# A systematic literature review of swimming performance prediction: methods, datasets, techniques and research trends

*Una revisión sistemática de la literatura sobre predicción del rendimiento en natación: métodos, datos, técnicas y tendencias*


## Authors

Ari Tri Fitrianto [1]
Oddie Barnanda Rizky [2]
Edi Rahmadi [3]
Asary Ramadhan [3]

[1] University of Islam Kalimantan MAB (Indonesia)
[2] University of Bengkulu (Indonesia)
[3] Universitas Lambung Mangkurat (Indonesia)

Corresponding author:
Ari Tri Fitrianto
aritri.fitrianto@uniska-bjm.ac.id



## How to cite in APA

Tri Fitrianto, A., Barnanda Rizky, O., Rahmadi, E., & Ramadhan, A. (2025). A systematic literature review of swimming performance prediction: methods, datasets, techniques and research trends. *Retos*, *67*, 482–497. https://doi.org/10.47197/retos.v67.112197



## Abstract

Introduction: Predicting swimming success in competitive sports, primarily the outcomes of forthcoming Olympic swimming competitions.
Objective: This paper provides an extensive and systematic review of research in swimming performance prediction published from 2014 to 2024.
Method: This swimming performance prediction research was conducted with a Systematic Literature Review (SLR). In addition, to create article boundaries by the research topic that was reviewed, this study used projects for systematic reviews and meta-analysis (PRISMA) guidelines to conduct the systematic review. There are 21 journal publications are the result of the extraction of selected studies for identification and analysis to describe research topics or trends, datasets, techniques, methods, evaluations, and problems in this research field.
Results: The results of the analysis provide an in-depth explanation of the topics or trends that are the focus of their studies in the field of swimming performance prediction, provide references to public datasets, and explain the techniques and methods that researchers often use to compare and develop methods.
Discussion: A predictive mathematical model is a favourite technique because it integrates complex biologica health, biomechanical variables, providing precise performance prediction. Additionally, to improve both the accuracy and interpretability of predictions need for hybrid approaches that combine mathematical models with more advanced techniques, such as machine learning and explainable artificial intelligence (XAI).
Conclusions: Swimming performance prediction plays a crucial role in enhancing training programs, guiding athlete selection, and evaluating progress.



## Keywords

artificial intelligence; picoc; statistical model; swimming



## Resumen

Introducción: Predecir el éxito en la natación en deportes competitivos, principalmente los resultados de las futuras competiciones olímpicas de natación.
Objetivo: Este artículo ofrece una revisión exhaustiva y sistemática de la investigación sobre la predicción del rendimiento en natación publicada entre 2014 y 2024.
Metodología: La investigación sobre la predicción del rendimiento en natación se realizó mediante una Revisión Sistemática de la Literatura (SLR). Además, para establecer los límites de los artículos según el tema de investigación, este estudio utilizó las directrices PRISMA para realizar la revisión sistemática. El resultado de la extracción de estudios seleccionados fue la identificación y análisis de 21 publicaciones científicas que describen los temas de investigación, conjuntos de datos, técnicas, métodos, evaluaciones y problemas en este campo.
Resultados: El análisis proporciona una explicación detallada de los temas y tendencias que centran los estudios en la predicción del rendimiento en natación, ofrece referencias a conjuntos de datos públicos y explica las técnicas y métodos empleados por los investigadores.
Discusión: El modelo matemático predictivo es una técnica popular, ya que integra variables biológicas y biomecánicas complejas, proporcionando predicciones precisas. Además, para mejorar tanto la precisión como la interpretabilidad de las predicciones, es necesario el uso de enfoques híbridos que combinen modelos matemáticos con técnicas más avanzadas, como el aprendizaje automático y la inteligencia artificial explicable (XAI).
Conclusiones: La predicción del rendimiento en natación es fundamental para mejorar los programas de entrenamiento, orientar la selección de atletas y evaluar su progreso.



## Palabras clave

Inteligencia artificial; picoc; modelo estadístico; natación

## Introduction

Swimming performance is a multifactorial phenomenon (de Anda Martín et al., 2024; Morais et al., 2022). It is similar to anthropometrics, hydrodynamics, kinematics and energetics which are scientific domains of multifactorial phenomena (Morais et al., 2021). Several authors have estimated and/or measured variables in various scientific domains (i.e. anthropometric, hydrodynamic, kinematic and energetic) that are easy to collect and/or can predict performance and detect talented swimmers (Lobato et al., 2023; Marinho et al., 2020). The main criterion for the success of swimming performance is to complete a certain race distance in the shortest possible time. Research has found that swimming speed is the product of the paddle's speed and length, which means increasing one of the two will increase the swimmer's performance (Nurmukhanbetova et al., 2023). It has also been suggested that increasing swimming speed requires high frequency, duration, and intensity. Where a high total volume of exercise might be expected (Armen et al., 2024).

Furthermore, swimming performance is highly dependent on muscle strength (Apriyano et al., 2025; Sadewa et al., 2024) and closely connected to physiological adaptations that are induced by the athlete's training program (Nugent et al., 2019). The area of swimming performance prediction research has been studied since the 20th century, which was discussed openly by (Banister & Calvert, 1980; Busso et al., 1990, 1997; Chatard & Stewart, 2011; Fitz-Clarke JR et al., 1991; Hohmann, 1992) with a linear mathematical concepts like linear differential equations or regression analysis. However, biological adaptation is a complex non-linear problem because the adaptation of a biological system leads to changes in the system itself, that is, the adaptive behavior can change. Further, it is commonly known that double training input does not lead to double performance output. Hence, linear models can only approximate the non-linear adaptive behavior in a very small range of the modeled performance output (Edelmann-Nusser et al., 2002). That is why, today swimming performance prediction is a significant research topic in the sport science field (Staunton et al., 2024a).

To explore more opportunities in research in this field, this research using Systematic Literature Review (SLR) technique to get the results of the exploration of a more systematic, measurable, and diverse topics that reviews more diverse and wider. The advantages of SLR over traditional review techniques are the use of scientific methods and their systematic work (Okoli & Schabram, 2012) to minimize bias, and the results are clear and can be accounted for. Because of its definite method of operation, SLR can confidently provide input to policymakers. Various studies conducted from 2010 to the end of 2024 are categorized into research questions (RQ) group.

## Method

### *Review Method*

This swimming performance prediction research was conducted with a Systematic Literature Review (SLR). SLR is a method to identify, evaluate, and interpret the results of research that has been conducted as a whole that is relevant to the topic area or research questions that aim to provide answers to research questions (Okoli & Schabram, 2012). According to the reviewer's feedback during the revision of this study, in general, there are two parts carried out in SLR namely the first part is the planning phase and the second part is the implementation and report phase. In the planning phase there are 3 steps, the first is to identify the need for a systematic review, the second is to develop a review protocol and the last is to evaluate the preview. The next is in the implementation and report phase there are 4 steps, namely, the first search for primary, the second select primarys study, the third extract data from primary studies and the fifth disseminate results. The overall process of this SLR is presented in Figure 1.

