NAZARBAYEV UNIVERSITY
SCHOOL OF MEDICINE
MASTER IN SPORTS MEDICINE AND REHABILITATION

YENLIK ULTARAKOVA

MSMR 702

Professor Prashant K Jamwal, Associate professor,
School of Engineering and Digital Sciences

Dr. Yeltay Rakhmanov, Instructor, MD, PhD, School of Medicine

Astana, 2024–2025

MASTER THESIS

AI-BASED PRESCRIPTIVE SYSTEM FOR INDIVIDUALS OPTIMIZED AND NORMALIZED EVALUATION FROM ELITE-REFERENCED BUTTERFLY SWIMMING (AI-PIONEER). PROOF OF THE CONCEPT

Astana, 2024–2025

Abstract

Traditional methods of performance assessment in sports are often subjective, time consuming and require complex expertise to interpret the data accurately. To alleviate current limitations, artificial intelligence, utilizing advanced techniques, presents a promising solution by automating motion analysis, reducing human bias, and improving the accuracy of performance assessment.

This study aims to address the challenge of predicting temporally coherent IMU sequences representing optimal swimming butterfly stroke cycle for amateur athletes, conditioned on static body parameters and kinematic dry swimming parameters extracted from IMU sensors. The objective is to transform a low-dimensional input vector, consisting of 21 static features, into high-resolution temporal output sequences, represented as [T time steps × 48 IMU channels], where each of the 8 sensors records 6 values (three for acceleration and three for angular velocity).

The proposed Physics-Informed Conditional Variational Autoencoder (CVAE) system demonstrates the potential to simulate elite-referenced movement patterns for non-elite athletes, providing a tool for optimizing butterfly swimming technique.

While the ML architecture shows promise in learning from limited data and integrating biomechanical reasoning, the current implementation faces challenges, primarily due to the small dataset size which affects generalizability, the use of dry-land data collection instead of actual swimming, which affect the accuracy and applicability of the model in real swimming scenarios, and the low fidelity of generated sequences, which limits practical use. Thus, this study should be considered a proof of concept rather than a fully deployable solution.

Acknowledgments

I would like to thank my supervisor and co-supervisor, Professor Prashant Jamwal and Dr. Yeltay Rakhmanov, for their unwavering support, guidance, and encouragement throughout my studies at Nazarbayev University and beyond.

I am sincerely thankful to the faculty of the Master of Sports Medicine and Rehabilitation program at Nazarbayev University School of Medicine for providing a stimulating academic environment. Their dedication to teaching and mentoring has equipped me with the knowledge, professional and practical skills necessary to start my journey as a scientists and expert.

I am deeply grateful to my colleagues and fellow researchers Amna Khawaja (MSMR, 2022–2024), Ulan Sharipov (SEDS), Zeeshan (MSMR, 2024–2026) and Aray Zhaisanbek (SEDS) for their insightful discussions, collaboration, and moral support during challenging times.

From the bottom of my heart, I am forever grateful to my family for their unconditional love, sacrifices, and belief in me.

I dedicate this work to all those who strive to advance scientific knowledge of sports science.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AP	Arm Total Propulsion Phase
APB	Arm Propulsion Phase Beginning
APE	Arm Propulsion Phase End
AR	Arm Total Recovery Phase
Total LP	Leg Total Propulsion Phase
TTG	Total Time Gap
dV	Intra-cycle Velocity
COM	Center of Mass
SR	Stroke Rate
SL	Stroke Length
SI	Stroke Index
CFD	Computational Fluid Dynamics
MoCap	Motion Capture
IMU	Inertial Measurement Unit
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
BMI	Body Mass Index
PSA	Predictive Sports Analytics
PCA	Principal Component Analysis
KPIs	Key Performance Indicators
FSS	Free-Swimming Speed
MAE	Mean Absolute Error
CVAE	Conditional Variational Autoencoder
LSTM	Long Short-Term Memory
TCN	Temporal Convolutional Networks
NU IREC	Nazarbayev University Institutional Research Ethics Committee

cm	Centimeter
kg	Kilogram
cm^2	Square Centimeter
Hz	Hertz
m^2/s	Square Meter per Second
m/s	Meter per Second
sec	Second

Chapter 1

LITERATURE REVIEW

1.1 Swimming Theory and Technical Framework

Swimming is a widely practiced and highly competitive Olympic discipline, where the smallest margins of time determine the outcome of the race. In this high-stakes environment, improved and refined technique leads to a better performance outcome. According to Mooney (2016), race is divided into four main components or segments, where the “start” segment begins from the start signal until the head of a swimmer reaches the 15-meter length; "turn" is defined as the sequence of approach, rotation, push-off, glide and underwater phase; the “finish” segment refers to the last 5-10 meters of the race; and the “clean swimming” segment is usually defined as the rest locomotions of the race (Gonjo & Olstad, 2020; Marinho, 2020; Morais, 2019; Puel, 2023). While extensive research has primarily examined the start and turns, studies on the clean swimming have received less attention (Gonjo & Olstad, 2020), underscoring the need for a deeper analysis of the component. As indicated by (Barbosa, 2023), one of the main objectives of sports biomechanics is to characterize motion patterns and optimize their efficiency to enhance athletic performance.

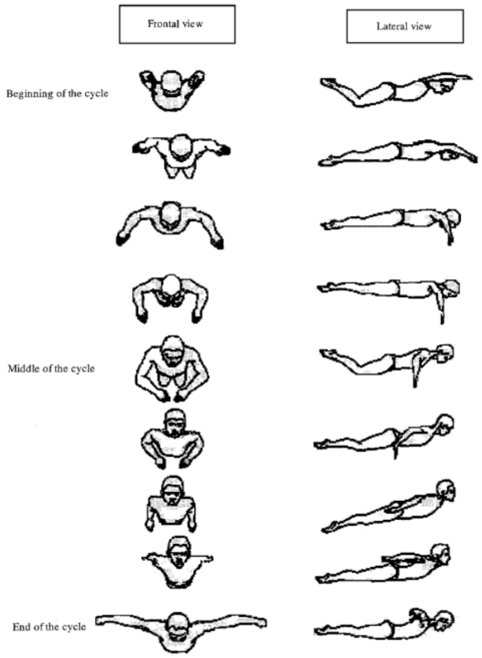
Swimmers face unique challenges as they compete while suspended in an aquatic environment, requiring them to propel their bodies by pushing against liquid rather than solid surfaces as fast as possible (Barbosa, 2010). According to

Maglischo (2003), this creates few significant disadvantages compared to land-based sports. Firstly, water provides less resistance to swimmers’ propulsive efforts, making it harder to generate forward motion. Secondly, due to its higher density, water imposes significantly greater resistance to swimmers’ forward progress than air does for land athletes. These factors highlight the biomechanical complexities of swimming performance and the need for specialized techniques to overcome these challenges. In this regard, an athlete's technical ability, comprising both propelling efficiency and the capacity to overcome drag, plays a crucial role in determining performance.

