimization of performance have remained the central focus of sports science research. Swimming performance is not a simple single-point output; rather, it is a complex product of non-linear superposition and dynamic coupling across several heterogeneous sub-phases: the start, turns, stroking (swim-through), and the finish [1]. Each phase involves unique biomechanical, physiological, and fluid dynamic principles, and the transitions and continuity between these phases are decisive for the final outcome. Therefore, conducting refined, quantitative analysis of the racing process to deeply understand the complex mapping between phase-specific performance and total time is vital for scientific training, technical diagnosis, and tactical formulation.

For a long time, swimming race analysis has primarily relied on traditional biomechanical measurements (e.g., high-speed videography, underwater filming, force plates) and classical statistical methods (e.g., correlation analysis, regression models) [2]. These methods established the empirical foundation of the field, revealing fundamental associations between key parameters—such as stroke rate, stroke length, and turn time—and overall performance. However, as competitive levels approach the limits of human potential and data acquisition technologies (e.g., Inertial Measurement Units (IMUs), pressure sensors, panoramic video systems) undergo explosive growth, the limitations of traditional paradigms have become increasingly prominent.

First, traditional biomechanical analysis, particularly video-based methods, often faces hurdles such as expensive equipment, complex deployment, time-consuming analysis, and delayed feedback. The unique nature of the underwater environment makes it difficult to directly apply precision technologies like optical motion capture, creating a bottleneck in data acquisition [3]. Second, traditional statistical models (e.g., linear regression) are often built on simplifying assumptions like linear relationships and feature independence. These struggle to capture the complex interactions and non-linear dynamic evolution among multiple factors (technique, physical fitness, tactics, psychology) inherent in swimming. Furthermore, feature selection depends heavily on the researcher's experience and prior knowledge, potentially overlooking deep patterns hidden in the data, which limits the predictive power and generalizability of such models.

Against this backdrop, Machine Learning (ML)—as the core branch of Artificial Intelligence—has brought a revolutionary paradigm shift to sports race analysis through its powerful data-driven modeling, complex pattern recognition, and high-dimensional non-linear fitting capabilities. Its evolution path clearly shows a trend moving from descriptive analysis (e.g., clustering to identify athlete types) to predictive modeling (e.g., forecasting race times or technical efficacy) and toward decision support systems (e.g., personalized training recommendations and real-time technical adjustments) [4].

This trend has been successfully validated in other cyclic racing events, such as track and field and cycling. For instance, in track and field, case studies have shown that AI-generated personalized training plans can yield measurable performance improvements of up to 15% for elite athletes [4]. In road cycling, research teams have utilized historical data and team performance characteristics to build ML models for predicting single-day race results, achieving accuracy that meets or even exceeds that of human experts. These cross-disciplinary successes provide a valuable paradigm and confidence for the deepened application of machine learning in swimming [4].

While ML applications in swimming are flourishing, the field remains in an exploratory stage with several deficiencies that require clarification: High Data Heterogeneity: Data sources are diverse, ranging from 2D/3D video sequences and IMU sensor streams to hydrodynamic force data and structured historical records (e.g., USA Swimming's database of over 4 million records). Effectively cleaning, aligning, and fusing this multi-modal, multi-rate, and multi-scale data to extract unified representative features remains a fundamental challenge. The Gap Between Interpretability and Practice: Although complex models like Deep Learning have excelled in tasks like swimmer detection and stroke rate identification (e.g., the SwimTrack challenge), their "black box" nature makes it difficult for coaches and athletes to trust the underlying logic, leading to caution in applying them to critical technical adjustments. While Explainable AI (XAI) methods like SHAP have been introduced to identify key variables, translating model outputs into actionable training insights requires further exploration. Fragmented Research: Existing studies are scattered across various algorithms, from SVMs and Random Forests to LSTMs and Transformers. However, these are often task-specific (e.g., stroke recognition only) and lack a systematic algorithmic architecture. There is currently no "Panorama" review that follows a complete ML pipeline: Data → Feature → Model → Application Scenario. This has resulted in a lack of unified performance benchmarks and consensus on "optimal" models for specific swimming tasks [5].