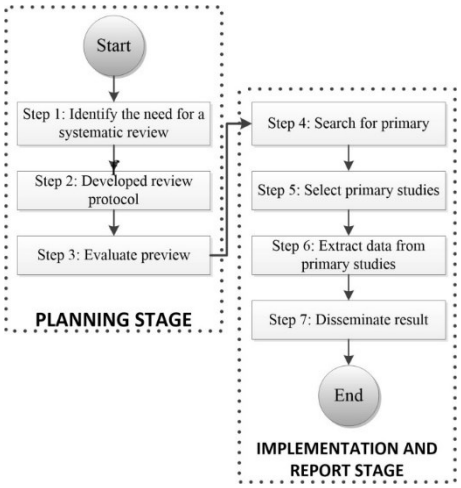
Table 1. PICOC Criteria

| | |
|---|---|
| Population | Swimming Performance Prediction |
| Intervention | The method in swimming performance prediction |
| Comparison | Different predictive models (Artificial Intellegence vs statical model |
| Outcomes | Swimming performance prediction |
| Context | Study in sports laboratories with small and large dataset |

### *Research Question (RQ)*

RQ is structured to facilitate the review process to be more focused and consistent. In general, the research questions were prepared using meaningful PICOC criteria (Population, Intervention, Comparison, Results, dan Context) (Kitchenham & Charters, 2007) to compute five inclusion/exclusion criteria (Rico-González et al., 2022). PICO criteria shown in Table 1.

Figure 1. Systematic Literature Review phase.



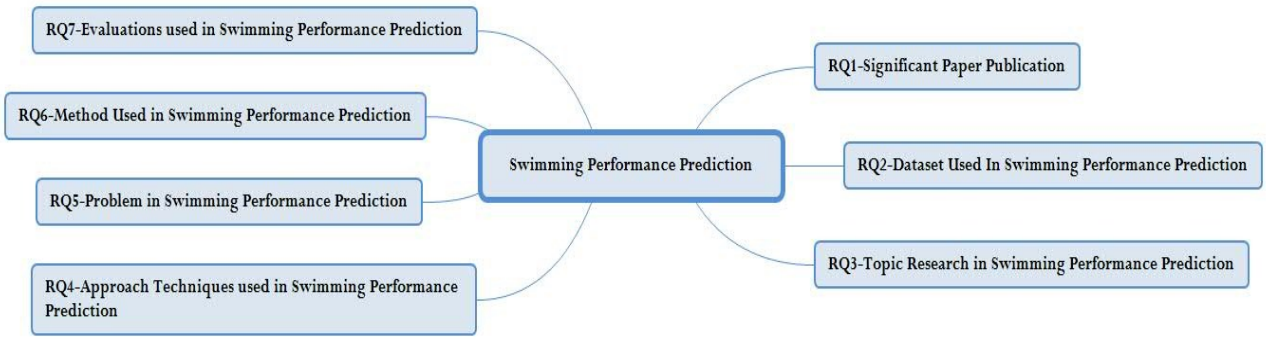

The research and motivation questions in this literature review are described in Table 2. RQ2 and RQ7 are the leading research study questions, while RQ1 and RQ3 are assigned to help evaluate the context of the main study. RQ2 shows the dataset and RQ4 to RQ7 shows the approach techniques, problems, methods, and evaluations. RQ1 and RQ3 provide a synopsis of certain research areas in swimming performance prediction.

Table 2. Research Question and Motivation

| ID | Research question | Motivation |
|---|---|---|
| RQ1 | What journals/conference papers explore the swimming performance prediction? | Identify the most significant journal/conference papers in the swimming performance prediction. |
| RQ2 | What datasets are used in swimming performance prediction? | Identify datasets commonly used in swimming performance prediction. |
| RQ3 | What journals/conference papers about swimming performance prediction? | Identify the research topic/trend of swimming performance prediction. |
| RQ4 | What approach techniques are used in swimming performance prediction? | Identify approaches that are often used in swimming performance prediction |
| RQ5 | What is currently the problem in a swimming performance prediction? | Identify problems that are occurred in swimming performance prediction. |
| RQ6 | What method is used in swimming performance prediction? | Identify the method used in swimming performance prediction. |
| RQ7 | What evaluation techniques are used in swimming performance prediction? | Identify what evaluations are carried out in the swimming performance prediction. |

To understand research questions more easily about summaries of swimming, an illustration has been done in the form of mind map in Figure 2.

Figure 2. Mind Map of Swimming Performance Prediction Review




## *Search strategy*

The data sources used in this review of swimming performance prediction are papers available on ScienceDirect.com, Springer, Taylor & Francis, PubMed and Google Scholar. These webpages are leading journals and conference sites suitable for reviewing swimming performance prediction research. To get a paper that matches the topic, the researchers entered the following keywords or synonyms of the keywords specified for the research topic being conducted. The following are the search strings used for the paper search process: ("swimming performance" or "swimming performances") and ("prediction" or "predicting") and ("athlete" or "swimmer"), and ("modeling" or "approach" or "technique").

Search string adjustments were made to reduce the list of irrelevant studies significantly. A search adjustment is required to meet each specific database requirement at each site. Specific requirements for database searches are based on titles, abstracts, and keywords. Search limited to year of publication: 2014-2024. The publications included are journal articles. The use of article language is limited to text in English.

## *Study selection*

When the paper article search stage was carried out, a large number of articles were filtered according to the criteria when performing the search adjustment process. The determination of the paper criteria included in the main study was obtained from the inclusion and exclusion process described in Table 3.

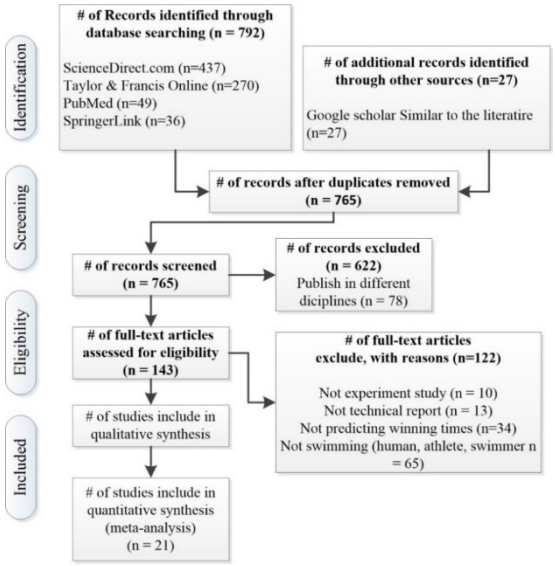
Table 3. Inclusion and Exclusion Criteria

| | |
|---|---|
| Inclusion criteria | Studies on swimming performance prediction include topic, problems, datasets, techniques and methods used.<br>Studies consisted of journals and papers from the conferences that specifically discussed the swimming performance prediction.<br>The publications of the studies were from 2014 to 2024. |
| Exclusion criteria | Studies that did not have experimental results and used unclear datasets.<br>Studies that discussed topics beyond swimming performance prediction.<br>Studies that weren't written in English. |

In addition, to create article boundaries by the research topic that was reviewed, this study used projects for systematic reviews and meta-analysis (PRISMA) guidelines to conduct the systematic review (Moher et al., 2009). The retrieval process, as shown in Figure 3, included an evaluation hierarchy that evaluated: the first, study based on journal titles, the second, based on abstracts, and the third, based on a full-text review in which journal articles were selected according to inclusion and exclusion criteria. The initial search resulted in 792 published journal articles, and after deleting the remaining duplicates, it resulted in 765 papers. A total of 622 papers were then issued because they were published in different disciplines, while another 143 papers were deleted because the full text was unavailable. A total of 143 remaining papers discuss the effect of strength training on adolescent swimming performance. One hundred twenty-two more papers were issued because they were not experimental stud-

ies, technical reports, or predicting winning times and not swimming (human, athlete). As a result, 21 papers were included in this systematic-review.