1.2 Biomechanical Indicators in Butterfly Swimming

Swimming biomechanics is described and easy-to-utilize framework for understanding propulsion and recovery phases. The upper limbs during each of the four competitive swimming strokes are divided into specific phases, defined by the alternated (as in the freestyle and backstroke) or simultaneous (butterfly and breaststroke) distinct sweeps of the arms. Butterfly swimming consists of a well-coordinated sequence of propulsive and non-propulsive phases involving both upper and lower bodily appendages. Strzała (2017) provides a thorough description of butterfly stroke kinematic structure, including the phase generationg propulsion - arm total propulsion (AP), and non-propulsive phase - arm total recovery phase (AR). The propulsive AP begins when the hands initiate an outward movement in pronation as a sequence of coordinated actions of the limbs and trunk, performed in a repeated, synchronous pattern including an identification of the the Arm Propulsion Phase Beginning (APB). This phase consists of three key movements: an outward sweep or insweep, an inward pull or upsweep/backsweep, and an outward push or outsweep/downsweep, which together generate forward thrust (Mason, 1991; Strzała, 2017). The phase concludes when the hands release water, marking the Arm Propulsion Phase End (APE). Following the propulsion phase, the arms start a non-propulsive phase known as the Arm Total Recovery Phase (AR), which begins “between the exit of the hand from the water and its subsequent entry into the water” (Qi, 2023). Since peak hydrodynamic drag occurs during recovery, minimizing resistance during this phase is crucial for efficiency. The Leg Total Propulsion Phase (Total LP) consists of two dolphin kicks per stroke cycle depicted in a 2:1 ratio. The first kick commences immediately preceding the moment of hand entry, aiding in maintaining forward momentum, while the second kick is executed close to the completion of the arm stroke, further enhancing propulsion. Each kick starts when the feet initiate a backward and downward movement and ends when the feet complete their downward motion. In butterfly swimming, propulsion is generated through the synchronized interaction between the upper and lower limbs, with each movement phase contributing to the overall efficiency of the stroke. Effective execution of the propulsion phases maximizes forward acceleration, while minimizing resistance during the recovery phases is critical for sustaining speed and reducing energy expenditure. Inter-limb coordination, specifically the timing between propulsive phases of the upper and lower limbs, is critical to ensure a fast recovery phase and maintain forward momentum, evaluated by the total time gap (TTG) (Barbosa, 2008; Seifert, 2008; Strzała, 2017) (Figure 1).

Figure 1. Different Phases of Butterfly Swimming



Note. Butterfly Hydrodynamics Optimization In Butterfly Swimming: Position, Drag Coefficient And Performance - Scientific Figure On Researchgate. [Accessed 6 Apr 2025]

1.3 Kinematics and Anthropometry in Butterfly Swimming

Kinematic parameters of clean swimming, such as intra- and inter-cycle velocity variations (Barbosa, 2008; Fernandes, 2024), segmental velocities, average velocity, stroke frequency, stroke rate and stroke length (Barbosa, 2010; Gonjo & Olstad, 2020; Morais, 2022), serve as critical indicators for evaluating technique efficiency. The optimization of swimming technique is inherently driven by the attenuating hydrodynamic resistance or drag forces and the increasing propulsion (Morais, 2022; Toussaint, 2011). In such a way, intra-cycle velocity (dV) and acceleration profile allow for an analysis of instantaneous velocity values of a specific point throughout a stroke cycle. Simultaneous swimming strokes exhibit greater variations in dV compared to freestyle and backstroke, consequently, butterfly swimming requires higher energy expenditure, imposing an increased mechanical workload on a swimmer (Barbosa, 2008; Fernandes, 2024). In contrast to other swimming strokes, in butterfly high variations of intra-cycle in average resultant impulse per phase (ARI) are observed due to a steep reduction in arm recovery phase (Barbosa, 2005). During symmetrical hand outsweep, there is an absence of an anchor point. The main objectives in the technical preparation of elite athletes are to achieve high propulsion force and decrease the intra-cycle variation in vertical displacement of the center of mass (COM) (Strzała, 2017). The findings indicated that elite swimmers demonstrate (1) greater arms and leg coordination (Boulesteix & Chollet, 2001), caused by a better streamlined body position during the first lower limbs action (Stosic, 2020; Strzała, 2017); (2) higher stroke rate (SR); (3) better agreement between the upsweep phase of the arm stroke and the second lower limb action augmenting body acceleration (Fernandes, 2024); (4) superior underwater dolphin kick technique, with higher peak angular velocities in the lower trunk, thigh, and shank (Matsuda, 2024). Contrary, less skilled swimmers tend to lengthen the outsweep phase, resulting to lower velocity achieved during upsweep phase, hence, reducing propulsion (Seifert, 2008). Few authors suggest stroke-enhancing strategies that help to improve propulsive forces and drag-reducing factors, such as mastering arm-to-leg coordination with reduced intervals of non-propulsive phase and applying gliding motion (Boulesteix & Chollet, 2001; Strzała, 2017).

Swimming is multifactorial type of physical activity influenced by various factors, such as biomechanics, anthropometrics, hydrodynamic forces, bioenergetics (Seifert, 2023). The study of the influence of body measurements on athletic performance dates to the 19th century, when scientists began systematically measuring human body proportions for various applications. Nowadays, it has become apparent that anthropometrics and somatic parameters contribute to overall performance (Grimston & Hay, 1986; Nevill et al., 2015; Peulić et al., 2023; Pourrahim Ghouroghchi & Pahlevani, 2020). Alves demonstrated that taller swimmers with longer arm spans perform better due to increased key stroke parameters, such as stroke length (SL) and stroke index (SI), whereas Morais (2021) concluded that higher anthropometrics are vital for better performance.

1.4 Performance Analysis in Swimming

Systematic monitoring of changes in swimming performance through the detailed examination of athletes' movement patterns contributes to the identification of technical errors, highlighting areas for improvement (Dalamitros, 2014). Technological breakthroughs in sports science allowed experts and coaches to utilize advanced tools and methods to gain deeper insights into movement analysis by extracting critical parameters and data on athlete performance. Methods and approaches for movement assessment evolved dramatically over the past five decades, transitioning from visual observation, limited by subjective interpretation of the observer, to fully automated real-time feedback technologies. According to Barbosa (2023), research in swimming performance analysis, whether fundamental or applied, is primarily focused on optimizing athlete performance and understanding the principles behind better swimming performance. As a principal tool of the coaching process, applied performance analysis allows for effective and accurate data-driven feedback (Nicholls, 2019) and is designed to augment performance; track progress over time; analyze the strengths and areas for improvement in the athlete and other competitors (Barbosa, 2023). Advanced technologies, as highlighted by Magalhaes (2015), are widely acknowledged as key instruments for performance assessment, enabling detailed analysis of movement patterns from a biomechanical perspective. Video analysis has been a widely recognized and predominant tool in sports research and coaching over the past four decades. Nevalainen (2016) demonstrated how digital camera-based analysis provided immediate feedback by analyzing key stroke metrics such as stroke length, velocity, and acceleration, enabling swimmers to refine their techniques and coaches to adjust training regimens effectively. Recent advancements in automated systems named DeepDASH developed by Hall (2021) have further revolutionized swimming analysis by automating swimmer`s head detection, tracking, and swimming style detection, significantly reducing manual annotation time while improving accuracy. Focusing on resistive hydrodynamic forces, Marinho (2010) successfully carried out numerical simulation techniques by utilizing computational fluid dynamics (CFD), a branch of fluid dynamics, to study the hydrodynamic drag acting on swimmers.

However, swimming is still one of the most challenging sports to quantitatively assess from the standpoint of biomechanics and physics due to the complexity of the dynamic biomechanical system that is constantly traveling through the aquatic and airspace environment (Barbosa, 2010). As a result, utilizing underwater video cameras for recording, modeling, and improving swimming strokes has become a widely adopted method in the practice of high-performance sports. Traditionally, this approach involves attaching tracking markers to specific body segments to estimate the locations of anatomical landmarks and applying algorithms to analyze motion patterns and extract performance-related biomechanical variables. However, some authors and coaches contend that this method has notable limitations, encompassing technical challenges like video quality degradation and data accuracy concerns (Ceseracciu, 2011; Veiga, 2022) and practical implications such as its time-consuming nature (Mooney, 2015). Moreover, dependency on high-resolution cameras required for 3D kinematic description can further increase the logistical and financial demands of the process (Magalhaes, 2015). In recent years, markerless mocap systems have emerged as a solution to overcome the limitations of time-consuming marker-based motion analysis (Lam, 2023). Despite named limitations, video analysis remains a powerful tool that provides a qualitative visual assessment of swimming technique, but alone it can be limited in delivering precise, quantitative data of swimmers' kinematic parameters. Quantitative assessment based on kinematic measures is crucial for providing objective performance feedback. The latest developments in quantitative performance analysis have highlighted wearable technologies as a promising alternative to traditional video-based approaches and a powerful supplementation for video-analysis that helps to overcome existing restraints, particularly in dynamic environments where video systems might face limitations due to water distortions and restricted camera angles. These innovations, as proposed by Mooney (2015), offer the potential to enhance performance monitoring and analysis by providing real-time, precise, and portable data collection capabilities. Building on Daukantas (2008), the initial feasibility studies comparing wearable inertial devices with traditional video-based analysis methods for tracking kinematic parameters demonstrated that wearable systems not only offer a viable alternative to the time-intensive process of video analysis but also exhibit the capability to replicate crucial performance metrics.

