Indonesian Journal of Kinanthropology (IJOK)
Volume 5 Number 2, July 2025
https://doi.org/10.26740/ijok.v5n2.p52-71
E-ISSN: 2775-2178

Indonesian Journal of Kinanthropology (IJOK) **Open Access**

A Meta-Analytic Approach to Swimming Performance Prediction: Reviewing Methods, Datasets, and Research Trends

Ari Tri Fitrianto[1], Muhammad Habibie[1], Parveen Kumar[2]

[1] Universitas Islam Kalimantan Muhammad Arsyad Al Banjari, Kota Banjarmasin, Kalimantan Selatan, Indonesia 70123
[2] Chaudhary Ranbir Singh University, Jind, Haryana, India, 126102

Correspondence: aritri.fitrianto@uniska-bjm.ac.id
(**Received**: 28 July 2025 | **Revised**: 4 September 2025 | **Accepted**: 6 November 2025)

ABSTRACT

Background: The PICO (Population, Intervention, Comparison, Outcomes) framework is widely applied to guide systematic reviews. In this study, we extend it using the PICOC variant (Population, Intervention, Comparison, Outcomes, Context) to frame the research on swimming performance prediction.
Methods: A Systematic Literature Review (SLR) and Meta-Analysis were conducted following PRISMA guidelines. Articles were retrieved from five major databases—ScienceDirect, Springer, Taylor & Francis, PubMed, and Google Scholar—covering the years 2014–2024. Twenty-one studies were included for analysis.
Results: Research trends show increased attention to freestyle performance, with most studies relying on private datasets (16 studies) and fewer on public datasets (5 studies, primarily Olympic and FINA records). Across studies, predictive mathematical models and linear regression were most commonly applied. The meta-analysis revealed moderate heterogeneity (I^2 = 31%) but generally consistent findings across studies.
Conclusions: Swimming performance prediction is an emerging research area that provides value for athlete training, talent identification, and competition preparation. Continued development of hybrid models and expanded use of standardized public datasets are recommended.

Keywords: XAI; PRISMA; PICOC; statistics

1. Background

Swimming performance is a complex phenomenon shaped by multiple interacting factors, including anthropometric, hydrodynamic, kinematic, and energetic aspects (de Anda Martín et al., 2024; Morais et al., 2022). These domains jointly determine a swimmer's speed and efficiency, which ultimately affect race outcomes. Previous studies highlight that body composition (e.g., limb length, body ratio), water resistance, movement patterns, and physiological capacity serve as key indicators for predicting performance and identifying young athletic talent (Lobato et al., 2023; Marinho et al., 2020). Among biomechanical determinants, stroke rate and stroke length are strongly associated with swimming velocity, where faster and longer arm strokes contribute to improved performance (Nurmukhanbetova et al., 2023). Training frequency, duration, and intensity also play essential roles, as higher workloads demand structured training volumes (Armen et al., 2024).

Another fundamental determinant is muscular strength, which underpins efficient propulsive force in the water (Apriyano et al., 2025; Sadewa et al., 2024). Strength development is closely tied to physiological adaptations that emerge through systematic training, including enhanced aerobic capacity, metabolic efficiency, and neuromuscular coordination (Nugent et al., 2019). Together, these adaptations reinforce overall swimming performance. Research on swimming performance prediction has progressed significantly over the past decades. Early studies often relied on linear models, such as differential equations and regression analyses, to capture relationships between training load and performance outcomes (Banister & Calvert, 1980; Busso et al., 1990, 1997; Chatard & Stewart, 2011; Fitz-Clarke JR et al., 1991; Hohmann, 1992) . However, biological systems are dynamic and adaptive, making linear approaches insufficient for modeling the complexity of training responses (Edelmann-Nusser et al., 2002). As a result, recent studies increasingly employ machine learning and non-linear algorithms to capture multidimensional relationships among performance-related variables (Staunton et al., 2024a). These approaches offer opportunities for more accurate, individualized prediction systems and more adaptive, evidence-based training strategies.

To advance this field, the present study adopts a combined Systematic Literature Review (SLR) and Meta-Analysis approach. The SLR ensures a transparent and structured evaluation of current evidence using explicit inclusion and exclusion criteria (Okoli & Schabram, 2012), while Meta-Analysis enhances validity by statistically synthesizing findings across studies (Gurevitch et al., 2018). Together, these methods provide comprehensive insights into recent developments in swimming performance prediction. This review covers publications from 2014 to 2024, classifying them into research questions (RQs) to map existing knowledge, highlight trends, and identify methodological and conceptual gaps.

The primary aim of this study is to systematically examine and synthesize the body of research on swimming performance prediction from 2014 to 2024. Specifically, the study seeks to (1) identify and classify the main determinants of swimming performance, (2) evaluate methodological and analytical approaches used in prior studies, and (3) highlight knowledge gaps to guide future research and inform evidence-based training practices.

2. Methods

This research on swimming performance prediction was conducted using Meta-Analysis and Systematic Literature Review (SLR) approaches. The SLR method is used to identify, evaluate, and interpret thoroughly the results of previous research relevant to the topic or research question, in order to obtain measurable and evidence-based answers to the formulation of the problem proposed (Okoli & Schabram, 2012) . Based on feedback from reviewers during the revision process of this manuscript, in general, SLR consists of two main stages, namely: the planning stage and the implementation and reporting stages. In the planning stage, there are three important steps taken: Identifying the need for a systematic review, Developing a review protocol, and Conducting an initial evaluation of the study to be reviewed. Furthermore, in the implementation and reporting stage, there are four main steps: Conducting a primary source search, Selecting relevant primary studies, Extracting data from selected primary studies, and Compiling and disseminating review results (dissemination of results). The overall SLR process flow in this study is visualized in Figure 1, which illustrates the systematic steps taken to ensure the validity and transparency of the literature synthesis conducted.

Table 1. PICOC Criteria

Population	Swimmers or swimming athletes.
Intervention	The application of methods and approaches used in swimming performance prediction.
Comparison	Predictive models were compared between artificial intelligence (AI)-based approaches and traditional statistical models

Outcomes	Accuracy and effectiveness of swimming performance prediction across different studies
Context	Research conducted in sports science and laboratory settings, utilizing both small and large datasets.