Figure 3. PRISMA flow chart of the study selection process




### *Data extraction*

The data extraction stage is a process that collects data from the main study to answer research questions. Table 4 below is the data extraction table used.

Table 4. Data Extraction

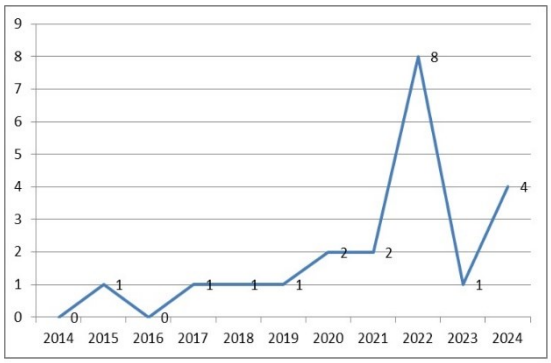
| Property | Research question |
|---|---|
| Publication | RQ1 |
| Swimming performance prediction dataset | RQ2 |
| Research topic or trend | RQ3 |
| Swimming performance prediction technique | RQ4 |
| Swimming performance prediction problem | RQ5 |
| Swimming performance prediction method | RQ6 |
| Swimming performance prediction evaluation | RQ7 |

## Results

### *Paper studies publication*

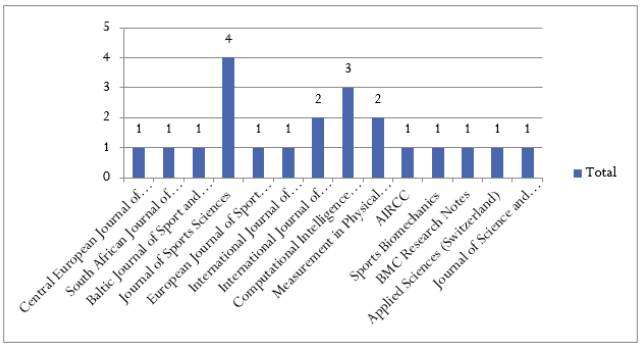
From the results of the screening process conducted as part of this study, it was found that there was a research paper specifically discussing the prediction of swimming performance. In total, a comprehensive review of the literature identified twenty-one papers that addressed various aspects of swimming performance prediction within the time frame from 2014 to 2024. These studies encompass a wide range of methodologies, including biomechanical analysis, physiological data modeling, machine learning techniques, and statistical approaches to assess and enhance swimming performance. To provide a clearer picture of the trend in research interest in this field, Figure 4 presents a graphical representation of the development in the number of published papers over the last ten years. This figure illustrates how the academic community has progressively engaged with the topic of swimming performance prediction, demonstrating fluctuations and potential growth patterns in publication numbers from year to year. From the trend depicted in the graph, it is evident that research on swimming performance prediction remains relevant and continues to attract significant scholarly attention.

Figure 4. Distribution over the past ten years selected studies




Most research on swimming performance prediction is in 2022, with as many as eight publications. This study experienced a significant increase in 2024, with four publications, of which there was only one publication in the previous year, 2023. From the graph, from 2014 to 2024, research on this study was less desirable, namely only 1 or 2 publications. This research began quite a lot in 2022 and continues to this day. In 2024, there were four publications. The publication of journal papers based on literature studies is shown in Figure 5.

Figure 5. Journal Publication and Distribution of Selected Studies

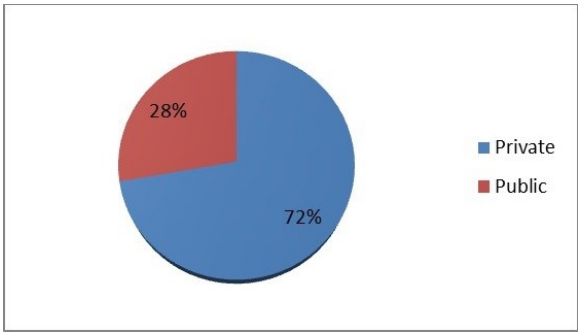



As explained in the background, in this review paper, the researcher took 100% of the journal paper. From the analysis of journal sources that publish publications in swimming performance prediction research, Sports Science and Computational Intelligence are the journal sources that publish the most research topics on swimming performance prediction.

## *Dataset*

In research, dataset is needed to test the performance of the proposed method. In the study of swimming performance prediction, various datasets have been used, which are divided into two groups of datasets: private and public. To see the comparison between private and public datasets that have been used for the last ten years can be seen in Figure 6. Public datasets were used less compared to private datasets. Of the 21 studies selected in the swimming performance prediction research, 16 used private datasets, and five used public datasets. The most favourite public dataset in this study was the men and women in the 100 m and 200 m breaststroke and butterfly stroke from the Olympic (www.olympic.org) dataset with a percentage of 60%. Then Spiideo 10%, web scraping 10%, and swimming events dataset for males and females. From 1990 to 2019, fina (www.fina.org) was 20%.

Figure 6. Distribution of Swimming Performance Prediction Datasets

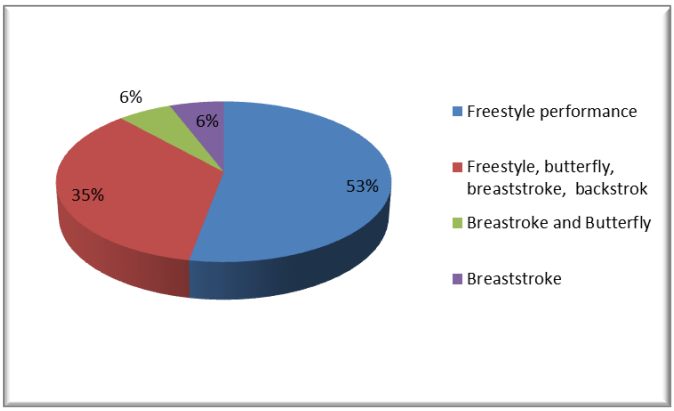



The private data that was widely used is a dataset of 42 male freestyle swimmers from corresponding authors upon request. Other private datasets are Forty-seven swimmers (males, n = 16, age 16.6 ± 1.2 y; females, n = 31, 15.2 ± 1.8 y) from CEFADE 28 2019, 3.061 race results (1.528 Men and 1.533 Women) from OLY and WLC (https://www.gracenote.com/sports/global-sports-data/) competitions between 2011 and 2019.

