JOURNAL LA MULTIAPP

VOL. 07, ISSUE 02 (500-510), 2026
DOI: 10.37899/journallamultiapp.v7i2.2981

Predicting ZNERGY Athletes' Swimming Times Using Random Forest Based on Lifestyle Physiological Exercise

Filemon Blessing Tambuwun[1], Audy Aldrin Kenap[1], Irene Realyta Halldy Trosi Tangkawarow[1]

[1]*Universitas Negeri Manado, Indonesia*

**Corresponding Author: Filemon Blessing Tambuwun*
Email: 22210175@unima.ac.id

Article Info

Article history:
Received 5 January 2026
Received in revised form 20 January 2026
Accepted 7 February 2026

Keywords:
Training Load
Feature Importance
Nutritional Compliance Index
Swimming Performance-Prediction
Random Forest

Abstract

This research was conducted to fill the gap in previous research in swimming performance prediction that rarely combines daily lifestyle aspects with training and physiological data, especially in the context of local swimming clubs in Indonesia that still rely on conventional approaches. The purpose of this research is to develop a prediction model for 50-meter freestyle swimming times using a holistic Random Forest Regressor algorithm. The method applied is a data-oriented methodology for prediction with longitudinal primary data from 10 ZNERGY Aquatic Swimming Club adolescent athletes (81 observations) collected through coach logbooks for training variables and time targets, and digital questionnaires for parents for physiological and lifestyle variables, including a specially developed Nutrition Compliance Index (IKG). The data was processed with Python (Scikit-Learn and Pandas), including pre-processing, lagged feature engineering, and model evaluation with an 80:20 split. The results show the model has high accuracy with an R^2 of 0.9591, MAE of 1.511 seconds, and RMSE of 2.254 seconds; The most important variable was the previous test time (45.66%), followed by training variables (33.67%), while lifestyle variables contributed little. The implication of this study is the availability of an evidence-based decision support system for coaches to design training programs objectively, optimize athlete progress, and prevent overtraining through a holistic approach that integrates multidimensional data.

Introduction

In the modern sporting environment, data has transformed into a fundamental asset, determining the difference between victory and defeat (Ofoghi et al., 2013; Fonti et al., 2023). In a measurable sport like swimming, fractions of a second can determine an athlete's ranking on the podium. Therefore, precision in designing training programs and predicting performance is essential. Mujika et al. (2023); Dann et al. (2024); Salas et al. (2012) highlight that new-generation models for predicting swimming times are now an essential evidence-based framework for coaches to allocate training resources effectively.

However, the reality on the ground shows significant differences. The average swimming club in North Sulawesi, particularly the ZNERGY Aquatic Swimming Club in Kotamobagu, remains heavily tied to conventional approaches. Coaches generally rely on their own knowledge and experience (Mesquita et al., 2010; Carter & Bloom, 2009). While invaluable, these are susceptible to cognitive biases and inconsistent assessments (Haselton et al., 2009).

ISSN: 2716-3865 (Print), 2721-1290 (Online)

The lack of quantitative tools makes it difficult to measure the specific impact of training variables on athlete progress, which risks performance stagnation or overtraining (Grandou et al., 2020; Halson & Jeukendrup, 2004). This situation underscores the urgency of modernizing coaching methods, given that without the support of valid data, the full potential of talented young athletes in this region may never be maximized.

The complexity of human physiology poses a particular challenge to conventional methods. (Bompa & Buzzichelli, 2022), in their recent book on strength periodization, explains that the body's adaptation to training load is complex and non-linear; increasing training volume does not always translate into improved performance without proper recovery management. The application of Machine Learning (ML)-based Sports Analytics offers a solution. This technology enables coaches to discover hidden patterns in complex historical data (Ghosh et al., 2023; Cossich et al., 2023).

Zacca et al. (2024) and Drew & Finch (2016) emphasize that a deep understanding of the relationship between training load and performance data is crucial. Without accurate monitoring, athlete adaptation to training load becomes difficult to predict. Furthermore, various computational algorithms have been tested to solve prediction problems like this in research conducted by: (Liu et al., 2024) the study conducted a comparative study of various ML algorithms and found that Random Forest provided the best performance in identifying swimming talent based on anthropometric and physiological phenotypes.