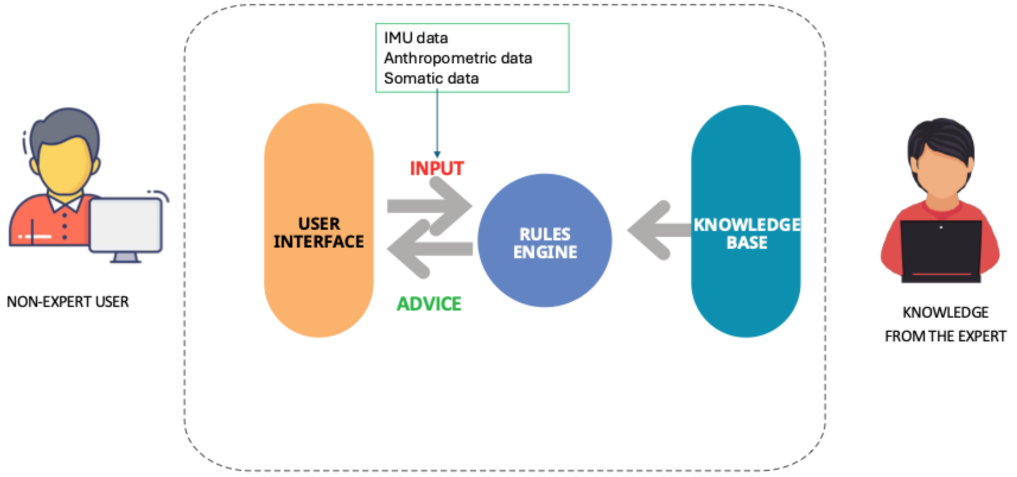
1.5 Wearable Technologies for Quantitative Data Collection

For the past decade, wearable devices have gained widespread acceptance among sports scientists and coaches for assessing athlete performance. Several studies have demonstrated the effectiveness of IMU sensors, technology, which measures the motion and orientation of an object, in swimming performance analysis and daily training settings. Hamidi Rad (2021) used IMUs to monitor the weekly progress of front crawl swimmers to evaluate goal metrics and further developed SmartSwim, IMU-based coaching assistance that facilitates quantitative assessment of swimmers performance. Furthermore, these findings demonstrated that IMUs can provide substantial data on both low-order (velocity, stroke length) and high-order (coordination dynamics) parameters across multiple swimming techniques, as well as inter-cycle velocity variation, a crucial parameter for understanding adaptive behaviors in swimming (Guignard, 2017; Hamidi Rad, Aminian, 2021; Hamidi Rad, Gremeaux, 2021). Bouvet (2024) in recent study has applied IMU sensors and functional clustering techniques for profiling biomechanical abilities in sprint front-crawl swimming and identified specific technical profiles of intra- and inter-cycle stroke regulation, concluding that combining high levels of both forms of variability is correlated with sprint performance. Similarly, Engel's (2021) findings demonstrated how automatic pattern recognition systems provide coaches with real-time feedback on movement execution and which performance parameters can be adjusted. In a recent study of a deep learning model, proposed by Delhaye (2022), designed for swimming analysis using a single IMU attached to the sacrum. Moreover, Fantozzi (2016) introduced an adapted framework for understanding biomechanics, employing IMMU sensors for assessing 3D joint kinematics during simulated swimming. The aim of this experiment was to develop a protocol to evaluate 3D upper limb kinematics using wearable sensors. An analogous framework was adapted by Cortesi (2019) to identify stroke phases in freestyle swimming using the IMU sensors. These findings underscore the potential of inertial devices to enhance the efficiency and precision of swimmer performance analysis.

1.6 AI And ML Applications in Sports

Artificial intelligence (AI) technologies have played a pivotal role in sports science by elevating decision-making strategies (Chen & Dai, 2024), game forecasting, performance enhancement, injury prevention, and motion tracking (Biró 2024; Chmait & Westerbeek, 2021). The earliest attempts to integrate AI in sports biomechanics date back to the 1990s. Lapham and Bartlett (1995) explored the potential of expert systems and neural networks in gait analysis, setting the foundation for further advancements in performance analysis and decision-making (Figure 2). Subsequent review article (Bartlett, 2006) reflected on how AI in sports biomechanics progressed over the decade, focusing on expert systems, artificial neural networks (ANN) and evolutionary computation. Bartlett described how a combination of a knowledge base, computational reasoning (using probability theory or fuzzy logic) and a user interface in the guise of a hypothetical expert system designed to analyze the biomechanical structure of movement by classifying technique. Another given example is how ANN, mimickimg brain neurons and using Kohen self-organizing maps, allow AI to learn from experience and detect patterns in movement, enabling classification, clustering and prediction. Evolutionary computation, an AI technique that uses artificial chromosomes to simulate evolutionary selection, has demonstrated potential in refining athletic performance through movement optimization. Bartlett remained warily optimistic about AI in sports biomechanics. However, the author distinctly acknowledges limitations, stating that expert systems remain underused; ANNs have seen some success but can only be used for testing, not further learning, requiring retraining with new large-scale data for adaptation; and evolutionary computation is still in its early stages. Bartlett suggests several directions for future research in AI applications within sports biomechanics, addressing better data collection, funding, and interdisciplinary collaboration.

Figure 2. Characteristics of Expert system



Note. This diagram illustrates the structure of an expert system designed to support non-expert users in making informed decisions based on biomechanical data. Adapted from javapoint.co

Advanced deep learning neural network architectures like convolutional neural networks (CNN) and graph convolution models enable automatic analysis of athletes' actions, frequencies, and trajectories. One of the well-known examples of an open-source framework called MediaPipe developed by Google helps build complex deep-learning models across various domains, including applications in sports, as it provides pipelines for pose estimation and body movement tracking (Tharatipyakul, 2024; Yu-Hung Hsu, 2024), hence, it is often used in AI-based fitness apps, virtual gym assistants (Dedhia, 2023) and extends to various sports where pose estimation is involved. Artificial intelligence is increasingly being applied to swimming as well. Recent research explores AI applications in swimming data analysis, posture recognition, and performance enhancement (S. Liu, 2021; Yu, 2024; Yu-Hung Hsu & Yu-Hung Hsu, 2024). Cao (2024) developed an intelligent program using ML and motion tracking to analyze swimming starts, which resulted in improved training quality through video processing. Another study, carried out by Sun (2024), employed intelligent sensors and neural networks to recognize swimming strokes with high accuracy. The findings led to the development of a teaching system for college swimming courses. An experienced AI technology, by simulating the judgment or behavior of an expert human, can be developed into an expert system that is designed for resolving complex problems in any particular domain (*AI and Expert Systems: All You Need to Know*, 2024). The system heavily relies on human expertise and is characterized as a set of reasoning rules. By simulating various stroke modifications, this model may determine the most effective adjustments for each swimmer, leading to a personalized, data-driven assessment method that potentially can maximize performance outcomes. Moreover, this approach in athletic training allows sports analysts to develop personalized training programs (Derie, 2020) adapted to each athlete’s unique biomechanics. Additionally, AI can play a crucial role in injury prevention by analyzing movement patterns and predicting risks. By modifying stroke technique based on AI predictions, athletes can avoid overuse injuries (Wanivenhaus, 2012). As AI continues to evolve in sports science, it promises to unlock new dimensions of performance optimization and personalized training strategies (Bodemer, 2023).