Given these gaps, this review aims to provide a systematic taxonomy of the machine learning spectrum in swimming race analysis, moving beyond a simple list of cases to construct a clear technical map [6]. We propose an innovative four-dimensional analysis framework (See **Table 1**).

Table 1. An innovative four-dimensional analysis framework.

Layer	Focus Area
Data Layer	Multi-source data types, acquisition methods, preprocessing challenges, and fusion strategies.
Feature Layer	Extraction of time-domain, frequency-domain, and high-level semantic features for the start, turn, stroke, and finish phases.

Continued

Model Layer	Categorization and performance comparison of supervised, unsupervised, deep learning, and ensemble learning (e.g., WoCC) models.
Application Layer	Mapping model outputs to technical diagnosis, performance prediction, training load optimization, and talent identification.

Finally, this review provides a critical analysis and future outlook, identifying directions such as real-time feedback systems, the coupling of XAI with coaching decision systems, and Reinforcement Learning for personalized strategy generation. We expect this work to provide a theoretical roadmap and methodological "toolbox" for researchers and coaches, driving ML in swimming from experimental exploration toward systematic, practical application.

2. The Theory and Practice of Phase Division in Competitive Swimming

The decomposition of a continuous fluid motion into discrete, manageable phases is the fundamental prerequisite for any rigorous biomechanical or tactical analysis. In the context of competitive swimming, phase division does more than provide a unified framework for data alignment; it serves as a "diagnostic scalpel" that allows researchers and coaches to isolate the specific mechanical efficiencies or failures that manifest at different velocities and environmental constraints [7] [8].

2.1. Macro-Level Standardization and Hydrodynamic Transitions

From a macro-perspective, the swimming community has converged on a four-stage standardization: the Start, Swimming (Clean Swimming), Turn, and Finish. While these categories appear intuitive, their boundaries are defined by fundamental shifts in the athlete's hydrodynamic environment [9] [10].

The Start Phase (signal to 15 m) represents a complex transition from terrestrial ballistics to underwater fluid dynamics. Biomechanically, this phase is subdivided into the dry-land component (reaction time and impulse on the block), the flight phase, and the entry/underwater phase. The latter is governed by the Froude Number theory, which posits that swimmers must reach a specific depth (typically 0.6 m to 1.0 m) to minimize Wave Drag, the resistance caused by surface wake. Machine learning models, particularly those utilizing Convolutional Neural Networks (CNNs), are now employed to identify the "Breakout" point—the exact moment a swimmer transitions from underwater dolphin kicking to surface swimming—to evaluate whether the athlete maximized the low-drag environment of the deep water [11] [12].

Similarly, the Turn Phase is a non-linear process of momentum reversal. It involves a controlled deceleration (approach), a complex angular momentum exchange (the rotation), and a high-velocity push-off. The integration of Inertial Measurement Units (IMUs) with Recurrent Neural Networks (RNNs) has enabled the automated segmentation of these sub-micro-phases by detecting specific peaks in the longitudinal acceleration signal, allowing for the precise measurement of "wall contact time" versus "underwater glide efficiency" [13] [14].

2.2. Micro-Technical Segmentation: The Ambiguity of Stroke Cycles

When the analysis descends to the level of a single Stroke Cycle, the lack of a unified definition for phase division becomes a significant hurdle. This inconsistency is not merely a failure of standardization but a reflection of the complexity of Unsteady Hydrodynamics [15].

In the most basic Two-Phase Model, the cycle is split into the Propulsive Phase and the Recovery Phase. However, modern research identifies that propulsion is not a binary state. During the "Propulsive Phase," forces are generated not only through drag-based pressure on the palm (the paddle effect) but also through the generation of high-energy vortices around the forearm (the lift effect) [15].