(Aurélien, 2019) explained that the main advantage of the Random Forest architecture lies in its ensemble learning method, which is able to handle non-linear relationships between variables and prevent overfitting on limited datasets. This capability is very relevant to the dynamic biological characteristics of humans. In addition, (Mehrabi et al., 2025) succeeded in developing a performance prediction model by combining anthropometric data and peak performance age, which statistically proved that the age and innate physical condition of an athlete greatly determine the performance of the athlete.

On the other hand, athlete health and safety are also a major focus in sports analytics (Amendolara et al., 2023) highlighting the use of ML to predict injury risk, while Barry et al. (2022), evaluated training load monitoring practices internationally. They found that predictive algorithms were able to detect injury risk patterns with high accuracy, validating the use of historical data as a preventative tool. However, there is a gap in the existing literature.

These studies tend to emphasize purely technical biomechanical or physiological aspects, often neglecting daily lifestyle variables. However, a recent study by Bardan & Lakhdari, (2025); Capra et al. (2024); AlKasasbeh & Akroush (2024), emphasized that balanced nutrition and adequate rest play essential roles in supporting sports performance and emotional balance in child and adolescent swimmers.

Based on this gap analysis, this study aims to address this gap in the literature by developing a holistic swimming performance prediction model. The novelty of this research is the use of 3 main data dimensions: 1. training program data (volume and intensity), 2. physiological data (demographic and talent), 3. lifestyle data (sleep quality and nutrition). Specifically, this study also introduces the Nutrition Compliance Index (IKG) variable as a quantitative metric to measure the quality of nutritional intake and nutrition of athletes, which has not been widely explored in previous swimming performance prediction model studies due to methodological challenges in converting complex dietary habit data into a standardized numerical format.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

This study aims to address the problem of uncertainty in athlete evaluation by using an accurate Random Forest Regressor model, as well as to find and identify which variables have the most influence (Feature importance) on the 50-meter freestyle swimming time record (due to limited data, the main case here is only the prediction of the 50-meter freestyle). This solution is expected to be a valid and objective Decision Support System for coaches, so that it can replace speculative assumptions with evidence-based recommendations in producing high-achieving athletes in the future.

Methods

This study applies a quantitative research design with a computational experiment approach. In scientific literature, computational experiments are defined as a research method that uses artificial systems to predict and analyze the behavior of complex systems through controlled and repeatable experiments ("Computational Experiment - an overview | ScienceDirect Topics," n.d.). The paradigm used is data-driven predictive modeling, where historical patterns of data are used to project future results. The research procedure is systematically designed to enable replication, starting from data collection from the Trainer, data processing (such as cleaning), feature engineering (creating lagged features), model training, to evaluating model performance and interpreting results. The machine learning method used is Supervised Learning with a regression algorithm, considering that the predicted target variable is continuous data (time in seconds) (Raharjo, n.d.). The entire computational process is carried out using Python as a programming language and using the Scikit-learn Library for modeling and the Pandas library for data manipulation, as recommended by (Abdullah & Kusumastuti, 2025) for efficient implementation of machine learning algorithms. The data source we obtained was primary longitudinal (time-series) data. Athletes were selected using purposive sampling due to limited access to active athletes at the ZNERGY Aquatic Swimming Club, with criteria including regular training participation and willingness to monitor lifestyle. This resulted in a sample of 10 adolescent athletes (8 males, 2 females). Gender imbalances could potentially influence the results due to physiological differences such as muscle strength, maturation rate, and body composition. Focusing on adolescents introduces variability from growth spurts and hormonal/neuromuscular changes, which may confound the relationships between training, lifestyle, and performance variables. This approach also risks selection bias due to the homogeneity of training culture, coaching style, and club environment, thus limiting generalizability to other contexts. The small sample size (81 observations) increases the likelihood that the pattern of results is influenced more by individual variation than systematic effects, which is acknowledged to avoid overgeneralization. Objective monitoring of training variables, such as using standardized measuring instruments, is crucial for understanding the relationship between propulsive force and swimming speed (Lopes et al., 2023).