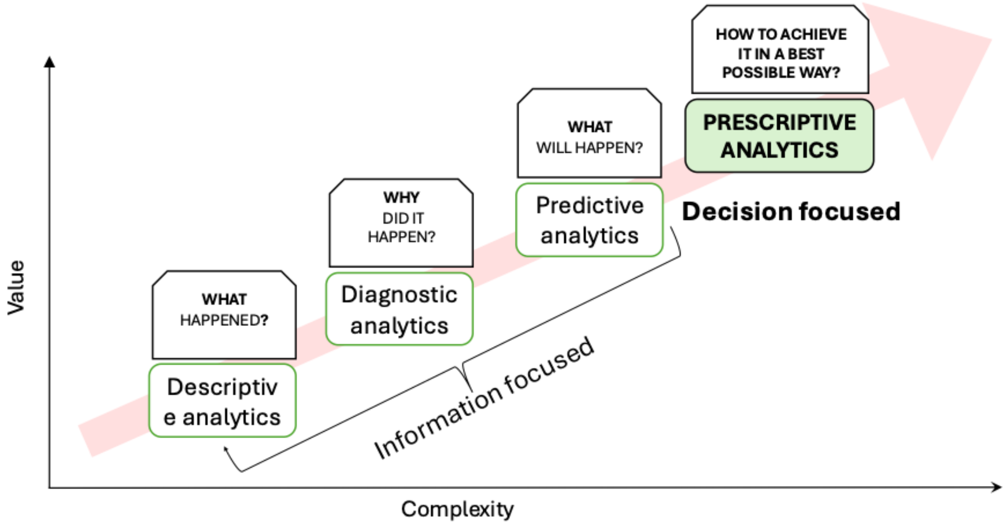
1.7 Data Analytics In Sports

By employing advanced computational methods, data analytics converts raw information into actionable insights. Sports practitioners heavily rely on data to extract objective quantitative information from athletes to streamline informed decision-making, optimize training plans, monitor athletes' health and prevent injuries. For the past decades, the increasing integration of technology, such as wearable sensors, motion capture systems, and machine learning algorithms, has revolutionized the way data is collected, analyzed, and utilized in sports. Evaluating an athlete’s fitness level provides essential insights into their physical capabilities, recovery needs, and overall adaptability to training loads (Kamarudin, 2022). Quantitative assessment methods have become an integral part of sports science and are widely utilized across various domains, including performance analysis, physiology for monitoring athlete health, talent identification and sports management. Within sports analytics, evaluating physical fitness level provides insight into an athlete's physical condition and recovery needs and is essential for talent identification and adjustment of training plans (Kamarudin, 2022). According to Hughes and Bartlett (2015), physical fitness consists of health-related and skill-related fitness components. While health-related fitness components serve as the foundation for assessing overall well-being, such as age, weight, body mass index (BMI), cardiovascular endurance, strength, body fat percentage (Cabarkapa, 2022), etc., skill-related fitness components pertain to the specific skills necessary for athletic performance. For example, in performance analysis, specialists focus on performance-related variables such as speed, power output, reaction time and many others that provide coaches with an overview of the athlete's current level of technical and tactical preparation (Exel, 2024). The indisputable benefits of monitoring an athlete's external and internal training loads deepen the understanding of the physiological mechanisms underlying their adaptability (Soligard, 2016).

1.8 Conceptual Data Analytics Framework in Performance Analysis

In light of the importance of a data-driven approach in sports, a structured theoretical foundation is essential. Robust conceptual data analysis frameworks are widely used in a broad spectrum of disciplines, defining key concepts and relationships, helping to structure complex information and facilitating multidisciplinary collaboration. Houtmeyers (2021) proposed a conceptual framework for elite sports based on a taxonomy used in business analytics literature, tailoring it to decision-making in elite sports. The author clearly emphasized the significance of data quality and integrity in sports analytics before implementing any type of data analysis, highlighting that dependable decision-making relies on accurate, valid, and well-structured data. Comprising four hierarchical stages—descriptive, diagnostic, predictive, and prescriptive analytics—each stage progressively increases in complexity and analytical depth, as depicted in Figure 3.

Figure 3 Conceptual Data Analytics Framework for Elite Sport



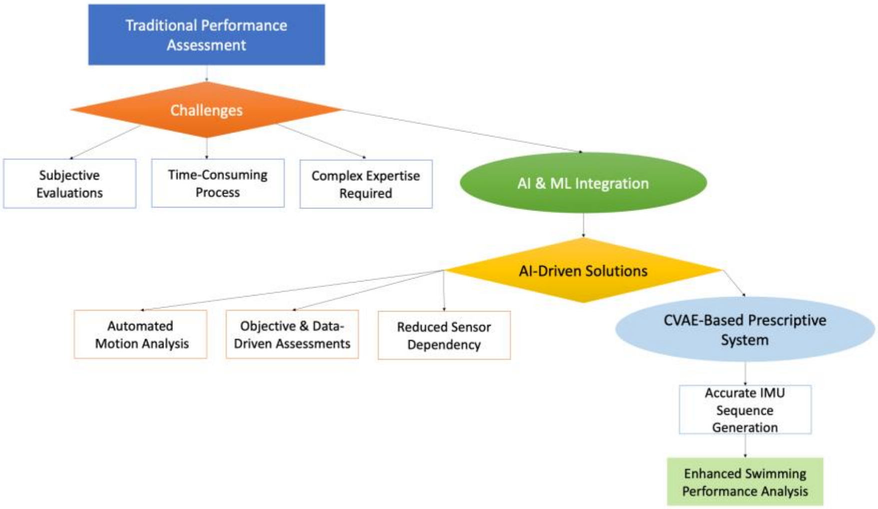
Note. The figure illustrates the evolution of data analytics from basic understanding to advanced decision-making, represented as a progression across value and complexity. Adapted from Houtmeyers, 2021

Descriptive statistics, utilizing current and historical data, helps to identify trends and relationships while summarizing collected information to answer the question, 'What happened?' (Dees, 2022). To uncover the root causes of outcomes and gain a deeper understanding of the problem, diagnostic analytics addresses the question, 'Why did it happen?' by bridging descriptive and predictive analytics. This iterative process employs a combination of techniques such as hypothesis testing, regression analysis, correlation analysis, time-series analysis, and model simulation to identify patterns, relationships, and anomalies in data. Nowadays, predictive sports analytics (PSA) is widely adopted in a wide range of sports, each with unique data and modeling requirements. It is using past statistics, match outcomes, player`s and team`s performance metrics to identify the peak of performance or predict net results. For instance, in team sports (football, rugby, handball, basketball, volleyball) models predict team and player performance, league standings, and even player positions on the field (Apostolou & Tjortjis, 2019; Pantzalis & Tjortjis, 2020). Structured scoring systems in tennis benefit from predictive models that estimate match outcomes (Sarcevic, 2022). Basketball, hockey, and cricket analytics focus on evaluating player and team performance, using advanced metrics to inform strategy and decision-making (M., 2011; Sarlis, 2020). A novel hybrid approach named CRITIC-VIKOR was proposed by Liu (2024). By using fuzzy logic-based models, it enhance prediction accuracy by effectively managing the ambiguity of data, offering a reliable integral tool to support decision-making. Additionally, models based on exponential power functions (Kissell, 2019), decision tree and logistic regression (Gifford & Bayrak, 2023) have demonstrated high accuracy in forecasting NFL game outcomes, providing an objective and transparent method for ranking teams and estimating win probabilities. In swimming performance analysis, PSA involves the use of data-driven techniques ranging from stroke classification to race performance assessment. Staunton (2024) utilized Principal Component Analysis (PCA), one of the key techniques in PSA, and multiple regression to improve performance prediction in short-course swimming by reducing redundant variables and identifying key performance indicators (KPIs), revealing strong agreement between predicted and actual outcomes (95% LOA), with free-swimming speed (FSS) and turn performance being crucial for longer races (200 m, 400 m, 800 m) and start performance for sprints (50 m, 100 m). Machine learning models (ML) and artificial neural networks (ANN) became popular over the past decades for their predictive capabilities. Powell (2025) applied ML models to forecast medal-winning performances in international swimming events, evaluating two statistical models: a linear regression and machine-learning approach (0.80% MAE) and a Bayesian framework (0.85% MAE), both demonstrating high accuracy (r = .99) in predicting results for the Budapest 2022 World Championships, providing a data-driven framework for performance forecasting at the Olympic Games in Paris 2024. The integration of wearable technologies with advanced AI techniques is becoming an indispensable component of technological advancement, shaping the future of innovation. One notable example is the study conducted by Delhaye (2022) which implemented a deep learning model aimed at analyzing the swimming activity and lap time using a single IMU attached to the sacrum to facilitate real-time tracking and assessment of swimmers during training.