To capture this, researchers utilize Four-Phase Models (Entry, Catch, Pull, Push). The "Catch" is perhaps the most critical transition point for machine learning algorithms to identify. Biomechanically, it is the moment when the hand's vector changes from a forward/downward search for "still water" to a backward-directed application of force. In the realm of feature engineering, this is often identified as the point of "slope mutation" in pressure sensor data or the reversal of angular velocity in the wrist.

2.3. Algorithmic Automation and Action Segmentation

The future of swimming analysis lies in moving from manual video annotation to End-to-End Action Segmentation. Traditional heuristic algorithms, which rely on fixed thresholds, often fail due to "technique drift" caused by fatigue or individual stylistic variations [15].

Current state-of-the-art approaches employ Temporal Convolutional Networks (TCNs) and Transformer architectures to perform automated labeling of sensor or video sequences. The Self-Attention mechanism in Transformers is particularly effective at identifying long-range dependencies—for instance, how a slight misalignment during the "Entry" phase affects the peak force produced during the "Push" phase several hundred milliseconds later. Furthermore, Unsupervised Clustering is being explored to discover "data-driven phases." By allowing the model to define its own segments based on structural patterns in the data, researchers may uncover micro-technical transitions that are invisible to the human eye but statistically significant for performance prediction [16].

2.4. Theoretical Foundations and Neuromuscular Control

The rationale behind these divisions is increasingly rooted in Neuromuscular Control Patterns. Although often overlooked, each phase corresponds to a specific "muscle synergy" or activation sequence. Theoretical models involving Central Pattern Generators (CPGs) suggest that the rhythmic nature of swimming is governed by neural oscillators. Therefore, the boundaries between phases can be redefined as the "inflection points" where the nervous system shifts its firing pattern from one group of synergists (e.g., the latissimus dorsi during the pull) to another (e.g., the deltoids during recovery) [17].

Ultimately, the choice of a phase division model must be analysis-driven. While a coach may require a simple breakdown of the "Catch-to-Push" duration to assess power, a biomechanist might require a multi-stage model to analyze the elbow's verticality during the "Late Pull." The challenge for machine learning is to build flexible frameworks that can adapt to these varying levels of granularity while maintaining a standardized data structure for cross-athlete comparison [18].

3. Key Indicators for Predicting Swimming Performance: Integration of Multi-Dimensional Feature Engineering and Biomechanical Modeling

Once the structural framework of phase division is established, the central challenge shifts to extracting quantitative indicators within these discrete temporal windows. In the context of Machine Learning (ML), this process is known as Feature Engineering. High-quality feature engineering not only enhances the predictive accuracy of models but also ensures that the outputs remain biomechanically interpretable, bridging the gap between raw data and coaching insights [18].

3.1. The Non-Linear Evolution of Core Kinematic Parameters

Kinematic indicators serve as the direct empirical foundation of performance evaluation. However, modern race analysis has pivoted from "mean value statistics" toward "instantaneous fluctuation analysis," providing a higher resolution of technical efficiency [19].

Intra-cyclic Velocity Fluctuations (IVV): While average velocity remains the primary determinant of race outcomes, IVV reveals the underlying mechanics of energy dissipation. Due to the pulsatile nature of propulsive forces and the continuous counteraction of fluid drag, velocity fluctuates cyclically within a single stroke [19]. Research indicates that IVV is positively correlated with the Energy Cost (CoT) of swimming; smaller fluctuations signify a more streamlined profile and higher propulsive continuity. By utilizing high-sampling-rate signals from Inertial Measurement Units (IMUs), ML models compute the Coefficient of Variation (CV) or Root Mean Square Error (RMSE) of the velocity curve to sensitively capture an athlete's ability to maintain velocity stability under the onset of fatigue [20].