Data collection was conducted using two main instruments to capture different dimensions of the variables. Stopwatch & Coach Logbook: Used to measure target variables (50m freestyle swimming time) and training variables (mileage/volume, number of sessions, and high-intensity training (sprint). Although data collection and measurement are taken by certified coaches to ensure validity, this approach is still very susceptible to human error such as reaction time bias on a manual stopwatch or lack of focus due to fatigue at the time of taking the time, motivation, or environmental factors.. Digital Questionnaire (Google Form): Used to record lifestyle variables and athlete height information. Although the questionnaire data was input by the athletes' parents, it still has the potential to contain recall bias and social desirability effects. To minimize this bias, we provided clear questionnaire guidelines to parents.The training data collection procedure is collected weekly. At the end of each week,

ISSN: 2716-3865 (Print), 2721-1290 (Online)

training data is summarized from the log book and a time test is also carried out, and the questionnaire is sent to parents to be filled out based on the athlete's activities. The research variables are classified into four categories for holistic analysis: (1) Target Variable (Y): 50-meter freestyle swimming time, measured in seconds using a standard competition stopwatch; (2) Training Variables: Weekly training distance (meters), number of sessions, and high-intensity training ratio (sprints). (Zacca et al., 2024) emphasize that these variables are key determinants of physiological adaptation; (3) Physiological Variables: Age, Sex, Height, and Aptitude Score. (Price, Cimadoro, & S Legg, 2024) state that anthropometry is strongly correlated with performance; (4) Lifestyle Variables: Sleep Quality and Nutritional Compliance Index (GIK). (Romdhani et al., 2021) found that sleep efficiency significantly impacts physical recovery. The GCI variable is calculated using the following composite formula:

$$IKG = \left(\frac{Positive\ Nutrition\ Score}{Maximum\ Positive\ Target}\right) - \left(\frac{Junk\ Food\ Penalty\ Score}{Junk\ Food\ Tolerance\ Limit}\right)$$

This formula was developed as a specific operational definition for this study, with the aim of normalizing questionnaire data from parents into a consistent numeric scale for the random forest algorithm. This formula was adopted from the Diet Quality Index International (DQI-I), which uses a method of assigning points to each component, for example, points are awarded for consuming one food group. By adopting this method, we can determine the score for the Nutrition Compliance Index. The data analysis process was carried out in several stages using the Python programming language:

Data Pre-processing

Data cleaning, handling missing values, converting time formats from minutes:seconds to float, and calculating nutrition and sleep data values from numeric units mixed with sentences into a value that can be processed by the model.

Feature Engineering

Application of Lagged Feature techniques to capture autocorrelation in time-series data. The "Previous Test Time" feature was created by shifting the target data back one period (t–1).

Modeling

Using the Random Forest Regressor algorithm. This algorithm was chosen due to its resilience to overfitting on small datasets and its ability to handle non-linear relationships, as described by (Liu et al., 2024). The hyperparameter configuration was set at n_estimators with a number of 100, max_depth with a number of 5, and min_samples_leaf with a number of 3. Based on empirical testing results (Fitriani et al., 2025), they found that the performance of the Random Forest algorithm was significantly influenced by hyperparameter settings such as n_estimators, which means the number of trees, and max_depth, which means the maximum tree depth. In their experiments, using a larger number of trees (e.g., 100) and limiting the tree depth proved effective in increasing model stability and preventing overfitting.

Model Evaluation

The dataset was divided with a ratio of 80:20 (Train:Test). Performance was evaluated using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2) metrics, referring to the standard evaluation method used by (Susanto et al., 2025). This study adheres to ethical principles for research involving human subjects. Written permission was obtained from the management of the ZNERGY Aquatic Swimming

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Club. Informed consent was also obtained from the athletes' parents/guardians for the use of lifestyle data for academic research purposes only.

Results and Discussion

This study successfully created a predictive model for swimming performance utilizing multidimensional data from ZNERGY Aquatic Swimming Club athletes. The initial modeling process began with examining the characteristics of the dataset to ensure the quality of the training data. Based on descriptive statistical analysis of 81 longitudinal observations, significant performance variation was identified, with the fastest time ranging from 26.71 seconds to the slowest time of 75.04 seconds (with an average of 37.36 seconds). This heterogeneity reflects the dynamics that occur in athlete development at the club level, which encompasses various developmental phases, from beginner to advanced. This variation in data is crucial for machine learning algorithms to learn robust generalization patterns. As explained by (Li & Mu, 2024), the application of feature selection-based algorithms such as Random Forest to complex datasets has proven effective in overcoming accuracy issues typically encountered by traditional statistical methods. The following table presents descriptive statistics from the existing dataset.