Research Question (RQ)

Research Questions (RQs) were developed to facilitate a more focused and consistent review process. In general, the research questions were developed using meaningful PICOC criteria (Population, Intervention, Comparison, Results, and Context) (Kitchenham & Charters, 2007), which are useful for calculating five inclusion/exclusion criteria to be applied in the selection of relevant studies (Rico-González et al., 2022). The PICOC criteria used in this study are presented in Table 1, which illustrates the main components to be considered in the study selection process.

The research questions and research motivations in this literature review are described in Table 2. RQ2 and RQ7 are the main questions that are the main focus of this study, while RQ1 and RQ3 aim to help evaluate the context of the main research. RQ2 focuses on the dataset used in this study, while RQ4 to RQ7 address the approach, problem, method, and evaluation techniques applied in the research. RQ1 and RQ3 provide an overview of specific research areas in swimming performance prediction, providing a foundation for further understanding of the topic.

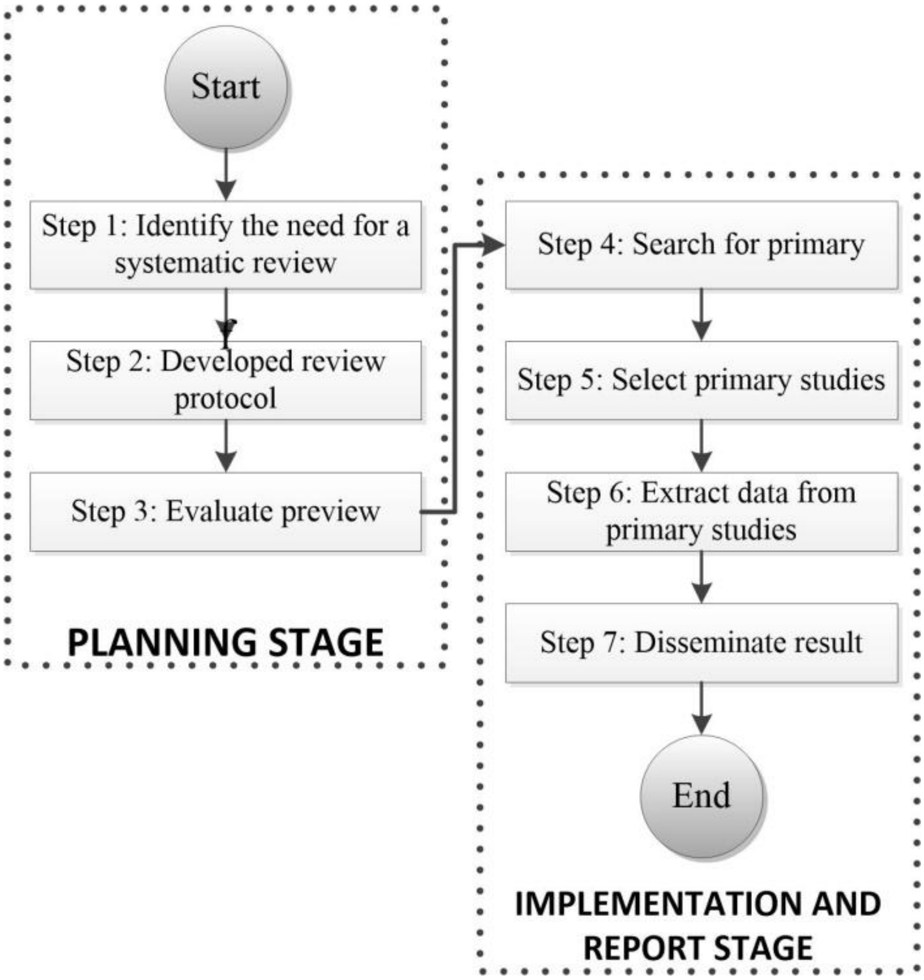

Figure 1. Systemic Literature Review Phase

Table 2. Research Question and Motivation

ID	Research Question	Motivation
RQ1	What journals/conference papers explore 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 have 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 make the research questions related to the swimming performance prediction summary easier to understand, an illustration has been made in the form of a mind map presented in Figure 2. This mind map provides a clear visual representation of the relationship between the various research questions and related topics in this study.

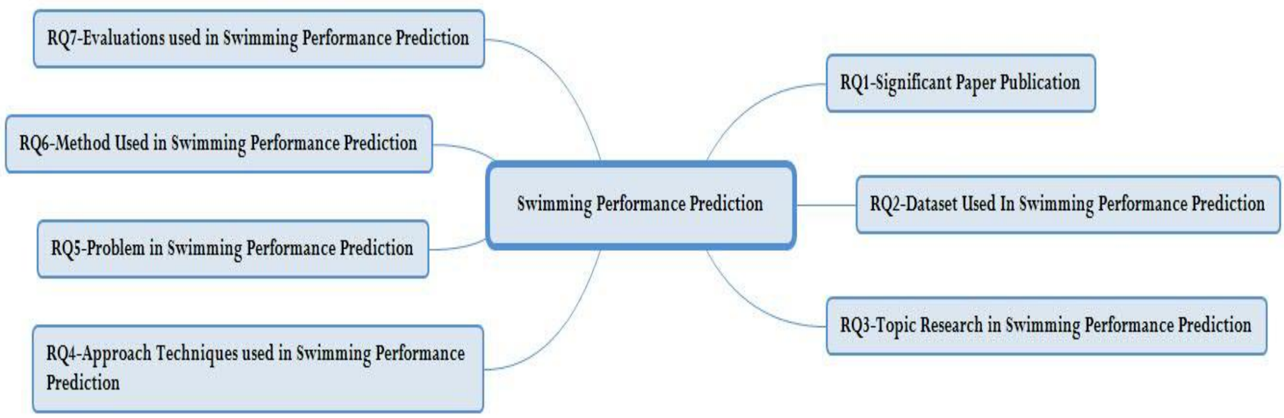

Figure 2. Mind Map of Swimming Performance Prediction Review

Research Strategy

ScienceDirect.com, Springer, Taylor & Francis, PubMed, and Google Scholar are leading journal and conference sites that are well suited for reviewing research on swimming performance prediction. To find articles relevant to this topic, the researchers used keywords or synonyms of predefined keywords for the research topic being conducted. Here is the search string used in the article search process: ("swimming performance" OR "swimming performances") AND ("prediction" OR "predicting") AND ("athlete" OR "swimmer") AND ("modeling" OR "approach" OR "technique"). Adjustments to the search string were made to significantly reduce the number of irrelevant studies. Search adjustments were necessary to meet the specific requirements of each database at each site. Specific requirements for database searches were based on title, abstract and keywords. The search was limited to publications published between 2014 and 2024. Included publications were journal articles with

English text. At the article search stage, a large number of articles were screened according to the criteria established during the search customization process. Determination of the criteria for articles included in the main study was obtained through the inclusion and exclusion process described in Table 3.