## *Research topics or trends*

Research on swimming performance prediction covers various research topics or trends. In the last ten years, there have been four research topics or trends in swimming performance prediction: freestyle performance, freestyle, butterfly, breaststroke and backstroke, breastroke and butterfly, and breaststroke. The distribution of research trends or topics in swimming performance prediction is presented in Figure 7. By examining Figure 7, one can observe how specific research topich and trends within swimming performance prediction have gained or diminished in prominence over time.

Figure 7. Distribution of Swimming Performance Prediction Datasets




The most popular research topic or trend in the last ten years is freestyle performance, which has reached 53%. Swimming performance prediction for freestyle performance is the most popular topic because it is more challenging than other stroke types. This is due to its importance in competitive swimming and the complexity of the factors that influence it. Researchers focused on several determinants, including physiological, biomechanical, and anthropometric characteristics. These factors are critical as they affect a swimmer's speed, efficiency and overall performance (Morais et al., 2021). The next favourite research topic after freestyle performance is swimming performance prediction for three types of strokes: freestyle, butterfly, and breaststroke performance. Each type of stroke has unique biomechanical and physiological characteristics that affect swimmers' performance differently. This study aims to understand how these factors interact to maximize speed, efficiency, and overall performance in competition (Espada et al., 2022; Fone & van den Tillaar, 2022; Staunton et al., 2024c). More detailed information about the four topics or trends in the text summary that the researchers have carried out can be found in Table 5.

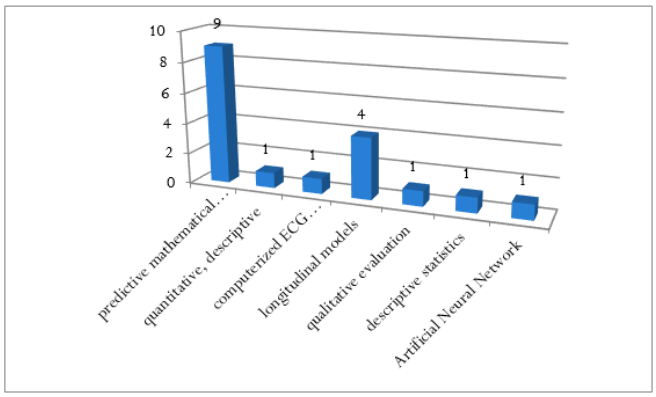
Table 5. Detailed Topics or Trends in Swimming performance Prediction

| Topic/Trends | Research |
|---|---|
| Freestyle performance | (Staunton et al., 2024c), (Carvalho et al., 2024), (Staunton et al., 2024b), (Born et al., 2024)(Crowley et al., 2022), (Morais, Barbosa, Forte, et al., 2023), (Hołub et al., 2021), (de Jesus et al., 2019), (Zuozienė & Poderys, 2018) |
| Freestyle, butterfly, breaststroke, backstroke | (Sridana et al., 2024), (Guo et al., 2022), (Fone & van den Tillaar, 2022), (Post et al., 2022), (Mitchell et al., 2020), (Mabweazara et al., 2017) |
| Breaststroke and Butterfly | (Hołub et al., 2021), (Nicol et al., 2022) |
| Breaststroke | (Abbott et al., 2020), (Staunton et al., 2024c) |

## *Approach Techniques*

From the literature obtained over the last ten years, there are seven approaches or techniques used in swimming performance prediction: predictive-based mathematical model, quantitative-descriptive, computerized ECG registration, longitudinal models, qualitative evaluation, descriptive statistics and artificial neural network. The distribution of the swimming performance prediction approach in the last ten years can be seen in Figure 8.

Figure 8. Distribution of Techniques Applied in Swimming Performance Prediction




The most favourite technique approach used in swimming performance prediction was the predictive mathematical model, which had nine studies. A predictive mathematical model is a favourite technique because it integrates complex biologica health (Lima et al., 2025), biomechanical variables , providing precise performance prediction (Podrihalo et al., 2021). These models help in optimizing training, strategizing competition plans, and understanding performance trends (Mujika et al., 2023). Predictive mathematical models for swimming performance are advanced computational tools designed to forecast an athlete's performance based on various influencing factors (Yuan & Han, 2022a).

However, although the predictive mathematical approach is the favourite, there are better approaches than the predictive mathematical approach. Predictive mathematical models used for swimming performance prediction have several limitations (Mujika et al., 2023) (Dormehl et al., 2017). One significant weakness is their often-limited generalization ability (Demirkan et al., 2023), especially when using linear and non-linear regression models (Imbach et al., 2022). These models are typically easier to interpret but need help capturing the complex, non-linear relationships often present in athletic performance data, leading to potentially less accurate predictions. For instance, the simplicity of these models might need to account for individual variability and the multifaceted nature of swimming performance (Santos et al., 2023), including the influence of biomechanics, training regimens, and physiological factors (Donato et al., 2003).

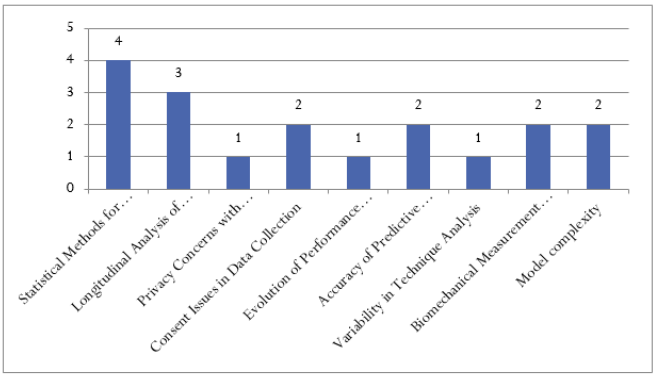
Moreover, while more complex models such as neural networks can offer improved accuracy, they often suffer from the "black box" problem (Sun et al., 2020), where the decision-making process becomes opaque. This lack of interpretability can be problematic for coaches and athletes who need clear guidance from predictive insights. Additionally, the reliability of these models can be compromised when the input data is noisy or incomplete, which is a common issue in sports analytics. These weak-

nesses highlight the need for hybrid approaches that combine mathematical models with more advanced techniques, such as machine learning and explainable artificial intelligence (XAI) (Carvalho et al., 2024), to improve both the accuracy and interpretability of predictions. This lack of interpretability can be problematic for coaches and athletes who need clear guidance from predictive insights. Additionally, the reliability of these models can be compromised when the input data is noisy or incomplete, which is a common issue in sports analytics. These weaknesses highlight the need for hybrid approaches that combine mathematical models with more advanced techniques like machine learning and explainable artificial intelligence (XAI) to improve both the accuracy and interpretability of predictions (Carvalho et al., 2024; Mujika et al., 2023; Zhao et al., 2023).

### *The problem in swimming performance prediction*

Based on swimming performance prediction research from 2012 to 2024, several problems have become challenges that have been trying to overcome. Details of the problems in the text summary over the last ten years can be seen in Figure 9. The most common problem faced in swimming performance prediction over the last ten years is statistical methods for performance prediction. Statistical methods for predicting swimming performance are challenging due to the multifactorial nature of the sport and the complex interactions between various performance-influencing factors. Unlike more straightforward sports, swimming performance is influenced by various variables, including biomechanics, physiology, psychology, and even environmental conditions. Traditional statistical models, such as linear or non-linear regression, often need help with this complexity. These models typically assume a linear relationship between variables, which may not capture the non-linear interactions and dependencies present in swimming (Morais, Barbosa, Gonjo, et al., 2023). For example, the relationship between a swimmer's technique and their performance can vary significantly based on the individual's physical characteristics, training history, and the specific swimming stroke being analyzed. Moreover, the dynamic nature of swimming, where performance can change over time and be affected by numerous contextual factors, makes it difficult to create models that generalize well across different populations of swimmers.