Table 1 Descriptive Table of Dataset

Variables	Unit	Minimum	Maximum	Average
Time records	Seconds	26.71	75.04	28.13
Weekly training distance	Meters	6,100 m	31,600 m	16,393 m
Number of training sessions	Session	5	7	5.45
Athlete age	Year	9	16	13.1
Nutritional compliance index	Score (-1 to 1)	-0.38	1	0.31

The model's effectiveness was evaluated through a testing scenario that separated 20% of the total data to be used as test data that was not recognized by the previous model. The coefficient of determination (R^2) value of 0.9591 indicates a very high level of model fit to the observed data. The combination of training, physiological, and lifestyle variables in this dataset. The level of prediction error is also relatively low with a Mean Absolute Error (MAE) of 1.511 seconds and a Root Mean Squared Error (RMSE) of 2.254 seconds which is statistically moderate, but still potentially significant in the context of competitive sprint swimming where small time differences can affect athletes' rankings and achievements.

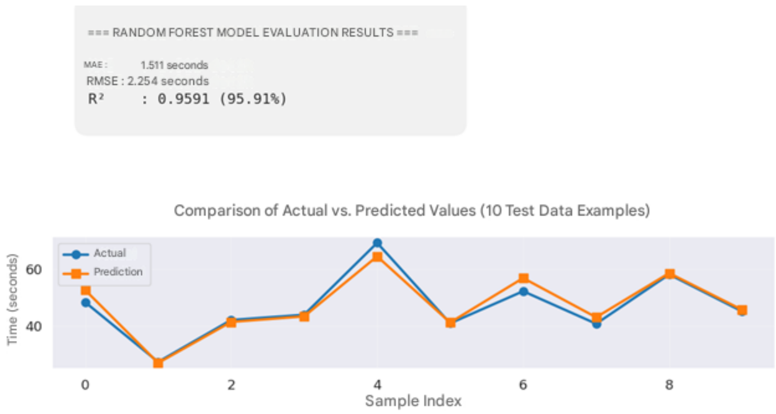

Figure 1. Model Evaluation Results

ISSN: 2716-3865 (Print), 2721-1290 (Online)

However, given the small sample size and the single train-test split (80:20), this high R^2 value indicates a risk of overfitting, where patterns may be dataset-specific rather than broadly predictive. The difference between MAE and RMSE indicates the presence of outliers, but overall, the model provides internally consistent estimates for coaches to monitor weekly progress, although transferability to other contexts has not been tested; further validation, such as cross-validation, is recommended to test for stability.

--- 10 HASIL PREDIKSI VS AKTUAL (DATA UJI) ---

	ID	CATATAN WAKTU TES MINGGU LALU	JARAK LATIHAN	CATATAN WAKTU TES	WAKTU SEBENARNYA (y_test)	PREDIKSI WAKTU (y_pred)	ERROR (detik)
1	1	27.29	20000	27.2	27.2	27.109	-0.0909993
2	1	27.2	22500	27.46	27.46	27.1063	-0.353724
27	4	29.74	31600	29.61	29.61	29.4978	-0.112221
55	7	41.67	11600	40.77	40.77	41.4963	0.726288
37	5	43.47	15000	40.81	40.81	43.0724	2.26244
62	8	40.31	15000	40.96	40.96	41.2257	0.265705
33	5	40.48	15400	42.11	42.11	41.4451	-0.664908
52	7	44.24	11000	43.94	43.94	43.3862	-0.553834
41	6	46.33	11300	45.11	45.11	45.6583	0.548311
69	9	52.18	11000	48.36	48.36	52.6136	4.25361

Figure 2. Prediction Results and Actual Data

Random Forest's strength lies in its ability to identify variable influences using feature importance. The dominance of the "Previous Test Time" variable (45.66%) reflects the presence of temporal autocorrelation in performance data, where previous times are naturally strong predictors of subsequent times. While this characteristic is valid in sports performance data, the dominance of historical variables is more indicative of short-term performance continuity than of causal mechanisms that can be directly modified through training interventions. The relatively high training variable (33.67%) is due to the physiological adaptation overload mechanism that coaches can directly control to increase performance. Physiological variables are relatively low (16.92%) due to the unbalanced sample and lack of female data. While lifestyle variables (GMI and Sleep Score) contribute very little (<1%) because their impact is more directed towards long-term recovery than direct weekly predictions, they are still important to prevent overtraining.