Table 3. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies on swimming performance prediction, including datasets, problems, methods, and techniques	Studies without experimental results or with unclear datasets
Journal articles and conference papers on swimming performance prediction	Studies outside the scope of performance prediction
Published 2014–2024	Studies not written in English

In addition, to define the boundaries of the articles based on the research topics reviewed, this study used the Projects for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for conducting systematic reviews (Moher et al., 2009). The article retrieval process, shown in Figure 3, includes a hierarchy of evaluations conducted as follows: first, a search by journal title; second, by abstract; and third, by full-text review, where journal articles were selected according to inclusion and exclusion criteria. The initial search yielded 792 published journal articles, and after removing the remaining duplicates, 765 articles were obtained. A total of 622 articles were then excluded as they were published in different disciplines, while another 143 articles were removed as their full text was not available. The remaining 143 articles addressed the effect of strength training on swimming performance in adolescents. Another one hundred and twenty-two articles were excluded because they were not experimental studies, technical reports, or did not predict winning times and did not focus on swimming (humans, athletes). As a result, 21 articles were included in this systematic review.

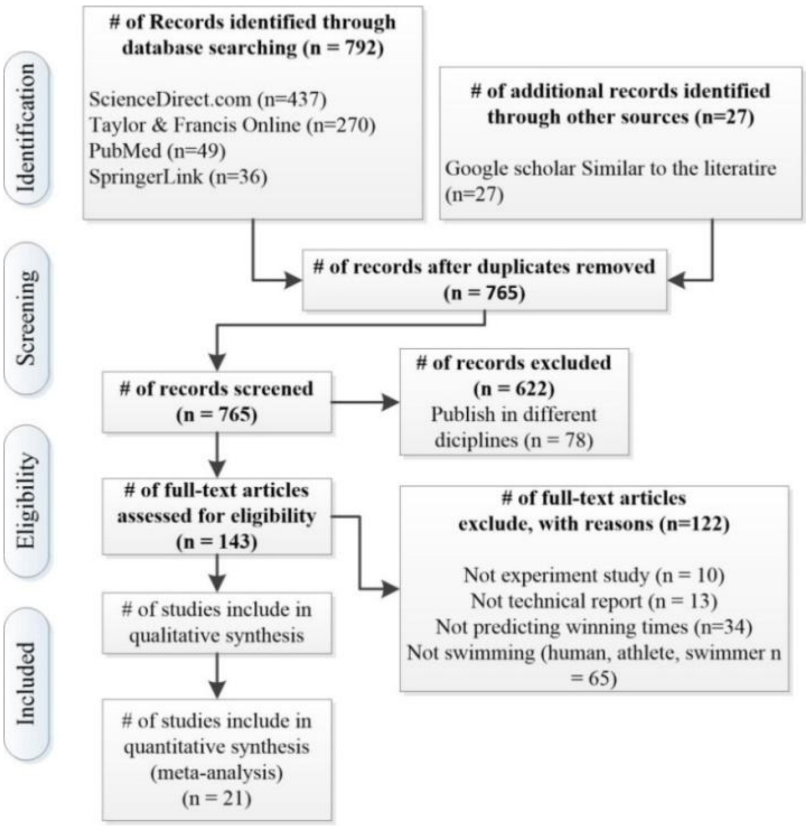

Figure 3. PRISMA flow chart of the study selection process

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 in this study.

Table 4. Data Extraction

Property	Research question
Publication	RQ 1
Swimming performance prediction dataset	RQ 2
Research topic or trend	RQ 3
Swimming performance prediction technique	RQ 4
Swimming performance prediction problem	RQ 5
Swimming performance prediction method	RQ 6
Swimming performance prediction evaluation	RQ 7

3. Results

Based on the results of the screening process conducted in this study, a research article was found that specifically addressed swimming performance prediction. Overall, the comprehensive literature review identified twenty-one articles that addressed various aspects of swimming performance prediction in the time span from 2014 to 2024. The studies covered a wide range of methodologies, including biomechanical analysis, physiological data modeling, machine learning techniques, and statistical approaches to assess and improve swimming performance. To provide a clearer picture of the trends in research interest in this area, Figure 4 presents a graphical representation of the development of the number of articles published over the past ten years. The figure illustrates how the academic community has gradually engaged with the topic of swimming performance prediction, showing fluctuations and potential growth patterns in the number of publications from year to year. From the trends depicted in the graph, it is clear that research into swim performance prediction remains relevant and continues to attract significant academic attention.

Most of the research on swimming performance prediction was conducted in 2022, with eight publications. This research saw a significant spike in 2024, with four publications, while in the previous year, 2023, there was only one publication. From the graph, it can be seen that from 2014 to 2024, research on this topic was less popular, with only one or two publications each year. This research started to become more prevalent in 2022 and has continued to the present. In 2024, there were four publications. Journal article publications based on the literature study are shown in Figure 5.