Methodology Gap and Motivation for AI-PIONEER

While data analysis has proven effective in identifying trends, patterns, relationships, causes of past outcomes and forecasting outcomes in sports performance, a gap remains in fully integrating AI as an expert system to make reliably informed decision-making that adapts dynamically to training needs. There is a broad consensus among researchers that expertise is a fundamental component of effective athletic development. In order to elevate the training process, AI paired with advanced technologies must be designed not only to predict outcomes but also to encompass expert-level decision-making. Traditional methods of performance assessment are often subjective, time-consuming and require complex expertise to interpret the data accurately. As stated by Houtmeyers (2021), elite sports remain predominantly data-informed rather than fully data-driven. Despite the growing number of research studies exploring the application of AI and ML in human movement science, minimal to no research has been done on prescriptive systems for swimming performance assessment, as existing studies mainly focused on *predictive*, *descriptive* or *diagnostic* analytics. Significant gaps still exist in regard to the optimization of swimming stroke technique based on kinematic parameters. The integration of artificial and machine learning techniques presents a promising solution to these challenges by automating motion analysis, reducing human bias, and improving the accuracy of performance assessment. This necessitates the development of AI models that utilize cutting-edge techniques to provide prescriptive or actionable insights with the depth and adaptability of human expertise. Prescriptive systems suggest a series of the best possible outcomes and the consequences of each decision option available (Houtmeyers, 2021). An AI-driven prescriptive system for evaluating swimming performance based on kinematic parameters holds the potential in enhancing athlete training and performance assessment.

Figure 4. Framework for AI-Based Prescriptive Performance Assessment in Swimming



Note. This flowchart illustrates how performance assessment can moved from traditional methods to advanced AI-based solutions, leading to a multifaceted system that helps improve athlete performance.

Premise of the Hypothesis

The system development requires the choice of appropriate ML models and algorithms, inclusion of measurable biomechanical data, and performance contributing factors to ensure accurate and reliable results. Picking the right types of data for training the AI is key to how well it works. Studies suggest that spatiotemporal data, along with body measurements and physical characteristics, should be used to assess swimming technique and performance. One major challenge is making sure all these different types of data are aligned and useful for AI analysis. It's also important to select the most relevant features and high-quality data for AI to spot the differences from elite-level swimming standards. Solving these issues will help create a strong AI system that can give helpful feedback to improve butterfly swimming performance.

Primary Research Questions

i. How can AI be utilized to develop a prescriptive system for evaluating butterfly swimming performance?
ii. Which ML models or algorithms are most effective for analysing and optimizing butterfly swimming techniques?
iii. How can human biomechanical data be integrated into an AI system to ensure accurate and reliable performance evaluation?

Secondary Research Questions

i. What types of data are critical for training the AI-based prescriptive system?
ii. How can spatiotemporal swimming data be normalized and aligned with anthropometric and somatic parameters for effective AI model inputs?

iii. Which metrics or features should the AI prioritize to accurately assess deviations from elite-referenced swimming techniques?

Research Question and Hypothesis

- Can a CVAE-based prescriptive system accurately generate IMU sequences from anthropometric and somatic data to facilitate motion analysis while reducing reliance on extensive sensor input?
- A CVAE-based prescriptive system generating IMU sequences from anthropometric and somatic data, enabling motion analysis with reduced sensor dependency.

Objectives

a) To develop a CVAE-based prescriptive system capable of generating IMU sequences using anthropometric and somatic data.
b) To evaluate the accuracy of the predicted IMU sequences in comparison to actual sensor data.

Specific Aims

The primary aim of this study is to develop and validate an AI-based prescriptive system for optimizing butterfly swimming performance based on elite-referenced motion data. Ethical approval will be obtained from the Nazarbayev University Institutional Research Ethics Committee (IREC), followed by the recruitment of four competitive butterfly swimmers with over seven years of continuous training experience to represent elite-level performance standards. Data collection will involve recording anthropometric, somatic, and spatiotemporal kinematic data from these elite athletes using a wearable 3D motion capture system (Ultium Motion System, Noraxon U.S.A., Inc.) during dry swimming trials. Additionally, anthropometric and somatic parameters will be collected from amateur swimmers to serve as input for the model. The collected data will undergo preprocessing, including orientation estimation using the Madgwick filter, position extraction via the Kalman filter, Z-score normalization of all IMU and anthropometric data, and temporal synchronization of stroke cycles. For data labeling, IMU sequence data will be cleaned and labeled, anthropometric and somatic parameters will be assigned as conditional inputs, and IMU sequences will serve as ground truth outputs for model training. The final phase involves training and validating a Conditional Variational Autoencoder (CVAE) to generate IMU sequences from input features. The model's accuracy and reliability in replicating elite-referenced motion profiles will be evaluated, with the goal of reducing dependence on extensive sensor setups and providing precise, data-driven feedback for performance improvement.

1. Participant Recruitment & Ethical Approval

a) Obtain ethical clearance from the NU Institutional Research Ethics Committee (IREC).

b) Recruit 4 competitive butterfly swimmers with over 7 years of continuous training experience to represent elite-referenced performance.

2. Data Collection

2.1. Elite Athletes (for Model Training): record anthropometric parameters, somatic parameters, and kinematic (spatiotemporal) data using a wearable 3D Motion Capture system (Ultium Motion System, Noraxon U.S.A., Inc.) during dry swimming trials.

2.2. Amateur Swimmers (for Model Input): collect anthropometric and somatic parameters for use as input in the model.

3. Data Preprocessing

i. Apply Madgwick Filter for orientation estimation (quaternion format).

ii. Use a Kalman Filter to extract position data.

iii. Normalize all IMU and anthropometric data using Z-score normalization.

iv. Detect stroke cycles and perform temporal synchronization of input-output sequences.

4. Data Labeling

i. Clean and label IMU sequence data.

ii. Assign anthropometric and somatic parameters as conditional inputs.

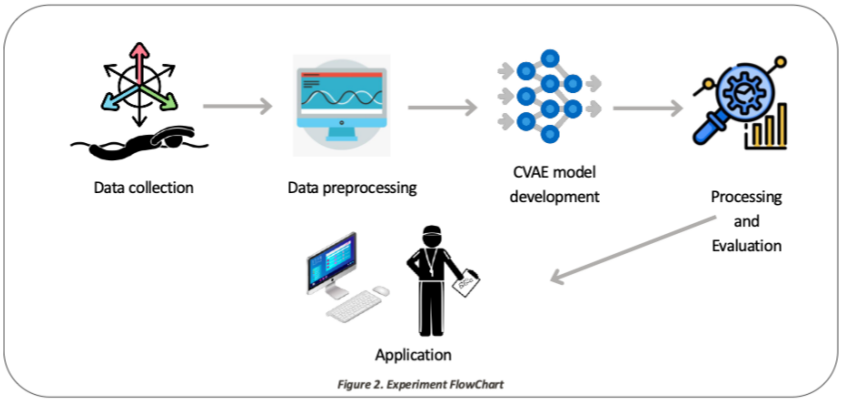
iii. Set the IMU sequences as ground truth output for model training.

5. Model Validation

a) Train and validate a Conditional Variational Autoencoder (CVAE) to generate IMU sequences based on anthropometric and somatic input.

b) Evaluate the accuracy and reliability of the model in replicating elite-referenced motion profiles, aiming to reduce dependency on extensive sensor setups.

Figure 5. Flowchart of Experimental Design



Chapter 2

MATERIALS AND METHODS

2.1 Ethical Consideration

The study was conducted in accordance with ethical guidelines for human research. Prior to participation, all participants were informed about the study's purpose, procedures, and potential risks and provided written informed consent, confirming their voluntary participation and the right to withdraw at any stage without consequences. On **18/12/2024**, the Nazarbayev University Institutional Research Ethics Committee (NU-IREC) approved the submission 994/12112024 "Artificial Intelligence Based Expert System of Butterfly Swimming Assessment.