Figure 9. Distribution of Problem in Swimming Performance Prediction




The data used in these models can also be noisy and difficult to interpret, further complicating predictions.

The prediction performance technique often used for statistical problems method for performance prediction in the last ten years is the predictive mathematic model, comparable to the approach technique most often used in the last ten years, the predictive mathematics model (see Figure 8).

### *Methods*

The swimming performance prediction research method used there includes various types of approach techniques. The distribution of methods for this study from 2014 to 2024 is described in Figure 10.

Based on literature studies, linear regression has been the most widely used method in swimming performance prediction for the last ten years. Linear regression is the favorite because it is easy to implement and understand. It provides a precise mathematical relationship between dependent and

independent variables, making it easier to interpret the results. Often, sports performance datasets are relatively small. Linear regression works well with smaller datasets compared to more complex models that may require large amounts of data to train effectively. Linear regression helps identify the most significant variables affecting performance. This can be crucial for coaches and athletes to understand which factors to focus on for improvement. Linear regression serves as a good baseline model. More complex models can be compared against it to justify their use by demonstrating significant performance improvements (Costa et al., 2010; Pyne et al., 2004; Veiga et al., 2013).

Figure 10. Distribution of Methods Used in Swimming Performance Prediction

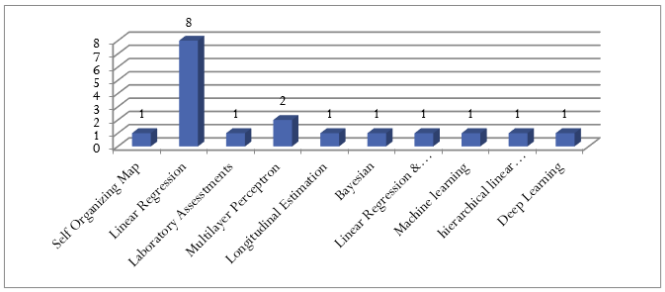



Table 6 provides a comprehensive overview of the critical points identified in 21 studies on swimming performance prediction conducted over the past ten years.

Table 6. Summary of Topic, Problem, Approach Technique, and Method in Swimming Performance Prediction.

| Topics/Trends | Problems | Techniques | Methods | References |
|---|---|---|---|---|
| Freestyle performance | Statistical Methods for Performance Prediction | predictive mathematical model<br>computerized ECG registration<br>longitudinal models<br>descriptive statistics | Self Organizing Map<br>Linear Regression<br>Non linear Regression<br>Bayesian<br>Linear Regerssion<br>hierarchical linear model (HLM)<br>Linear Regression<br>Deep Learning<br>Linear Regression | (Wilk et al., 2015)<br>(Mabweazara et al., 2017)<br>(de Jesus et al., 2019)<br>(Wu et al., 2022)<br>(Mujika et al., 2023)<br>(Morais, Barbosa, Forte, et al., 2023)<br>(Staunton et al., 2024c)<br>(Carvalho et al., 2024)<br>(Born et al., 2024) |
| | Longitudinal Analysis of Training Impact.<br>Consent Issues in Data Collection.<br>Accuracy of Predictive Algorithms | predictive mathematical model | Linear Regression | (Amara et al., 2021) |
| Freestyle, butterfly, breaststroke, backstroke | Model complexity.<br>Longitudinal Analysis of Training Impact | quantitative, descriptive longitudinal models<br>qualitative evaluation<br>predictive mathematical model<br>Artificial Neural Network | Linear Regression<br>Multi-layer perceptron<br>Linear Regression & Non-Linear<br>Machine Learning | (Mabweazara et al., 2017)<br>(Abbott et al., 2020)<br>(Demarie et al., 2022)<br>Learning (Yuan & Han, 2022b) |
| Breaststroke and Butterfly | Biomechanical Measurement Precision | predictive mathematical model | Linear Regression | (Amara et al., 2021) |
| Breaststroke | Variability in Technique Analysis.<br>Privacy Concerns with Performance Data | longitudinal models | Longitudinal Estimation | (Mitchell et al., 2020) |

This table serves as a valuable reference for understanding the various aspects of research in this field. It details the specific research topics explored within these studies, highlighting the key issues and challenges addressed by researchers. Additionally, the table outlines the different approach techniques that have been employed to tackle these problems, offering insights into the methodologies used to enhance the accuracy and effectiveness of swimming performance predictions. Furthermore, the methods associated with these approach techniques are also described, showcasing how researchers have systematically applied various analytical and computational strategies to improve the under-

standing and forecasting of swimming performance. Through this structured presentation, Table 6 provides a clear and informative summary of the advancements made in this domain, helping to identify trends, gaps, and future research opportunities in swimming performance prediction.

### *Evaluation in swimming performance prediction*

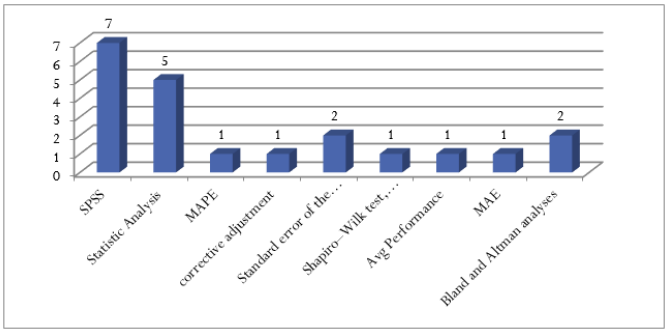
Evaluations in swimming performance prediction encompass various approaches to assess the accuracy and reliability of predictive models. Coaches or experts in swimming may be involved in evaluating the predicted performances. They compare the model's predictions with their knowledge and experience to assess its practical utility (Morais, Barbosa, Gonjo, et al., 2023). Based on studies in the past ten years, there have been various approaches to evaluating the results of swimming performance, namely evaluation in terms of cross-validation (Staunton et al., 2024c).

Evaluation with the statistical metrics is often done by Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) (Hamidi Rad et al., 2021). They are commonly used metrics to measure the difference between the predicted performance and the actual performance. A lower MSE or RMSE indicates better model accuracy (Tomaszewski et al., 2024). In addition, the correlation coefficient is often used to assess the strength and direction of the linear relationship between predicted and actual performances. A high correlation suggests that the model predictions are closely aligned with actual results (Carvalho et al., 2024). Cross-validation evaluation is done by K-Fold Cross-Validation (Liu et al., 2024). This evaluation is used to validate the predictive model by dividing the data into K subsets and using each subset as a validation set while the remaining K-1 subsets are used for training. It helps ensuring that the model generalizes well to unseen data (Liu et al., 2024).