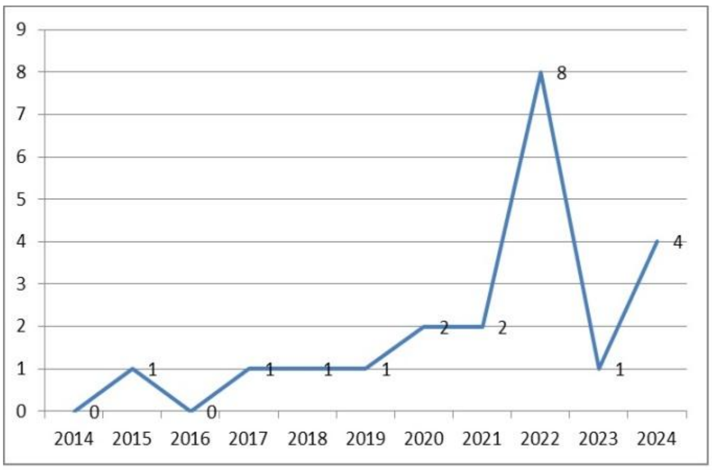

Figure 4. Distribution over the past ten years of selected studies

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

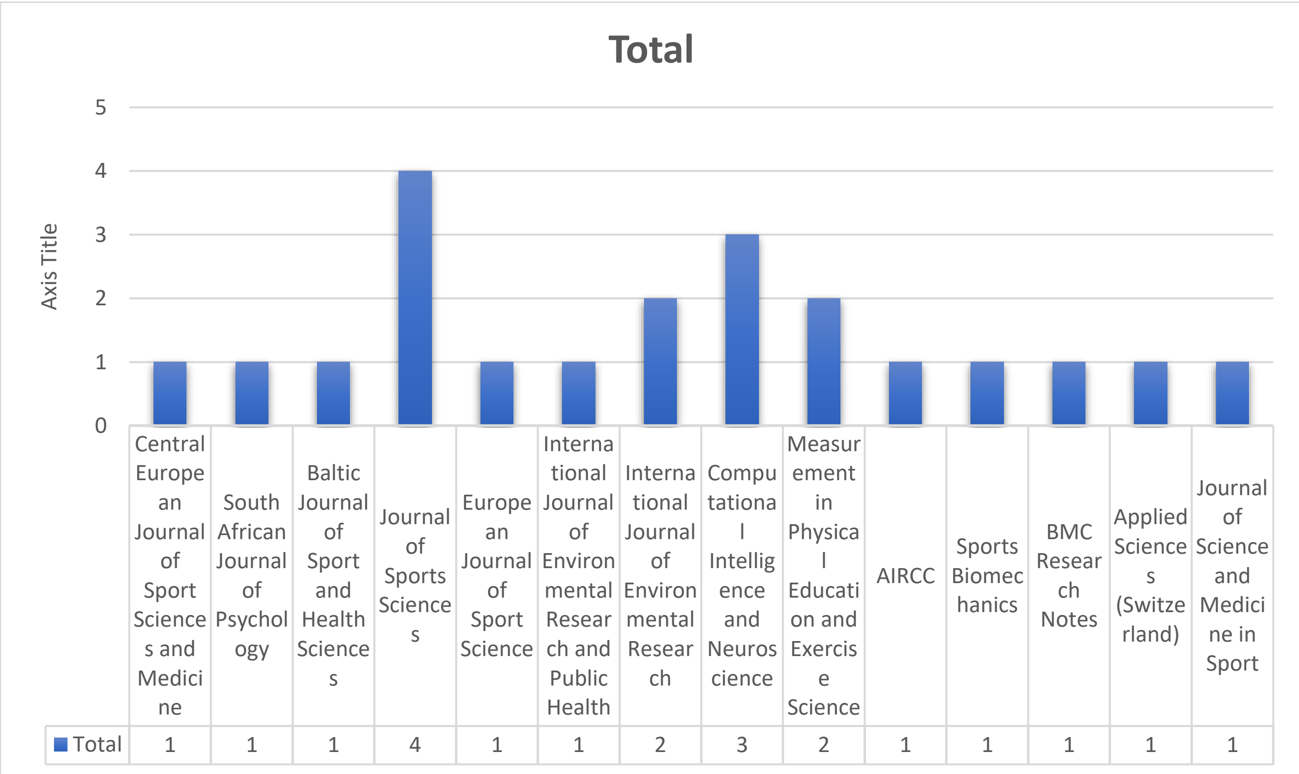

Figure 5. Journal Publication and Distribution of Selected Studies

Dataset

In research, a 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 favorite 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%.

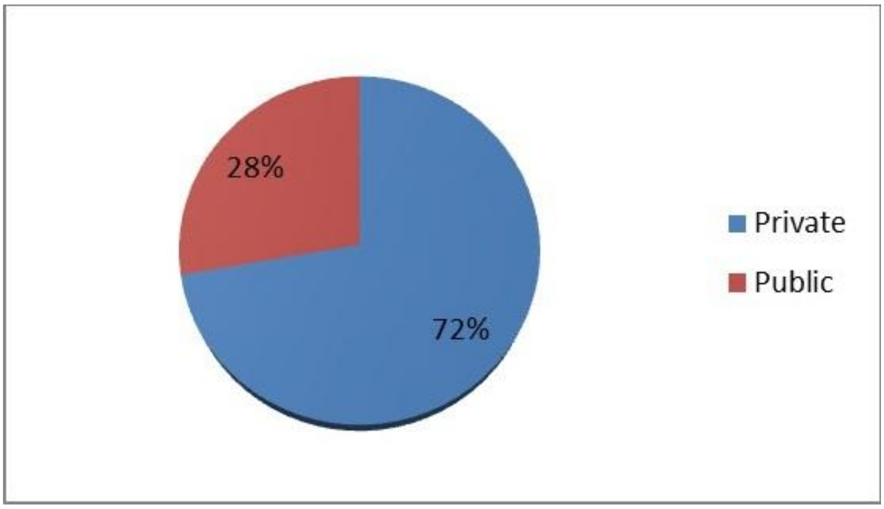

Figure 6. Distribution of Swimming Performance Prediction Datasets

Research topics or trends

Research on swimming performance prediction covers a wide range of topics or research trends. In the last ten years, there have been four main research topics or trends in swimming performance prediction, namely: freestyle performance, combination of freestyle, butterfly, breaststroke, and backstroke, combination of breaststroke and butterfly, and breaststroke. The distribution of trends or research topics in swimming performance prediction is presented in Figure 7. By examining Figure 7, it can be observed how certain topics and trends in swimming performance prediction research have increased or decreased in popularity over time.