The Dynamic Coupling of Stroke Rate (SR) and Stroke Length (SL): The fundamental equation $V = SR * SL$ is deceptively simple. In elite competition, these two variables exhibit a complex, non-linear antagonistic relationship where gains in one often come at the expense of the other. In feature engineering, single metrics like SR or SL are often insufficient to characterize performance. Advanced models incorporate Stroke Gait Features, calculating the rate of change in the SR-SL ratio to identify an athlete's “efficiency signature.” This allows models to discern whether an athlete is sacrificing efficiency for frequency—as typically seen in the final 5 m sprint—or maintaining an optimal propulsive distance per cycle [21].

3.2. Deep Mining of Kinetic and Biomechanical Parameters

Kinetic features explore the origin of motion—the application of force and energy—which is vital for understanding the underlying athletic capacity and physiological ceiling of a swimmer [22].

Impulse and Explosive Power Representation: Impulse is the integral of force over time and serves as a proxy for effective work. During the start and turn phases, the Countermovement Jump (CMJ) Impulse generated against the block or wall has been identified by ML models—particularly through SHAP value analysis—as one of the most significant predictive variables. By extracting features from the force-time curve, such as time-to-peak force and the Rate of Force Development (RFD), models can quantify the efficiency of an athlete’s explosive power utilization and their ability to transition land-based strength into aquatic propulsion [23].

4. Conclusions: Toward an Intelligent Synergy of Theory and Data

The integration of Machine Learning into competitive swimming analysis represents a transformative leap from subjective technical intuition toward a rigorous, data-driven diagnostic paradigm. As demonstrated throughout this study, the effectiveness of these advanced algorithms is fundamentally predicated on a robust theoretical foundation. By systematically aligning standardized phase division with high-density biomechanical features, we move beyond mere performance prediction toward a profound understanding of the “why” behind elite athletic achievement [24].

The core of this evolution lies in the transition from viewing swimming as a collection of isolated variables to treating it as a complex, non-linear dynamic system. The dynamic coupling of propulsion and drag, the delicate trade-off between stroke rate and stroke length, and the physiological constraints on mechanical efficiency are no longer “black boxes” of sports science. Instead, through the application of deep learning architectures and Explainable AI (XAI) tools like SHAP, these intricate relationships are now quantifiable and actionable [25].

Looking forward, the future of swimming science resides in the creation of holistic digital twins—models that fuse multi-modal data streams (video, IMUs, and physiological sensors) to provide real-time, personalized tactical interventions. By bridging the gap between sophisticated algorithmic modeling and the practical needs of the poolside environment, we empower coaches and athletes with a “methodological toolbox” capable of pushing the boundaries of human potential. Ultimately, the synergy of biomechanical depth and computational breadth will define the next era of aquatic excellence, ensuring that the pursuit of a hundredth of a second is guided by both the art of coaching and the precision of intelligence [26].

Acknowledgements

The present study was funded by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (No. 2024C03260 and No. 2023C03197), the Ministry of Education Humanities and Social Science Research Youth Fund Project (No. 22 YJC890056), the National Social Science Fund of China (No. 24 CTY025), the Zhejiang Provincial College Students' Scientific and Technological Innovation Activity Program (New Seedling Talent Program) (No. 2025R404A011) and Zhejiang Normal University's 2025 Provincial Undergraduate Innovation and Entrepreneurship Training Program (No. S202510345121).

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] González-Ravé, J.M., González-Mohino, F., Hermosilla Perona, F., Rodrigo-Carranza, V., Yustres, I. and Pyne, D.B. (2025) Biomechanical, Physiological and Anthropometric Determinants of Backstroke Swimming Performance: A Systematic Review. *Sports Medicine—Open*, **11**, Article No. 68. https://doi.org/10.1186/s40798-025-00868-z

[2] Figueiredo, P., Pendergast, D.R., Vilas-Boas, J.P. and Fernandes, R.J. (2013) Interplay of Biomechanical, Energetic, Coordinative, and Muscular Factors in a 200 m Front Crawl Swim. *BioMed Research International*, **2013**, Article ID: 897232. https://doi.org/10.1155/2013/897232