Another evaluation is a task-based approach, which may involve simulating competition scenarios where the predicted performances are compared to actual results in competitive settings. It helps understanding how well the model performs under real-world conditions. Different scenarios, such as changes in training intensity or technique adjustments, are used to evaluate how well the model adapts and predicts performance under varying conditions. Techniques such as Shapley Additive explanations (SHAP) are increasingly used to interpret the results of complex models (Edelmann-Nusser et al., 2002) (Carvalho et al., 2024). These techniques help understand which factors most influence the predicted outcomes, making the evaluation more transparent and trustworthy. These evaluation methods help fine-tune predictive models and ensure their reliability and applicability in real-world swimming performance assessments.

From the studies over the last ten years, the evaluation approach most often used is statistical methods for performance prediction and longitudinal analysis of training impact. In terms of statistical methods for performance prediction, the measure that is often used is SPSS. Meanwhile, in terms of longitudinal analysis of training impact, the measure that is often used is Statistical Analysis. Further distribution details can be seen in Figure.11.

Figure 11. Distribution Statistical Analysis




## Conclusions

Swimming performance prediction is an exciting research topic among the sports science community that helps generate invaluable information in various aspects, especially to adjust the training program to improve specific weaknesses or maximize the potential strengths of a swimmer. It helps designing more effective and specific exercises. Swimming performance prediction allows coaches and

athletes to monitor progress over time. By comparing actual results with predictions, they can evaluate the effectiveness of the training program and make the necessary adjustments. It can also be an essential tool for athlete selection for teams or competitions. By relying on predictive data, coaches can make more objective decisions when selecting athletes with the most potential to deliver the best results.

The main idea of this paper is to present the latest research and progress that has been achieved in this field using the Systematic Literature Review (SLR) method. The SLR method proves that this approach can provide a more structured, broad, and diverse review, ranging from trends/topics, datasets, approach techniques, problems, and methods to evaluations available as a guide for future work. This method also summarizes the relationship between trends/topics, problems, and challenges on each topic, technique, and method used in a single unit, making it easier to explore and re-analyze. This is important to provide in-depth insights and more holistic guidance for researchers and practitioners who wish to develop further research in this area.

The predictive mathematical approach is preferred because it integrates complex physiological, biomechanical, and environmental variables, providing precise performance forecasts. In the research topic of swimming performance prediction, some future work that can be done include I) Combining statistical techniques and techniques based on artificial neural networks. II) Rarely Used Datasets: It includes favourite datasets such as the men and women in the 100 m and 200 m breaststroke and butterfly stroke from the Olympic (www.olympic.org) dataset to test the swimming performance prediction method before being tested on a personal dataset to assess the performance of the method.

## References


Abbott, S., Moulds, K., Salter, J., Romann, M., Edwards, L., & Cobley, S. (2020). Testing the application of corrective adjustment procedures for removal of relative age effects in female youth swimming. *Journal of Sports Sciences*, *38*(10), 1077–1084. https://doi.org/10.1080/02640414.2020.1741956

Amara, S., Chortane, O. G., Negra, Y., Hammami, R., Khalifa, R., Chortane, S. G., & van den Tillaar, R. (2021). Relationship between swimming performance, biomechanical variables and the calculated predicted 1-rm push-up in competitive swimmers. *International Journal of Environmental Research and Public Health*, *18*(21). https://doi.org/10.3390/ijerph182111395

Apriyano, B., Zainuddin, Z. A., Hashim, A. H. M. H., Sayyd, S. M., Mazlan, A. N., Wenando, F. A., Argantos, Ockta, Y., & Anisa, M. F. (2025). Endurance of leg muscle strength and endurance of arm muscle strength to the ability of swimming speed 200 meters breaststroke. *Retos*, *62*, 327–334. https://doi.org/10.47197/retos.v62.109079

Armen, M., Neldi, H., Suud, A., & Alben, C. (2024). *The differences effects of lunges vs. squats exercise programs on the swimming speed 50-meters butterfly style: a quasi-experimental study*. *2041*(2014), 1344–1350.

Banister, E. W., & Calvert, T. W. (1980). Planning for future performance: implications for long term training. *Canadian Journal of Applied Sport Sciences. Journal Canadien Des Sciences Appliquees Au Sport*, *5*(3), 170–176. http://europepmc.org/abstract/MED/6778623

Born, D. P., Stöggl, T., Lorentzen, J., Romann, M., & Björklund, G. (2024). Predicting future stars: Probability and performance corridors for elite swimmers. *Journal of Science and Medicine in Sport*, *27*(2), 113–118. https://doi.org/10.1016/j.jsams.2023.10.017

Busso, T., Denis, C., Bonnefoy, R., Geyssant, A., & Lacour, J. R. (1997). Modeling of adaptations to physical training by using a recursive least squares algorithm. *Journal of Applied Physiology*, *82*(5), 1685–1693. https://doi.org/10.1152/jappl.1997.82.5.1685

Busso, T., Häkkinen, K., Pakarinen, A., Carasso, C., Lacour, J. R., Komi, P. V, & Kauhanen, H. (1990). A systems model of training responses and its relationship to hormonal responses in elite weight-lifters. *European Journal of Applied Physiology and Occupational Physiology*, *61*(1), 48–54. https://doi.org/10.1007/BF00236693

Carvalho, D. D., Goethel, M. F., Silva, A. J., Vilas-Boas, J. P., Pyne, D. B., & Fernandes, R. J. (2024). Swimming Performance Interpreted through Explainable Artificial Intelligence (XAI)—Practical Tests and Training Variables Modelling. *Applied Sciences (Switzerland)*, *14*(12), 1–15. https://doi.org/10.3390/app14125218

Chatard, J. C., & Stewart, A. M. (2011). Training load and performance in swimming. *World Book of Swimming: From Science to Performance*, *January 2011*, 359–373.

Costa, M. J., Marinho, D. A., Reis, V. M., Silva, A. J., Marques, M. C., Bragada, J. A., & Barbosa, T. M. (2010). Tracking the performance of world-ranked swimmers. *Journal of Sports Science and Medicine*, *9*(3), 411–417.

Crowley, E., Ng, K., Mujika, I., & Powell, C. (2022). Speeding up or Slowing Down? Analysis of Race Results in Elite-level Swimming from 2011-2019 to Predict Future Olympic Games Performances. *Measurement in Physical Education and Exercise Science*, *26*(2), 130–140. https://doi.org/10.1080/1091367X.2021.1952592

de Anda Martín, I. O., Pérez, I. M., Mena, E. B., & Del Arco Paniagua, A. (2024). Impact of psychological and physiological factors on endurance performance prediction. *Retos*, *61*, 774–784. https://doi.org/10.47197/retos.v61.106874

de Jesus, K., de Jesus, K., Ayala, H. V. H., dos Santos Coelho, L., Vilas-Boas, J. P., & Fernandes, R. J. P. (2019). Predicting centre of mass horizontal speed in low to severe swimming intensities with linear and non-linear models. *Journal of Sports Sciences*, *37*(13), 1512–1520. https://doi.org/10.1080/02640414.2019.1574949