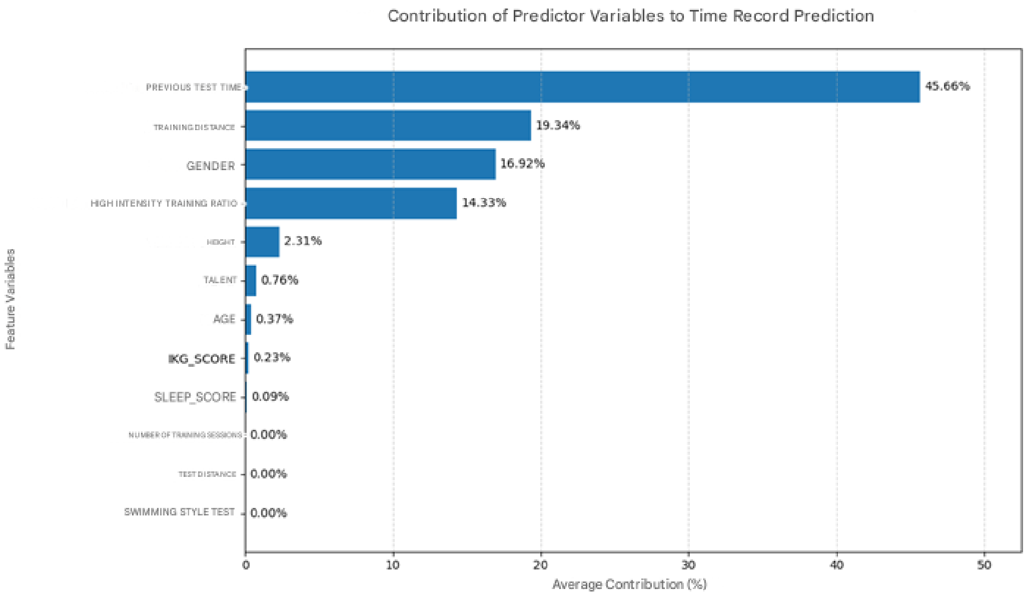

Figure 3. The Most Influential Variables

ISSN: 2716-3865 (Print), 2721-1290 (Online)

To test the model's results as a decision support system, a training scenario simulation was conducted. In the case of a 16-year-old male athlete with a starting time of 32.55 seconds, the model predicted that increasing training volume to 25,000 meters/week at a constant intensity would significantly reduce the time to 28.36 seconds. This predicted speed increase of 4.28 seconds is theoretically supported by (Ruiz-Navarro et al., 2025), who stated that aerobic capacity (developed through high training volume) is the primary foundation for speed endurance in sprinting.

This simulation demonstrates the potential of using the model as an exploratory tool in training planning based on historical data patterns, but it cannot yet be considered a fully reliable prescriptive tool for directly determining training decisions. The high accuracy of the model (95.91%) confirms the superiority of the data-driven approach over intuition alone, as it successfully captures temporal autocorrelation (dominant at "Previous test time" at 45.96%) and the overload adaptation mechanism of training variables (33.67%).

The low contribution of physiological variables due to sample imbalance, and the very low contribution of lifestyle factors due to the less dominant effects of long-term recovery on weekly predictions, effectively explain the nonlinear pattern of performance in adolescent athletes. These findings are consistent with a study (Liu et al., 2024) that reported similar accuracy using Random Forest for swimming talent identification, demonstrating the algorithm's reliability in handling non-linear sports data.

The model's ability to capture these complex patterns provides a solution to the challenge of performance variability, which is often difficult to predict manually. The dominance of the "Previous Test Time" variable in the feature analysis confirms the autocorrelation of sports performance. This aligns with findings (Turfi, 2023), which analyzed Olympic swimming performance and concluded that split time trends consistently exhibit a fairly consistent progressive pattern from year to year, making historical data a very strong predictor.