2.2 Participants

Four male student-athletes 20.75 ± 2.87 years old specializing in butterfly swimming with no less than seven years of training experience participated in the study. The average personal best 50m butterfly swimming time of participants is 29.03 ± 2.31. Athletes had no injuries in the 7 months prior (Table 1).

Table 1. Participant Demographics

Participant	Gender	Age	Years of Swimming	Personal Best 50m Butterfly (s)
1	Male	18	7	30.92
2	Male	19	7	30.21
3	Male	21	12	25.60
4	Male	25	8	29.40

Note. All participants are male competitive swimmers.

2.3 Data Collection

To support the development of individualized performance models, detailed anthropometric and somatic measurements were collected from each swimmer. Anthropometric parameters were assessed using standard manual measurement protocols, whereas somatic parameters were analyzed using the InBody 770 bioelectrical impedance analyzer (Table 2). Kinematic data of dry swimming were obtain using a motion capture system and an array of Noraxon (Ultium Motion System, Noraxon U.S.A., Inc.) IMU sensors.

.*2.3.1 Anthropometric and Somatic Variables for Model Developement*

These parameters were chosen for their relevance in evaluating swimmers’ body proportions, stroke efficiency, and range of motion.

Table 2. Parameters for Cvae Model Developement

FEATURE NAME	UNIT	DESCRIPTION
BODY PARAMETERS		
Height	cm	Total body height
Weight	kg	Body mass
Arm Span	cm	Distance from fingertip to fingertip with arms extended
Leg Length	cm	Distance from hip to foot
Torso Length	cm	Distance from shoulder to hip
Shoulder Width	cm	Distance between shoulders
Hand Size	cm^2	Surface area of the hand
Foot Size	cm^2	Surface area of the foot
Body Fat %	%	Percentage of body mass from fat
Muscle Mass	kg	Estimated muscle mass
SWIMMING PARAMETERS		
Stroke Rate	Hz (strokes/sec)	Number of strokes per second
Stroke Length	m/stroke	Distance traveled per stroke
Stroke Index	m^2/s	Efficiency metric (Stroke Length × Stroke Rate)
Stroke Symmetry	%	Measures difference between left & right strokes
Kick Frequency	Hz	Number of kicks per second
Kick-Amplitude	cm	Vertical displacement of feet per kick
Breathing Pattern	Ratio	Number of strokes per breath
Swim Speed	m/s	Average velocity of a swimmer
Cycle Duration	sec	Time taken for one full stroke cycle
Arm Recovery Time	sec	Time taken to recover the arm above water
Propulsion Time	sec	Time when the arm generates forward force

2.3.2 IMU Data Collection for Model Training

The participants performed dry swimming exercises under standardized conditions. The testing area was equipped with a motion capture system and an array of Noraxon (Ultium Motion System, Noraxon U.S.A., Inc.) IMU sensors, strategically placed to track kinematic parameters. A calibrated reference frame was established to align sensor data with real-world motion. The protocol involved each participant underwent a standardized warm-up protocol consisting of dynamic stretching and neuromuscular activation exercises to ensure consistency in muscle recruitment and movement execution. The trials consisted of simulated butterfly strokes, performed in a prone position on an elevated surface on a couch to mimic in-water conditions while facilitating precise motion tracking. Each subject completed three trials of a 30-second simulated butterfly swimming movement at self-selected competitive intensities, with a two-minute rest period between trials to prevent fatigue-induced variability. The experimental setup was designed to simulate the biomechanical conditions of the in-water butterfly stroke while minimizing external variability. Participants were instructed to perform coordinated arm and leg movements following the prescribed butterfly stroke technique, ensuring synchronization of the Arm Propulsion Phase (AP), Arm Recovery Phase (AR), and Leg Total Propulsion Phase (LP). Real-time feedback was provided via motion tracking software to monitor adherence to correct form. Instructors observed the sessions to confirm compliance with the established protocol and to ensure data integrity.

Before data collection, sensor calibration was conducted to eliminate drift and ensure accuracy. This included:

- baseline static calibration (participants held a neutral posture for five seconds to establish reference values);
- dynamic calibration (participants performed a standardized movement sequence to verify sensor responsiveness).

Kinematic data were collected using Noraxon IMU sensors, which were placed on hands (L, R) wrist (L, R), elbow (L, R), shoulder (L, R), trunk, lower back, head, pelvis, thighs (L, R), knees (L, R) and feet (L, R) to capture movement patterns. The sensors recorded data at a sampling rate of 5 Hz, providing detailed motion analysis. Data acquisition was conducted using myoResearch MR3, and all measurements were synchronized to ensure accuracy.

2.3.3 IMU Signal Preprocessing.

Preprocessing involved signal filtering using a low-pass Butterworth filter (cutoff frequency 5 Hz) to remove noise and enhance motion signal clarity.

The collected data were processed using myoResearch MR3, applying resampling and gravity compensation to eliminate noise and extract relevant kinematic parameters. Statistical analysis was performed using, with a significance level set at.

2.4 CVAE METHODOLOGY

The purpose of the method

The methodology aims to develop a model that takes a swimmer's anthropometric values and IMU sensor data to generate a time-motion sequence, comprising position and orientation vectors for each body segment at any given moment, by adapting the motion patterns of an elite swimmer to the individual's parameters.

Data Acquisition and Preprocessing

The model development required two distinct datasets: one comprising elite swimmers and another composed of non-elite, amateur swimmers. The dataset for elite swimmers served as the training base for the CVAE model. It included both input features and ground-truth outputs. Input features consisted of anthropometric and somatic variables such as height, weight, arm span, leg length, hand and foot size, torso-to-leg ratio, muscle mass, and body fat percentage. The outputs were time-series IMU recordings collected via eight body-mounted sensors, each capturing linear acceleration and angular velocity across three axes (x, y, z), resulting in a 48-dimensional vector per time step. These recordings were collected over full butterfly stroke cycles and sampled at a frequency of 100 Hz.

In addition to raw sensor data, derived swimming parameters such as stroke rate, stroke length, kick frequency, propulsion time, swim speed, and stroke symmetry were extracted. The data were then segmented into individual stroke cycles and temporally aligned using interpolation to ensure consistency in sequence length across samples. Noise and anomalies were smoothed using biomechanical filtering techniques to ensure that the time-series data adhered to realistic motion patterns.

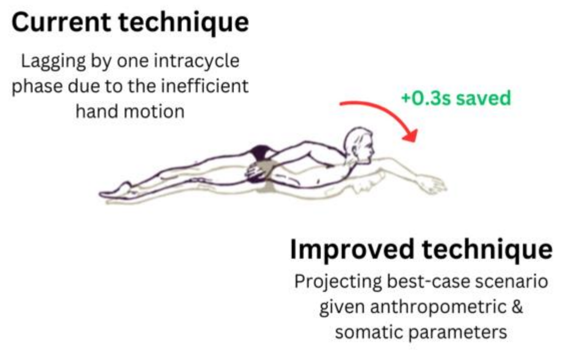
As an input, only the anthropometric and somatic variables of amateur swimmers were collected. Trained CVAE model synthetically generated IMU sequences for these individuals during the inference phase.

All variables were normalized to ensure consistency and enable efficient model training. Standardization (z-score normalization) and min-max scaling were selectively applied depending on the distribution of each feature. Special consideration was given to scaling joint data ranges appropriately to maintain physical realism across the feature space.

Basic principles & key components.

This study proposes a novel methodological framework based on a Conditional Variational Autoencoder (CVAE) to generate biomechanically valid and temporally consistent IMU sensor sequences for butterfly swimming strokes. The generated sequences are specifically tailored to the anthropometric and somatic characteristics of individual non-elite swimmers. The approach integrates principles of machine learning, biomechanics, and data normalization to establish an expert assessment system that can simulate idealized swimming performance and derive actionable metrics for personalized technique improvement (Figure 6).