Demarie, S., Chirico, E., & Galvani, C. (2022). Prediction and Analysis of Tokyo Olympic Games Swimming Results: Impact of the COVID-19 Pandemic on Swimmers' Performance. *International Journal of Environmental Research and Public Health*, *19*(4). https://doi.org/10.3390/ijerph19042110

Demirkan, E., Özkadi, T., Alagöz, I., Çağlar, E. Ç., & Çamiçi, F. (2023). Age-related physical and performance changes in young swimmers: The comparison of predictive models in 50-meter swimming performance. *Baltic Journal of Health and Physical Activity*, *15*(2). https://doi.org/10.29359/BJHPA.15.2.04

Donato, A. J., Tench, K., Glueck, D. H., Seals, D. R., Eskurza, I., & Tanaka, H. (2003). Declines in physiological functional capacity with age: A longitudinal study in peak swimming performance. *Journal of Applied Physiology*, *94*(2), 764–769. https://doi.org/10.1152/japplphysiol.00438.2002

Dormehl, S. J., Robertson, S. J., Barker, A. R., & Williams, C. A. (2017). Confirming the Value of Swimming-Performance Models for Adolescents. *International Journal of Sports Physiology and Performance*, *12*(9), 1177–1185. https://doi.org/10.1123/ijspp.2016-0506

Edelmann-Nusser, J., Hohmann, A., & Henneberg, B. (2002). Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Science*, *2*(2), 1–10. https://doi.org/10.1080/17461390200072201

Espada, M. C., Santos, F. J., Conceição, A., Louro, H., Ferreira, C. C., Reis, J. F., Pessôa-Filho, D. M., & Pereira, A. (2022). The Effects of 12 Weeks In-Water Training in Stroke Kinematics, Dry-Land Power, and Swimming Sprints Performance in Master Swimmers. *Journal of Men's Health*, *18*(9). https://doi.org/10.31083/j.jomh1809186

Fitz-Clarke JR, Morton, R. H., & Banister, E. W. (1991). Optimizing athletic performance by influence curves. *Journal of Thermal Biology*, *71:11*, 51–58. https://doi.org/https://doi.org/10.1152/jappl.1991.71.3.1151

Fone, L., & van den Tillaar, R. (2022). Effect of Different Types of Strength Training on Swimming Performance in Competitive Swimmers: A Systematic Review. *Sports Medicine - Open*, *8*(1). https://doi.org/10.1186/s40798-022-00410-5

Guo, W., Soh, K. G., Zakaria, N. S., Hidayat Baharuldin, M. T., & Gao, Y. (2022). Effect of Resistance Training Methods and Intensity on the Adolescent Swimmer's Performance: A Systematic Review. *Frontiers in Public Health*, *10*(April), 1–10. https://doi.org/10.3389/fpubh.2022.840490

Hamidi Rad, M., Aminian, K., Gremeaux, V., Massé, F., & Dadashi, F. (2021). Swimming Phase-Based Performance Evaluation Using a Single IMU in Main Swimming Techniques. *Frontiers in Bioengineering and Biotechnology*, *9*(December), 1–10. https://doi.org/10.3389/fbioe.2021.793302

Hohmann, A. (1992). Analysis of delayed training effects in the preparation of the West german water polo team for the 1988 Olympic Games. *Biomechanics and Medicine in Swimming. Swimming Science VI*, 213–218. https://open-archive.sport-iat.de/bms/6_213-217_Hohmann.pdf

Hołub, M., Stanula, A., Baron, J., Głyk, W., Rosemann, T., & Knechtle, B. (2021). Predicting breaststroke and butterfly stroke results in swimming based on Olympics history. *Int. J. Environ. Res. Public Health*, *18*(12), 6621. https://doi.org/10.3390/ijerph18126621

Imbach, F., Perrey, S., Chailan, R., Meline, T., & Candau, R. (2022). Training load responses modelling and model generalisation in elite sports. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-05392-8

Kitchenham, B., & Charters, S. M. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report, Ver. 2.3 EBSE Technical Report. EBSE*, *January 2007*.

Lima, L. P., Pontes-Silva, A., Oliveira, I. P. de, Quaresma, F. R. P., & Maciel, E. da S. (2025). Mathematical model for predicting handgrip strength in Quilombola children and adolescents: a cross-sectional study. *Retos*, *62*, 26–30. https://doi.org/10.47197/retos.v62.108681

Liu, C., Xu, B., Wan, K., Sun, Q., Wang, R., Feng, Y., Shao, H., Liu, T., & Wang, R. (2024). Improved prediction of swimming talent through random forest analysis of anthropometric and physiological phenotypes. *Phenomics*, *4*(5), 465–472. https://doi.org/10.1007/s43657-024-00176-8

Lobato, C. H., de Lima Rocha, M., de Almeida-Neto, P. F., & de Araújo Tinôco Cabral, B. G. (2023). Influence of advancing biological maturation in months on muscle power and sport performance in young swimming athletes. *Sport Sciences for Health*, *19*(2), 487–494. https://doi.org/10.1007/s11332-022-01026-8

Mabweazara, S. Z., Leach, L., & Andrews, B. S. (2017). Predicting swimming performance using state anxiety. *South African Journal of Psychology*, *47*(1), 110–120. https://doi.org/10.1177/0081246316645060

Marinho, D. A., Ferreira, M. I., Barbosa, T. M., Vilaça-Alve, J., Costa, M. J., Ferraz, R., & Neiva, H. P. (2020). Energetic and biomechanical contributions for longitudinal performance in master swimmers. *Journal of Functional Morphology and Kinesiology*, *5*(2). https://doi.org/10.3390/jfmk5020037

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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., Gøtzsche, P. C., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7). https://doi.org/10.1371/journal.pmed.1000097

Morais, J. E., Barbosa, T. M., Forte, P., Bragada, J. A., Castro, F. A. d. S., & Marinho, D. A. (2023). Stability analysis and prediction of pacing in elite 1500 m freestyle male swimmers. *Sports Biomechanics*, *22*(11), 1496–1513. https://doi.org/10.1080/14763141.2020.1810749

Morais, J. E., Barbosa, T. M., Forte, P., Silva, A. J., & Marinho, D. A. (2021). Young Swimmers' Anthropometrics, Biomechanics, Energetics, and Efficiency as Underlying Performance Factors: A Systematic Narrative Review. *Frontiers in Physiology*, *12*(September). https://doi.org/10.3389/fphys.2021.691919

Morais, J. E., Barbosa, T. M., Gonjo, T., & Marinho, D. A. (2023). Using Statistical Parametric Mapping as a statistical method for more detailed insights in swimming: a systematic review. *Frontiers in Physiology*, *14*(June), 1–12. https://doi.org/10.3389/fphys.2023.1213151

Morais, J. E., Barbosa, T. M., Neiva, H. P., Marques, M. C., & Marinho, D. A. (2022). Young Swimmers' Classification Based on Performance and Biomechanical Determinants: Determining Similarities Through Cluster Analysis. *Motor Control*, *26*(3), 396–411. https://doi.org/10.1123/mc.2021-0126