Furthermore, the significant influence of the training load variable (33.67%) provides empirical evidence for periodization theory, suggesting that volume and intensity manipulation are key mechanisms coaches can control to implement measurable overload phases to drive performance improvements. Although lifestyle factors such as Nutrition Compliance Index and sleep quality have small predictive contributions in the weekly model, their importance should not be overlooked. In-depth research by (Wang et al., 2017; Soltani et al., 2012) shows that demographic factors and daily habits, such as sleep efficiency, significantly impact an individual's physical condition. While nutritional deficiencies or inadequate sleep may not directly impair performance in a single test session, over the long term, they can hinder recovery and increase the likelihood of overtraining. Therefore, integrating this data is crucial for providing a comprehensive understanding of athlete well-being and preventing bias in predictions arising from fatigue.

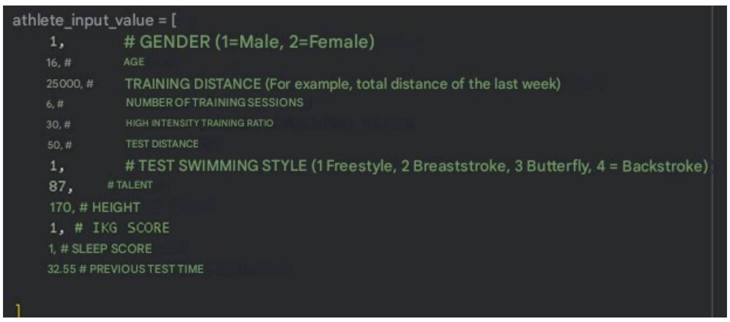

Figure 4. Input Data for Simulation

ISSN: 2716-3865 (Print), 2721-1290 (Online)

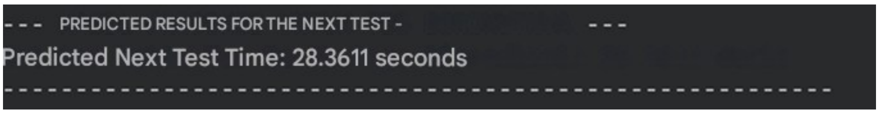

Figure 5. Simulation Prediction Results

Conclusion

This study successfully reduced subjectivity in evaluating swimmer performance by developing a Random Forest Regressor-based predictive model that demonstrated good internal predictive performance on observed data. Empirical testing results show that this model is able to predict ten-meter freestyle swimming times with a Coefficient of Determination (R^2) value reaching 95.91% and an average error (Mean Absolute Error) of only 1.511 seconds. This makes the model a more objective evaluation tool than conventional approaches, given the data and validation scheme used. A summary of the analysis shows that historical performance data and training load metrics (both volume and intensity) are the most important factors contributing to short-term predictions. Meanwhile, lifestyle-related variables, such as diet and sleep quality, although having a smaller direct impact, still play a crucial role in supporting long-term health and preventing bias due to overtraining. The practical implication of this research is the potential development of a data-driven decision support system that can be used exploratively by coaches in training planning, allowing coaches to simulate training scenarios (What-If Analysis) in designing measurable and data-driven performance targets. However, this study also has limitations related to the relatively small sample size and the use of self-reported data on lifestyle variables.

Suggestion

Therefore, future research is recommended to increase the number of participants and utilize wearable sensor technology to obtain more accurate and real-time biomechanical data.

Acknowledgment

The author would like to express his gratitude to God Almighty for the blessings and grace given to the author so that the author could complete this Research article. The author would also like to thank Dr. Audy A. Kenap, ST, M.Sc, as Academic Supervisor, and Dr. Irene R. H. T. Tangkawarow, ST, MISD., as Vice Dean 1 of the Faculty of Engineering, Manado State University and Research Supervisor, for their full support and guidance. The author would also like to thank the author's family and girlfriend who always help, pray, and strengthen the author in every problem the author faces while carrying out this research.

References

Abdullah, R. W., & Kusumastuti, R. (2025). Analisis Pengolahan Ekstraksi Fitur Citra Untuk Klasifikasi Jenis Apel Menggunakan Scikit-Learn Dengan Algoritma K-Nearest Neighbor. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *12*(1), 165-174. https://doi.org/10.25126/jtiik.2025129149

AlKasasbeh, W., & Akroush, S. (2024). Investigating the interrelationships among food habits, sports nutrition knowledge, and perceived barriers to healthy eating: a study of adolescent swimmers. *Frontiers in Nutrition*, *11*, 1381801. https://doi.org/10.3389/fnut.2024.1381801

Amendolara, A., Pfister, D., Settelmayer, M., Shah, M., Wu, V., Donnelly, S., ... & Bills, K. (2023). An overview of machine learning applications in sports injury prediction. *Cureus*, *15*(9).