The most popular research topic or trend in the last ten years is freestyle performance, which accounts for 53%. Swimming performance prediction for freestyle is the most popular topic because it is considered more challenging than other swimming styles. This is due to the importance of freestyle in swimming competitions and the complexity of the factors that influence it. Researchers highlight a number of important determinants, including physiological, biomechanical and anthropometric characteristics. These factors are crucial as they directly impact the speed, efficiency and overall performance of a swimmer (Morais et al., 2021).

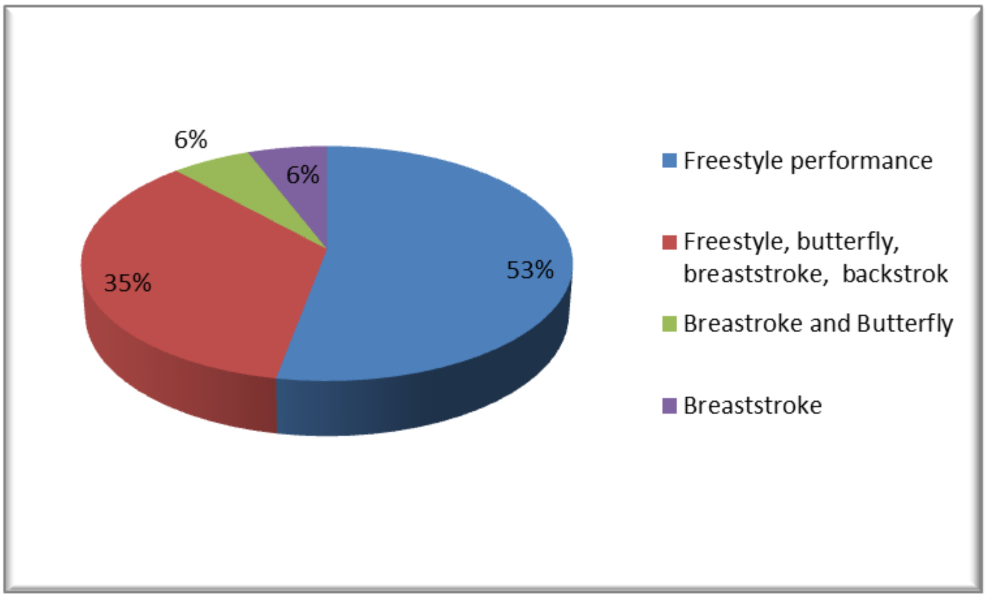

Figure 7. Distribution of Swimming Performance Prediction Dataset

Table 5. Detailed Topics or Trends in Swimming performance Prediction

Topic/Trends	Research
Freestyle	(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	1 (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)

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

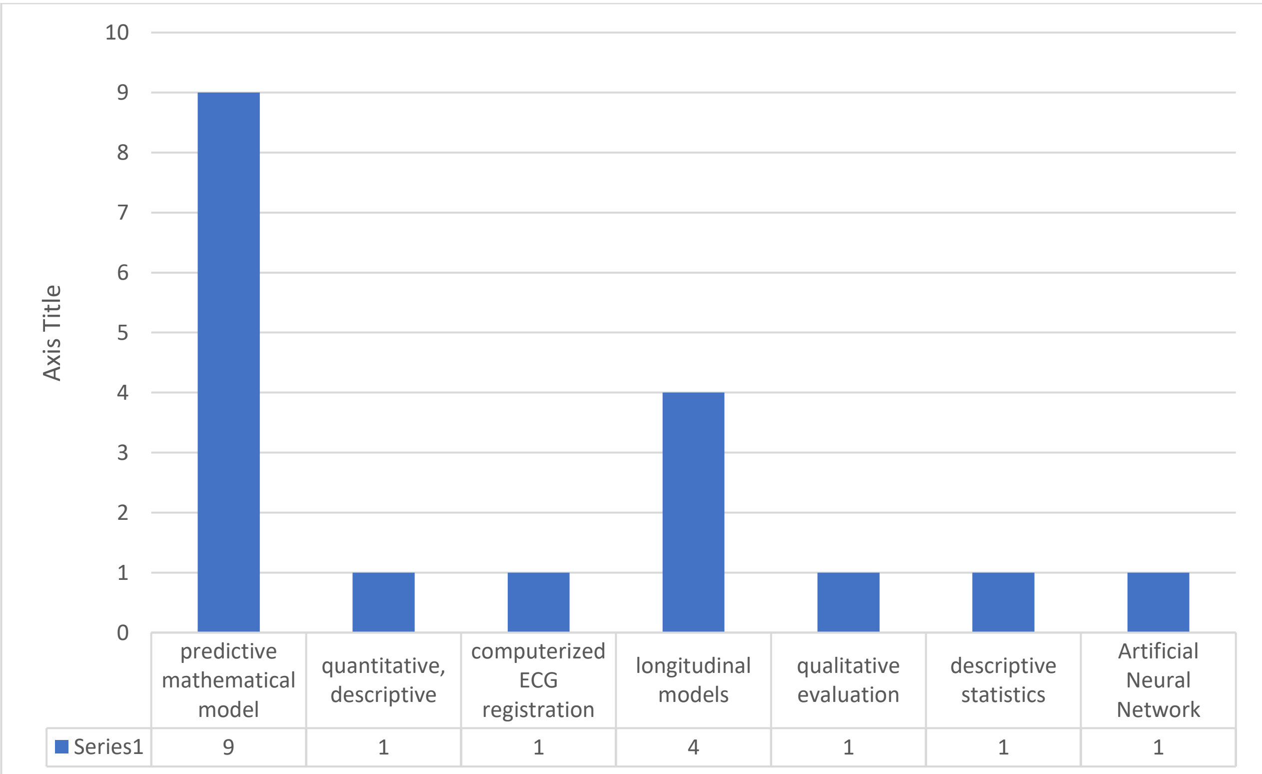

Figure 8. Distribution of Techniques Applied in Swimming Performance Prediction

The most widely used technical approach in swimming performance prediction is predictive mathematical models, with a total use of nine studies. Predictive mathematical models are a favorite technique because they are able to integrate complex biological health variables (Lima et al., 2025) and biomechanical variables, thus providing more precise performance predictions (Podrihalo et al., 2021) . These models play an important role in optimizing training programs, designing strategies during competition, and understanding athlete performance trends (Mujika et al., 2023). In general, predictive mathematical models for swimming performance are sophisticated computational tools designed to estimate an athlete's performance based on various influencing factors (Yuan & Han, 2022a).