Figure 6. Prototype Visualization of Technique Optimization Using Motion Engine

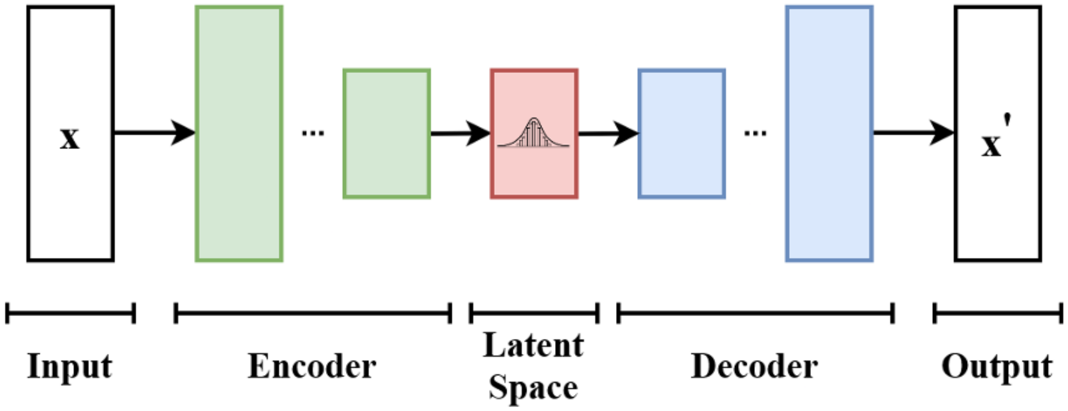


Note. Illustration of current vs. Improved butterfly stroke technique. The motion engine visually demonstrates technique inefficiences and project biomechanically optimal recomendations based on individual parameters. Improved technique is derived from IMU data generated by CVAE model.

CVAE Model Architecture

CVAE is the type of architecture that consists of 3 main parts: encoder, decoder and latent space. Like a traditional autoencoder, CVAE learns 2 functions: an encoding function (which could be thought of as compressing the information) and a decoding function (analogously decompressing). CVAE is also similar to the variational autoencoder in a way that both encode information to the latent space in the form of a probability distribution (Figure 7).

Figure 7. Variational Autoencoder Structure



The reasoning behind that type of mapping lies in the necessity to make latent space structured. It is modeled as a multivariate Gaussian distribution, parameterized by a mean vector (μ) and a log-variance vector (σ2). A 3- dimensional latent space was selected based on the complexity of the biomechanical data and the need for interpretability. Latent Variable Sampling:

$$z = \mu + \varepsilon \times \sqrt{\sigma^2}\,, where\ \varepsilon \sim \mathcal{N}(0, I)$$

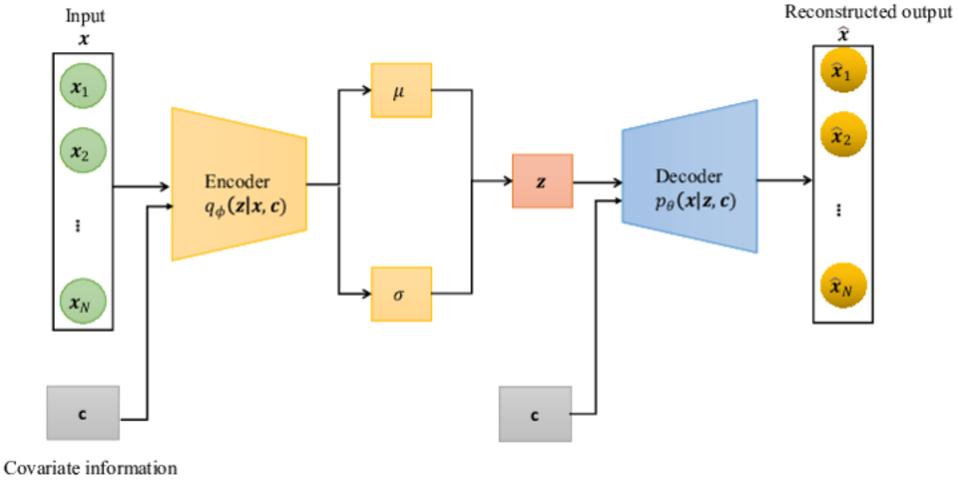
Note. Reparameterization Trick

Without it, we would map the result from the encoder directly to the latent space that could lead to unstructured and discontinuous space that is only able to directly reconstruct the input. The sampled latent vector zzz, along with the swimmer’s anthropometric and somatic condition vector, is then passed to the decoder. The decoder utilizes Temporal Convolutional Networks (TCNs) with dilated convolutions, which are well-suited for modeling long-range temporal dependencies while enabling parallel computation. The decoder reconstructs a time-series IMU sequence that is biomechanically valid and customized to the input body profile.

However, such modification allows sample z-score from latent space around that probability distribution meaning now output is not one-to-one reconstruction but rather generation that is similar to the input. Important caveat here is that such generation is often considered “blurry” (e.g. in a context of image generation) due to the nature of the normal distribution. Yet, in the context of generating motion dynamics of a body part, such blurry output is considered acceptable since the aim is generate general motion patterns rather than sharp & realistic outputs.

Now when it comes to Conditional VAE (Figure 7), one important distinction from VAE lies in the ability of the first to condition the input to a class label.

Figure 8. Conditional Variation Autoencoder Structure



Note. Adapted from Abdelli, 2023

Loss Function Design and Training

Three main components, for composite loss function, were included to make sure the model not only learns meaningful representations but also adheres to biomechanical realism:

1. **Reconstruction Loss**: Mean squared error (MSE) between the reconstructed IMU sequences and the original sequences from elite swimmers.

$$\mathcal{L}\_rec = MSE(X\_true, X\_reconstructed)$$

2. **Kullback-Leibler Divergence**: Regularization term enforcing that the latent space distribution approximates a unit Gaussian distribution, facilitating efficient sampling.

$$\mathcal{L}\_KL = D\_KL(\mathcal{N}(\mu, \sigma^2) \mid\mid \mathcal{N}(0, I))$$

3. **Physics-Informed Losses**: These ensure biomechanical validity of the generated data and include:
    - *Sensor Consistency Loss*: Penalizes discrepancies between accelerometer and gyroscope derivative alignment.
    - *Biomechanical Constraint Loss*: Ensures that reconstructed sequences comply with known joint angle limits and propulsion symmetry.
    - *Temporal Smoothness Loss*: Encourages continuity by minimizing frame-to-frame motion discontinuities.

The total loss is a weighted sum of these components:

$$\mathcal{L}\_total = \lambda_1 \mathcal{L}\_rec + \lambda_2 \mathcal{L}\_KL + \lambda_3 \mathcal{L}\_physics$$

The model was trained using the elite swimmer dataset. Data were split into training, validation, and test subsets to monitor generalization performance. Data augmentation was performed via interpolation and synthetic transformation of anthropometric features to expand the dataset and reduce overfitting.

Inference and Application for Non-Elite Swimmers

Upon training completion, the CVAE model was deployed for inference on amateur swimmers. By providing only the anthropometric and somatic condition vector of a given amateur swimmer, the decoder generated a customized IMU sequence representing an idealized butterfly stroke cycle, tailored to that individual's physical profile.

From these generated sequences, essential performance metrics were computed, including stroke rate, stroke length, swim speed, propulsion time, and arm recovery time. These parameters were either directly calculated from the IMU data or inferred using regression models trained on elite swimmer data.

Validation

Model-generated sequences were validated through multiple layers of analysis. First, biomechanical plausibility was assessed by checking generated values against known ranges for acceleration, angular velocity, stroke symmetry, and propulsion dynamics. Second, generated sequences were compared with real IMU recordings from elite swimmers with similar body proportions to ensure qualitative similarity. Amateur swimmers who received generated stroke profiles underwent performance tracking to evaluate the practical impact of model-generated recommendations.

Deployment and Use Case Scenarios

The CVAE model offers several practical applications. Amateur swimmers and coaches can utilize the model to visualize optimized stroke cycles via 3D motion engines, compare current versus ideal performance, and receive prescriptive advice grounded in personalized biomechanical insights. This approach enables the delivery of evidence-based performance feedback even in resource-limited environments without requiring real-time sensor data collection from each athlete.