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Aurélien, G. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. *Concepts, tools, and techniques to build intelligent systems, 2nd ednn*.

Bardan, I., & Lakhdari, E. V. (2025). Relationship between Balanced Nutrition and Sports Performance in Swimmers. *Bulletin of the Transilvania University of Braşov. Series IX: Sciences of Human Kinetics*, 43-50. https://doi.org/10.31926/but.shk.2025.18.67.1.5

Barry, L., Lyons, M., McCreesh, K., Powell, C., & Comyns, T. (2022). International survey of training load monitoring practices in competitive swimming: how, what and why not?. *Physical Therapy in Sport*, *53*, 51-59. https://doi.org/10.1016/j.ptsp.2021.11.005

Bompa, T. O., & Buzzichelli, C. (2021). *Periodization of strength training for sports*. USA: Human Kinetics Publishers.

Capra, M. E., Stanyevic, B., Giudice, A., Monopoli, D., Decarolis, N. M., Esposito, S., & Biasucci, G. (2024). Nutrition for children and adolescents who practice sport: A narrative review. *Nutrients*, *16*(16), 2803. https://doi.org/10.3390/nu16162803

Carter, A. D., & Bloom, G. A. (2009). Coaching knowledge and success: Going beyond athletic experiences. *Journal of sport behavior*, *32*(4), 419.

Cossich, V. R., Carlgren, D., Holash, R. J., & Katz, L. (2023). Technological breakthroughs in sport: Current practice and future potential of artificial intelligence, virtual reality, augmented reality, and modern data visualization in performance analysis. *Applied Sciences*, *13*(23), 12965. https://doi.org/10.3390/app132312965

Dann, R., Duhig, S., Roberts, L., Kelly, V., Renshaw, I., & Headrick, J. (2024). A principled approach to skill acquisition in competitive surfing: Embracing representative learning design. *International Journal of Sports Science & Coaching*, *19*(6), 2534-2547. https://doi.org/10.1177/17479541241279044

Drew, M. K., & Finch, C. F. (2016). The relationship between training load and injury, illness and soreness: a systematic and literature review. *Sports medicine*, *46*(6), 861-883. https://doi.org/10.1007/s40279-015-0459-8

Fitriani, S., Budiman, E., Fadli, M., Surono, M., & Sulistiani, H. (2025, July). Optimalisasi Metode Random Forest menggunakan Particle Swarm Optimization dalam Prediksi Prestasi Mahasiswa. In *Prosiding Seminar Nasional Teknologi Komputer dan Sains* (Vol. 3, No. 1, pp. 406-415).

Fonti, F., Ross, J. M., & Aversa, P. (2023). Using sports data to advance management research: A review and a guide for future studies. *Journal of Management*, *49*(1), 325-362. https://doi.org/10.1177/01492063221117525

Ghosh, I., Ramasamy Ramamurthy, S., Chakma, A., & Roy, N. (2023). Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(5), e1496. https://doi.org/10.1002/widm.1496

Grandou, C., Wallace, L., Impellizzeri, F. M., Allen, N. G., & Coutts, A. J. (2020). Overtraining in resistance exercise: an exploratory systematic review and methodological appraisal of the literature. *Sports Medicine*, *50*(4), 815-828. https://doi.org/10.1007/s40279-019-01242-2

Halson, S. L., & Jeukendrup, A. E. (2004). Does overtraining exist? An analysis of overreaching and overtraining research. *Sports medicine*, *34*(14), 967-981. https://doi.org/10.2165/00007256-200434140-00003

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition*, *27*(5), 733-763. https://doi.org/10.1521/soco.2009.27.5.733

Li, Y., & Mu, Y. (2024). Research and performance analysis of random forest-based feature selection algorithm in sports effectiveness evaluation. *Scientific Reports*, *14*(1), 26275. https://doi.org/10.1038/s41598-024-76706-1

Liu, C., Xu, B., Wan, K., Sun, Q., Wang, R., Feng, Y., ... & 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