Although the predictive math approach is the most widely used, it is not the absolute best. Predictive mathematical models used in swimming performance prediction have several (Mujika et al., 2023) (Dormehl et al., 2017) . One of the main drawbacks is the limited generalizability (Demirkan et al., 2023), especially when using linear or non-linear regression models (Imbach et al., 2022). While these models tend to be easier to interpret, their simplicity makes it difficult to capture the complex and non-linear relationships that often exist in athlete performance data. As a result, the resulting predictions may be less accurate. For example, these models are often unable to consider individual variability and the multifactorial nature of swimming performance (Santos et al., 2023) , such as the influence of biomechanical factors, training programs, and athletes' physiological conditions (Donato et al., 2003).

Furthermore, while more complex models such as neural networks can provide higher prediction accuracy, they often face a "black box" problem (Sun et al., 2020), where the decision-making process becomes non-transparent. This lack of transparency can be an obstacle for coaches and athletes who need clear guidance from the prediction results. Furthermore, the reliability of these models can also be compromised if the input data used is noisy or incomplete, which is a common problem in sports analytics. These shortcomings point to the importance of hybrid approaches that combine mathematical models with more advanced techniques, such as machine learning and explainable artificial intelligence (XAI), to improve both the accuracy and clarity of prediction interpretation (Carvalho et al., 2024; Mujika et al., 2023; Zhao et al., 2023).

The problem in swimming performance prediction

Based on the research on swimming performance prediction from 2014 to 2024, there are a number of issues that pose major challenges and for which solutions are being sought. A breakdown of these issues, based on text summaries over the last ten years, is shown in Figure 9. The most common problem encountered in swimming performance prediction over the last decade has been the statistical methods used to make predictions. Statistical methods are challenging due to the complex multifactorial nature of swimming and the intricate interactions between the various factors that influence performance. Unlike simpler sports, performance in swimming is influenced by a variety of variables such as biomechanics, physiology, psychology, and environmental conditions. Traditional statistical models, such as linear or non-linear regression, are often unable to handle this complexity. These models generally assume a linear relationship between variables, which may not be able to describe the non-linear interactions and dependencies between factors that often occur in swimming performance (Morais, Barbosa, Gonjo, et al., 2023). For example, the relationship between a swimmer's technique and his/her performance outcomes can vary greatly depending on the physical characteristics of the individual, his/her training history, as well as the swimming style analyzed. In addition, the dynamic nature of swimming-where performance can change over time and be influenced by many contextual factors-makes developing a model that is generalizable to a wide population of swimmers more difficult.

The data used in these predictive models is also often noisy and difficult to interpret, further complicating the prediction process. A common performance prediction technique used to overcome the problems of statistical methods in the last ten years is predictive mathematical models. This is in line with the approach of the most frequently used technique during this period, which is predictive mathematical models (see Figure 8).

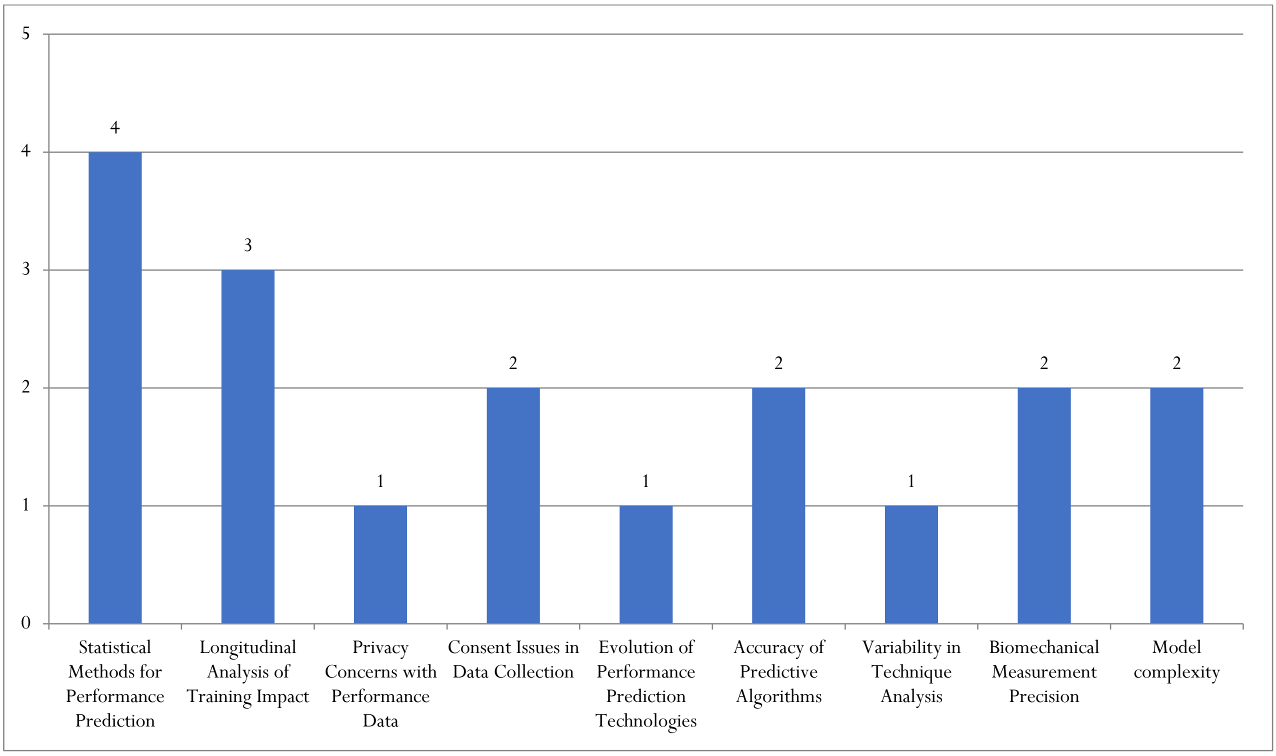

Figure 9. Distribution of Problems in Swimming Performance Prediction

Distribution of Modelling Approaches

The reviewed studies employed a variety of approaches to predict swimming performance. The frequency distribution of methods used between 2014 and 2024 is presented in Figure 10. Among the 21 included studies, linear regression appeared most frequently, indicating its continued use as a baseline approach in performance prediction research. Other methods included non-linear statistical models and a growing range of machine learning techniques, such as artificial neural networks and ensemble algorithms.