[3] Kholkine, L., De Schepper, T., Verdonck, T. and Latré, S. (2020) A Machine Learning Approach for Road Cycling Race Performance Prediction. In: Brefeld, U., Davis, J., Van Haaren, J. and Zimmermann, A., Eds., *Machine Learning and Data Mining for Sports Analytics*, Springer, 103-112. https://doi.org/10.1007/978-3-030-64912-8_9

[4] Targosinski, S., Henzi, A., Engmann, A.K., Rushing, E.J., Barth, A.A., Klein, H.J., *et al.* (2022) A Swim Test for Functional Assessment of Rodent Peripheral Nerve Regeneration. *Journal of Neuroscience Methods*, **379**, Article ID: 109663. https://doi.org/10.1016/j.jneumeth.2022.109663

[5] Cepeda-Pacheco, J.C. and Domingo, M.C. (2022) Deep Learning and 5G and Beyond for Child Drowning Prevention in Swimming Pools. *Sensors*, **22**, Article 7684. https://doi.org/10.3390/s22197684

[6] Sánchez La, B.C. (2024) Utilización de Machine Learning para proporcionar a los deportistas acuáticos herramientas innovadoras para la selección del aparejo a utilizar. Universitat Oberta de Catalunya (UOC).

[7] Arellano, R., Pardillo, S. and García, F. (1999) A System for Quantitative Measurement of Swimming Technique. *Biomechanics and Medicine in Swimming*, **VIII**, 1269-275.

[8] Xu, K., Yin, M., Zhong, Y., Huang, K., Xu, Y., Zhou, J., *et al.* (2025) Acute Effects of Assisted and Resisted Sprint Training on Subsequent Sprint Performance: A System-

atic Review and Meta-Analysis. *Journal of Strength & Conditioning Research*, **39**, e711-e720. https://doi.org/10.1519/jsc.0000000000005087

[9] Cortesi, M., Di Michele, R. and Gatta, G. (2019) Effects of Intracyclic Velocity Variations on the Drag Exerted by Different Swimming Parachutes. *Journal of Strength and Conditioning Research*, **33**, 531-537. https://doi.org/10.1519/jsc.0000000000002010

[10] Cahill, M.J., Oliver, J.L., Cronin, J.B., Clark, K., Cross, M.R., Lloyd, R.S., *et al.* (2020) Influence of Resisted Sled-Pull Training on the Sprint Force-Velocity Profile of Male High-School Athletes. *Journal of Strength and Conditioning Research*, **34**, 2751-2759. https://doi.org/10.1519/jsc.0000000000003770

[11] Hedges, L.V. and Olkin, I. (1985) Random Effects Models for Effect Sizes. In: Hedges, L.V. and Olkin, I., Eds., *Statistical Methods for Meta-Analysis*, Elsevier, 189-203. https://doi.org/10.1016/b978-0-08-057065-5.50014-2

[12] Stewart, A.M. and Hopkins, W.G. (2000) Consistency of Swimming Performance within and between Competitions. *Medicine & Science in Sports & Exercise*, **32**, 997-1001. https://doi.org/10.1097/00005768-200005000-00018

[13] Xu, K., Jukic, I., Cross, M.R., Hicks, D.S., Yin, M., Zhong, Y., *et al.* (2025) Effects of Resisted-Sprint Training on Sprint Performance and Mechanics: A Systematic Review and Meta-analysis Focusing on Load Magnitude. *Scandinavian Journal of Medicine & Science in Sports*, **35**, e70182. https://doi.org/10.1111/sms.70182

[14] Beretić, I., đurović, M. and Okičić, T. (2012) Influence of the Back Plate on Kinematical Starting Parameter Changes in Elite Male Serbian Swimmers. *Physical Education and Sport*, **10**, 135-140.