Mujika, I., Pyne, D. B., Wu, P. P.-Y., Ng, K., Crowley, E., & Powell, C. (2023). Next-generation models for predicting winning times in elite swimming events: Updated predictions for the Paris 2024 Olympic Games. *Int. J. Sports Physiol. Perform.*, *18*(11), 1269–1274. https://doi.org/10.1123/ijspp.2023-0174

Nicol, E., Pearson, S., Saxby, D., Minahan, C., & Tor, E. (2022). Stroke Kinematics, Temporal Patterns, Neuromuscular Activity, Pacing and Kinetics in Elite Breaststroke Swimming: A Systematic Review. *Sports Medicine - Open*, *8*(1). https://doi.org/10.1186/s40798-022-00467-2

Nugent, F., Comyns, T., Kearney, P., & Warrington, G. (2019). Ultra-Short Race-Pace Training (USRPT) In Swimming: Current Perspectives. *Open Access Journal of Sports Medicine*, *10*(October), 133–144. https://doi.org/10.2147/OAJSM.S180598

Nurmukhanbetova, D., Gussakov, I., & Yermakhanova, A. (2023). The influence of the low-volume high-intensity method training on the indicators of speed and strength qualities of young high skill level swimmers. *Retos*, *50*, 446–455. https://doi.org/10.47197/retos.v50.98492

Okoli, C., & Schabram, K. (2012). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *SSRN Electronic Journal*, *10*(2010). https://doi.org/10.2139/ssrn.1954824

Podrihalo, O., Podrigalo, L., Jagiełło, W., Iermakov, S., & Yermakova, T. (2021). Substantiation of methods for predicting success in artistic swimming. *International Journal of Environmental Research and Public Health*, *18*(16). https://doi.org/10.3390/ijerph18168739

Post, A. K., Koning, R. H., Visscher, C., & Elferink-Gemser, M. T. (2022). The importance of reflection and evaluation processes in daily training sessions for progression toward elite level swimming performance. *Psychology of Sport and Exercise*, *61*, 102219. https://doi.org/https://doi.org/10.1016/j.psychsport.2022.102219

Pyne, D. B., Trewin, C. B., & Hopkins, W. G. (2004). Progression and variability of competitive performance of Olympic swimmers. *Journal of Sports Sciences*, *22*(7), 613–620. https://doi.org/10.1080/02640410310001655822

Rico-González, M., Pino-Ortega, J., Clemente, F. M., & Arcos, A. L. (2022). Guidelines for performing systematic reviews in sports science. *Biology of Sport*, *39*(2), 463–471. https://doi.org/10.5114/BIOLSPORT.2022.106386

Sadewa, Y. R., Sumaryanto, S., Sumarjo, S., & Ika, K. Y. (2024). Relationship of muscle strength, power, and leg flexibility with the swim start of the butterfly style. *Retos*, *55*, 163–169. https://doi.org/10.47197/retos.v55.103106

Santos, C. C., Fernandes, R. J., Marinho, D. A., & Costa, M. J. (2023). From Entry to Finals: Progression and Variability of Swimming Performance at the 2022 FINA World Championships. *Journal of Sports Science and Medicine*, *July*, 417–424. https://doi.org/10.52082/jssm.2023.417

Sridana, R., Tomoliyus, Sukamti, E. R., Prabowo, T. A., & Abrori, R. B. (2024). The Effect of Coaching Style on Performance of Athletes Through Anxiety as Mediating Variable in Adolescent Swimmers. *Retos*, *55*, 241–248. https://doi.org/10.47197/RETOS.V55.103150

Staunton, C. A., Romann, M., Björklund, G., & Born, D.-P. (2024a). Diving into a pool of data: Using principal component analysis to optimize performance prediction in women's short-course swimming. *Journal of Sports Sciences*, *42*(6), 519–526. https://doi.org/10.1080/02640414.2024.2346670

Staunton, C. A., Romann, M., Björklund, G., & Born, D. P. (2024b). Diving into a pool of data: Using principal component analysis to optimize performance prediction in women's short-course swimming. *Journal of Sports Sciences*, *42*(6), 519–526. https://doi.org/10.1080/02640414.2024.2346670

Staunton, C. A., Romann, M., Björklund, G., & Born, D. P. (2024c). Streamlining performance prediction: data-driven KPIs in all swimming strokes. *BMC Research Notes*, *17*(1), 1–7. https://doi.org/10.1186/s13104-024-06714-x

Sun, X., Davis, J., Schulte, O., & Liu, G. (2020). Cracking the Black Box: Distilling Deep Sports Analytics. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 3154–3162. https://doi.org/10.1145/3394486.3403367

Tomaszewski, M., Lukanova-Jakubowska, A., Majorczyk, E., & Dzierżanowski, Ł. (2024). From data to decision: Machine learning determination of aerobic and anaerobic thresholds in athletes. *PLoS ONE*, *19*(8). https://doi.org/10.1371/journal.pone.0309427

Veiga, S., Cala, A., Mallo, J., & Navarro, E. (2013). A new procedure for race analysis in swimming based on individual distance measurements. *Journal of Sports Sciences*, *31*(2), 159–165. https://doi.org/10.1080/02640414.2012.723130

Wilk, R., Fidos-Czuba, O., Rutkowski, Ł., Kozłowski, K., Wiśniewski, P., Maszczyk, A., Stanula, A., & Roczniok, R. (2015). Predicting competitive swimming performance. *Central European Journal of Sport Sciences and Medicine |*, *9*(1), 105–112.

Wu, P. P.-Y., Garufi, L., Drovandi, C., Mengersen, K., Mitchell, L. J. G., Osborne, M. A., & Pyne, D. B. (2022). Bayesian prediction of winning times for elite swimming events. *Journal of Sports Sciences*, *40*(1), 24–31. https://doi.org/10.1080/02640414.2021.1976485

Yuan, R., & Han, Y. (2022a). Factor Analysis and Regression Prediction Model of Swimmers' Performance Structure Based on Mixed Genetic Neural Network. *Computational Intelligence and Neuroscience*, *2022*(1), 2052975. https://doi.org/https://doi.org/10.1155/2022/2052975

Yuan, R., & Han, Y. (2022b). Factor Analysis and Regression Prediction Model of Swimmers' Performance Structure Based on Mixed Genetic Neural Network. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/2052975

Zhao, H., Li, W., Gan, L., & Wang, S. (2023). Designing a prediction model for athlete's sports performance using neural network. *Soft Computing*, *27*(19), 14379–14395. https://doi.org/10.1007/s00500-023-09091-y

Zuozienė, I. J., & Poderys, J. (2018). Laboratory Assessments and Field Tests in Predicting Competitive Performance of Swimmers. *Baltic Journal of Sport and Health Sciences*, *3*(86), 115–119. https://doi.org/10.33607/bjshs.v3i86.276

## Authors' and translators' details:


| | | |
|---|---|---|
| Ari Tri Fitrianto | aritri.fitrianto@uniska-bjm.ac.id | Author |
| Oddie Barnanda Rizky | oddiebarnandarizky@unib.ac.id | Author |
| Edi Rahmadi | 2320129310001@mhs.ulm.ac.id | Author |
| Asary Ramadhan | as'ary.ramadhan@ulm.ac.id | author |
| Resna Suci Nurfalah | resnasucinurfalah@gmail.com | Translator |