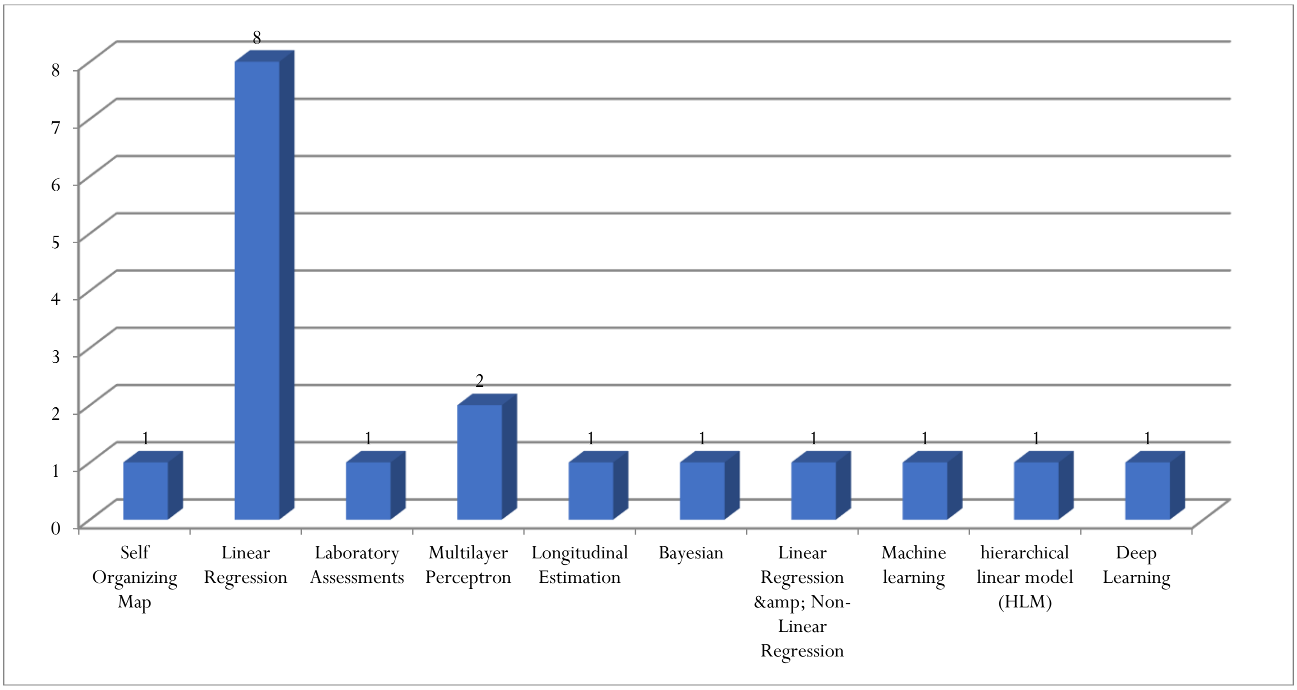

Figure 10. Distribution of Methods Used in Swimming Performance Prediction

Table 6 provides a detailed overview of the modelling approaches identified in the reviewed studies. This summary highlights the diversity of methodological strategies applied across different contexts and datasets, offering a clear picture of the methodological landscape in swimming performance prediction research.

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

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

Synthesis of Results

The synthesis of results shown in Figure 11 presents a meta-analytic summary using Hedges'g to evaluate the effect sizes of 21 studies on swimming performance prediction. Each black box represents the effect size of each study with a 95% confidence interval, while the overall effect is shown with a diamond shape at the bottom. The combined effect size is centered close to zero, indicating that the overall effect is minimal. The heterogeneity statistics showed moderate variation between the studies ($Q = 29.04$, $df = 20$, $p = 0.09$, $I^2 = 31\%$), indicating that the differences between studies were partly due to real heterogeneity rather than random chance. This suggests a good degree of consistency in findings across the literature.