[15] Xu, K., Blazevich, A.J., Boullosa, D., Ramirez-Campillo, R., Yin, M., Zhong, Y., *et al.* (2025) Optimizing Post-Activation Performance Enhancement in Athletic Tasks: A Systematic Review with Meta-Analysis for Prescription Variables and Research Methods. *Sports Medicine*, **55**, 977-1008. https://doi.org/10.1007/s40279-024-02170-6

[16] Morais, J.E., Barbosa, T.M., Silva, A.J., Veiga, S. and Marinho, D.A. (2021) Profiling of Elite Male Junior 50 M Freestyle Sprinters: Understanding the Speed-Time Relationship. *Scandinavian Journal of Medicine & Science in Sports*, **32**, 60-68. https://doi.org/10.1111/sms.14058

[17] Coloretti, V., Fantozzi, S., Gatta, G., Bonifazi, M., Zamparo, P. and Cortesi, M. (2025) Quantifying Added Drag in Swimming with Parachutes: Implications for Resisted Swimming Training. *Journal of Strength & Conditioning Research*, **39**, e701-e705. https://doi.org/10.1519/jsc.0000000000005109

[18] Cohen, J. (2013) Statistical Power Analysis for the Behavioral Sciences. Routledge.

[19] Shen, C., He, Z., Ye, B. and Zhu, H. (2025) The Effect of Competitive Level and Gender on the Interval Speed Characteristics and Pacing Strategies of High-Level 100 m Backstroke Athletes in China. *Applied Sciences*, **15**, Article 11195. https://doi.org/10.3390/app152011195

[20] Maglischo, E.W., Maglischo, C.W., Zier, D.J., *et al.* (1985) The Effect of Sprint-Assisted and Sprint-Resisted Swimming on Stroke Mechanics. *Journal of Swimming Research*, **1**, 27-33.

[21] Oğul, B., Uslu, S., Hindistan, I.E., Akdağ, E. and Cetin Özdoğan, E. (2023) The Effect of the Horizontal vs. Vertical PAPE Protocol on the Swim Start Performance in Adolescent Male. *Sports Biomechanics*, **24**, 1220-1235. https://doi.org/10.1080/14763141.2023.2200745

[22] Murray, A., Aitchison, T., Ross, G., Sutherland, K., Watt, I., McLean, D., *et al.* (2005) The Effect of Towing a Range of Relative Resistances on Sprint Performance. *Journal*

of Sports Sciences, **23**, 927-935. https://doi.org/10.1080/02640410400023332

[23] Zafeiridis, A., Saraslanidis, P., Manou, V., *et al.* (2005) The Effects of Resisted Sled-pulling Sprint Training on Acceleration and Maximum Speed Performance. *Journal of Sports Medicine and Physical Fitness*, **45**, 284-290.

[24] Prieske, O., Behrens, M., Chaabene, H., Granacher, U. and Maffiuletti, N.A. (2020) Time to Differentiate Postactivation "Potentiation" from "Performance Enhancement" in the Strength and Conditioning Community. *Sports Medicine*, **50**, 1559-1565. https://doi.org/10.1007/s40279-020-01300-0

[25] Cross, M.R., Lahti, J., Brown, S.R., Chedati, M., Jimenez-Reyes, P., Samozino, P., *et al.* (2018) Training at Maximal Power in Resisted Sprinting: Optimal Load Determination Methodology and Pilot Results in Team Sport Athletes. *PLOS ONE*, **13**, e0195477. https://doi.org/10.1371/journal.pone.0195477

[26] Finlay, M.J., Bridge, C.A., Greig, M. and Page, R.M. (2021) Upper-Body Post-Activation Performance Enhancement for Athletic Performance: A Systematic Review with Meta-Analysis and Recommendations for Future Research. *Sports Medicine*, **52**, 847-871. https://doi.org/10.1007/s40279-021-01598-4