Lopes, T. J., Sampaio, T., Oliveira, J. P., Pinto, M. P., Marinho, D. A., & Morais, J. E. (2023). Using wearables to monitor swimmers' propulsive force to get real-time feedback and understand its relationship to swimming velocity. *Applied Sciences*, *13*(6), 4027. https://doi.org/10.3390/app13064027

Mehrabi, A. N., Imani, H., Khademnoe, O., Khantan, M., Lundberg, T. R., & Gorzi, A. (2025). Modelling of performance prediction by analysis of elite swimmers' anthropometry, peak performance age and age-related performance progression. *PLoS One*, *20*(9), e0332306. https://doi.org/10.1371/journal.pone.0332306

Mesquita, I., Isidro, S., & Rosado, A. (2010). Portuguese coaches' perceptions of and preferences for knowledge sources related to their professional background. *Journal of sports science & medicine*, *9*(3), 480.

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. *International journal of sports physiology and performance*, *18*(11), 1269-1274. https://doi.org/10.1123/ijspp.2023-0174

Nash, C., & Sproule, J. (2011). Insights into experiences: Reflections of an expert and novice coach. *International Journal of Sports Science & Coaching*, *6*(1), 149-161. https://doi.org/10.1260/1747-9541.6.1.149

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013). Data mining in elite sports: a review and a framework. *Measurement in Physical Education and Exercise Science*, *17*(3), 171-186. https://doi.org/10.1080/1091367X.2013.805137

Romdhani, M., Hammouda, O., Smari, K., Chaabouni, Y., Mahdouani, K., Driss, T., & Souissi, N. (2021). Total sleep deprivation and recovery sleep affect the diurnal variation of agility performance: the gender differences. *The Journal of Strength & Conditioning Research*, *35*(1), 132-140. https://doi.org/10.1519/JSC.0000000000002614

Ruiz-Navarro, J. J., Santos, C. C., Born, D. P., López-Belmonte, Ó., Cuenca-Fernández, F., Sanders, R. H., & Arellano, R. (2025). Factors relating to sprint swimming performance: a systematic review. *Sports Medicine*, *55*(4), 899-922. https://doi.org/10.1007/s40279-024-02172-4

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological science in the public interest*, *13*(2), 74-101. https://doi.org/10.1177/1529100612436661

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Soltani, M., Haytabakhsh, M. R., Najman, J. M., Williams, G. M., O'Callaghan, M. J., Bor, W., ... & Clavarino, A. (2012). Sleepless nights: the effect of socioeconomic status, physical activity, and lifestyle factors on sleep quality in a large cohort of Australian women. *Archives of women's mental health*, *15*(4), 237-247. https://doi.org/10.1007/s00737-012-0281-3

Susanto, E. R., Inzaghi, M. R., Amarudin, A., & Neneng, N. (2025). Evaluasi Kinerja Model Random Forest Dalam Memprediksi Diabetes Berdasarkan Dataset Kesehatan di Indonesia. *Jurnal Pendidikan dan Teknologi Indonesia*, *5*(7), 1857-1866.

Turfi, A. I. (2023). Analisis Performa Renang Nomor 100meter Gaya Kupu-Kupu Pada Empat Kali Penyelengaraan Olimpiade. *Jurnal Prestasi Olahraga*, *6*(2), 67-74. https://doi.org/10.1234/jpo.v6i2.54917

Van Woezik, R. A., McLaren, C. D., Côté, J., Erickson, K., Law, B., Horning, D. L., ... & Bruner, M. W. (2021). Real versus ideal: Understanding how coaches gain knowledge. *International Sport Coaching Journal*, *9*(2), 189-202. https://doi.org/10.1123/iscj.2021-0010

Wang, S., Li, B., Wu, Y., Ungvari, G. S., Ng, C. H., Fu, Y., ... & Xiang, Y. T. (2017). Relationship of sleep duration with sociodemographic characteristics, lifestyle, mental health, and chronic diseases in a large Chinese adult population. *Journal of Clinical Sleep Medicine*, *13*(3), 377-384. https://doi.org/10.5664/jcsm.6484

Zacca, R., Demarie, S., & Morouço, P. (2024). Training and performance in swimming. *Frontiers in physiology*, *15*, 1402543. https://doi.org/10.3389/fphys.2024.1402543

ISSN: 2716-3865 (Print), 2721-1290 (Online)