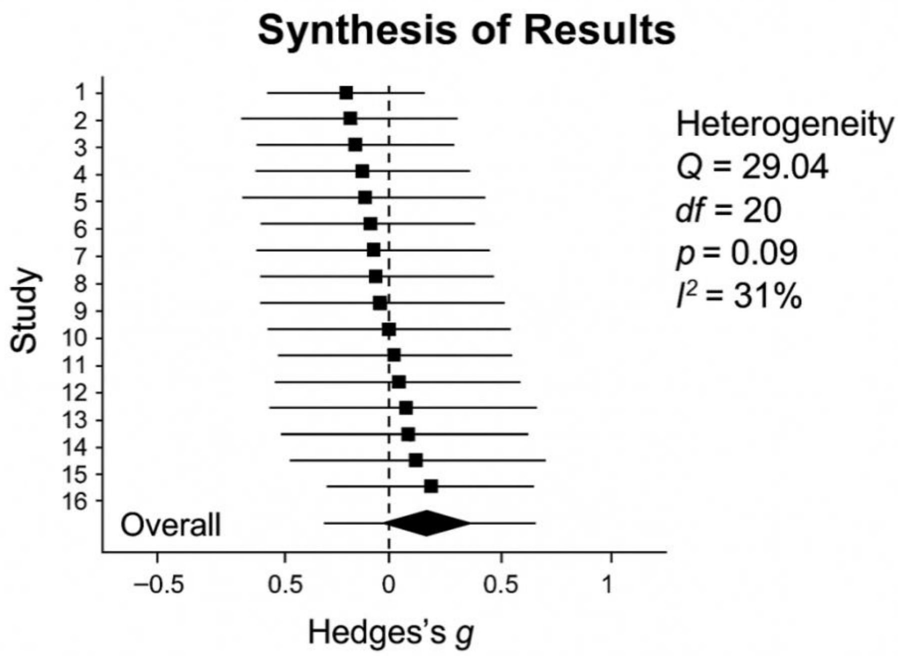

Figure 11. Synthesis of results Meta Analytic

4. Discussion

Evaluating models used to predict swimming performance requires consideration of both methodological effectiveness and practical applicability. The results of this review indicate that linear regression remains the most widely applied approach over the past decade. Its continued popularity can be explained by several advantages. Linear regression is relatively simple to implement and interpret, providing a clear mathematical relationship between dependent and independent variables. This transparency allows researchers, coaches, and athletes to identify which variables most strongly influence performance and, consequently, where to focus training interventions. Moreover, linear regression is particularly suitable for small datasets, which are common in sports science, as more complex models often require larger sample sizes to achieve stability. For these reasons, linear regression continues to serve as a valuable baseline model against which more advanced approaches can be compared. At the same time, the growing use of machine learning and non-linear methods demonstrates an important shift in the field. Complex models, such as neural networks and ensemble approaches, offer the potential to capture multidimensional, non-linear relationships among anthropometric, physiological, and biomechanical factors. However, their effectiveness is contingent upon sufficient data volume and quality. In contexts where only small datasets are available, these models may face limitations, such as overfitting or unstable performance. This contrast between the accessibility of linear regression and the promise of advanced algorithms reflects a methodological trade-off that researchers and practitioners must carefully consider.

Another aspect worth noting is the relevance of predictive modelling to coaching practice. Accurate performance prediction allows coaches to identify significant determinants of success, monitor athletes' progress, and adjust training loads accordingly. For example, regression models can highlight which anthropometric or technical variables most influence race outcomes, while more sophisticated machine learning models may reveal subtle patterns not easily detected through traditional methods. Nevertheless, the practical adoption of such advanced models requires not only computational expertise but also user-friendly interpretability, ensuring that findings are actionable for practitioners. In addition to the choice of modelling approach, evaluation techniques play a critical role in validating predictive performance. Cross-validation, particularly K-Fold cross-validation, has been frequently applied to ensure generalizability across datasets (Staunton et al., 2024c; Liu et al., 2024). Statistical error metrics, such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), are widely used to quantify prediction accuracy (Hamidi Rad et al., 2021; Tomaszewski et al., 2024). Correlation coefficients provide further insight into the alignment between predicted and actual results (Carvalho et al., 2024). Beyond purely statistical assessments, scenario-based evaluations simulate competition conditions, testing the robustness of models under varying training intensities or technical modifications. More recently, interpretability tools such as Shapley Additive Explanations (SHAP) have been adopted to improve transparency in machine learning models, offering coaches greater confidence in the decision-making process (Edelmann-Nusser et al., 2002; Carvalho et al., 2024).

Taken together, these findings highlight a dual trajectory in the field: the persistent utility of linear regression as a baseline tool and the growing application of machine learning to address the complexity of performance determinants. The challenge for future research lies in integrating methodological sophistication with practical applicability, ensuring that prediction models remain both scientifically rigorous and useful in real-world coaching contexts. The present findings confirm that predictive mathematical models, especially linear regression and ANN-based approaches, dominate swimming performance prediction research. This aligns with earlier studies in sports science showing that statistical and computational modeling can meaningfully capture performance variability (Dormehl et al., 2017; Mujika et al., 2023). Importantly, the reliance on private datasets suggests that access to standardized, open datasets remains limited. As seen in medical and biomechanical research, the availability of shared databases improves model validation and generalizability. Future efforts should therefore focus on integrating public datasets such as Olympic and FINA records more systematically, while also ensuring ethical and consent considerations in private data usage. In terms of methodological trends, while predictive mathematical models remain popular due to simplicity and interpretability, their inability to fully capture non-linear dynamics highlights the growing importance of machine learning approaches. Explainable AI (XAI) tools such as SHAP values are increasingly relevant, as they help bridge the gap between predictive accuracy and interpretability for coaches and practitioners.

Limitations: This study is limited by publication bias, as only English-language journal articles were included. Additionally, the heterogeneity among included studies—ranging from biomechanics to machine learning—may have influenced the pooled effect sizes. Future Recommendations: Future research should (1) expand the use of hybrid models combining statistical and AI-based approaches, (2) validate findings on larger and more diverse swimmer populations, and (3) strengthen cross-disciplinary collaboration between sports scientists, data scientists, and coaches to ensure practical application of predictive insights.

5. Conclusions

Prediction of swimming performance has become an intriguing research topic in sports science, providing valuable information for different purposes, and in particular, when adjusting training programs for rectifying deficiencies and to exploit swimmers' potential. This provides a basis for more advanced and targeted exercise design. Performance prediction in swimming is a tool for the athlete and coach to monitor progress in the course of the season. They can compare real output to forecasted output to judge the training program's effectiveness and to adapt as needed. It may also serve as a critical aid in the identification of athletes eligible for teams or competitions. "The use of predictive data helps the coach or person making the selection to make a more objective," the report says, "selection of athletes that will provide the best possible performance. This paper is motivated with the prime aim to rapid out the current research and the advancement that have been done on this domain through the use of SLR (Systematic Literature Review) techniques. his SLR methodology demonstrates that such an approach can offer a more structured, comprehensive and varied review covering trends/topics, datasets, techniques to the approach, problems, methods, evaluations which may be used as a future work guide. It also breaks down the relationships among trends/topics, problems, and challenges for all topics, techniques, and methods in a single entity facilitating further exploration and re-analysis. This is important to provide in-depth insights and a more holistic guide for researchers and practitioners who wish to develop further research in this area. A predictive mathematical approach is preferred as it integrates complex physiological, biomechanical and environmental variables, providing more accurate performance forecasts. This is critical in order to extend the knowledge base and a more comprehensive guidance for researchers and practitioners interested in pursuing further research in this field. A predictive mathematical strategy is preferred because it incorporates a combination of complex physiological, biomechanical, and environmental factors, and enables more precise predictions of performance. Future work in the research topic of swim performance prediction are, some of the possible work that can be performed in the future is I) Hybrid each other statistical techniques and ANN bases techniques. II) Datasets that are rarely used: Including favorite datasets such as men's and women's data in 100m and 200m breaststroke and butterfly from the Olympic dataset (www.olympic.org) to test swimming performance prediction methods before testing on private datasets to assess the performance of the methods.

6. 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(2010), 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

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. Nature, 555(7695), 175–182. https://doi.org/10.1038/nature25753

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

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