The model also supports additional functionality, such as estimating predicted swim times based on optimized technique parameters, thereby offering a quantitative assessment of potential performance gains from stroke corrections.

Future Studies

To improve robustness and expand utility, future work will focus on incorporating real-time feedback using wearable waterproof IMUs and increasing sample size to minimum 100 datasets. This will enable dynamic assessment and immediate coaching intervention during swim training. Additionally, a federated learning framework will be implemented to allow secure model updates from decentralized data sources, increasing anthropometric diversity in the training set while preserving swimmer privacy. Finally, the framework will be extended to support other swimming through domain-adapted retraining and transfer learning strategies.

Chapter 3

RESULTS

The results of this study mainly focus on visual outputs of generated sequences. The generated IMU sequences exhibit smooth waveforms that simulate actual trajectory for Head, Right Upper Arm, Left Thigh according to strategically placed IMU sensors for specific region (ax, ay, az). The acceleration and angular velocity curves follow realistic temporal dynamics, with peak activity corresponding to expected propulsion and recovery phases. When compared to actual IMU sequences from elite swimmers, the synthetic outputs display similar temporal structures and amplitude ranges, suggesting the model's ability to generalize biomechanically consistent stroke patterns. Sensors outputs (ax, ay, az) and generated sequences draw periodic peaks consistent to intra-cycle (within one cycle) with peaks consistent with arm propulsion cycles, while gyroscope data from the lower leg sensor reflects the undulatory motion of the dolphin kick.

Figure 10 shows the predicted (dashed red lines) and original (blue lines) head trajectories across the XY, XZ, and YZ planes. The model-generated sequences capture the general directional trend and spatial dynamics exhibited in the original IMU signals. Particularly in the YZ plane, the model performs with notable accuracy, following the curvature and range of vertical movement effectively.

In the XY and XZ projections, while the overall downward-forward progression is retained, minor deviations are observed, particularly around the curvature transitions, which suggest phase shifts or damping effects in generated motion. However, the trajectories remain biomechanically plausible and reflect essential swimming-specific kinematic characteristics.

Figure 10. Predicted vs. Actual Trajectory for Head

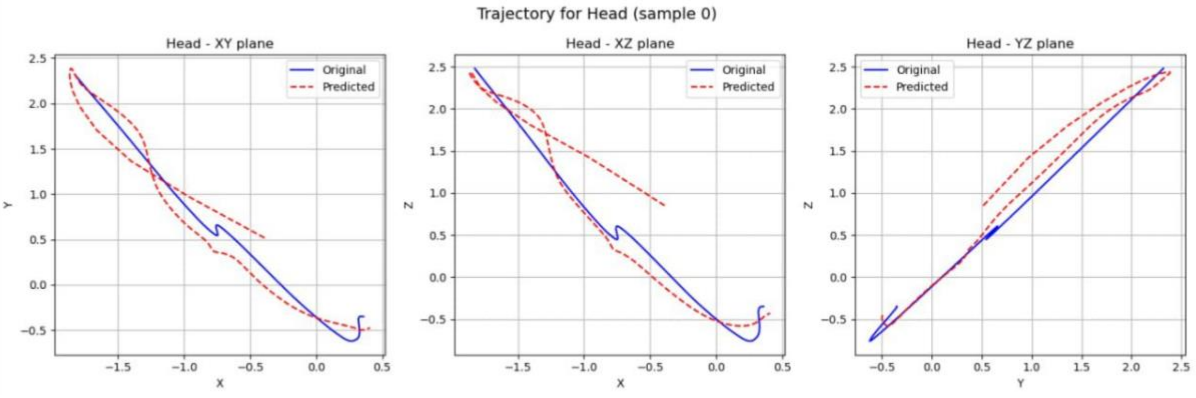
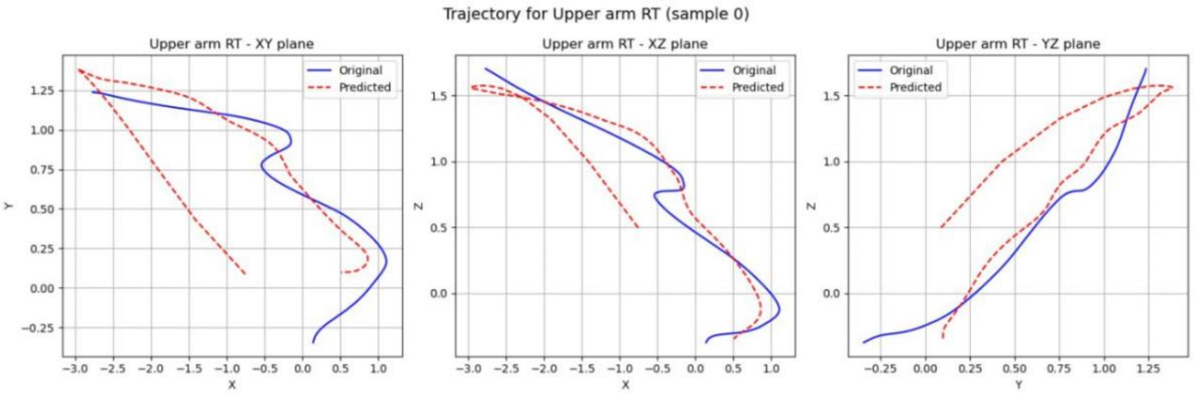


Figure 11 shows the trajectory comparisons for the right upper arm. The original trajectories exhibit complex curved paths in all three planes, especially in XZ, reflecting the whip-like recovery motion.

The CVAE model successfully reconstructs the structure of the motion, preserving the general sweep phase. However, the predicted sequences demonstrate smoother

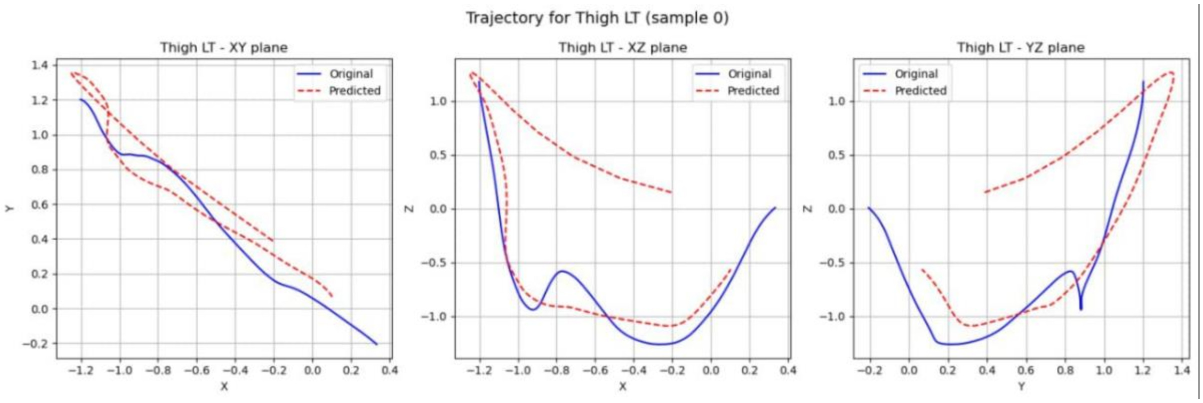
Figure 11. Predicted vs. Actual Trajectory for R Upper Arm



curves and occasionally under-represent sharp transitions seen in the original data. Despite these simplifications, the synthesized outputs align well with expected biomechanical patterns.

The generated output for the Left Thigh sensor (Figure 12), demonstrates nearly similar line curve as a real IMU sensor data. There are some deviations in X and Y planes, but the model is capable of repeating the motion pattern.

Figure 12. Predicted vs. Actual Trajectory for L Thigh



The qualitative similarity across all views suggests that the model is capable of generating biomechanically plausible sequences that retain essential movement dynamics. However, slight deviations in trajectory amplitude and phase alignment indicate areas for improvement in fine-grained spatial accuracy. While the current results are qualitative, the visual outputs from CVAE model confirm its capability to generate motion sequences.

DISCUSSION AND LIMITATIONS

This study introduces Conditional Variational Autoencoder (CVAE) as a novel approach to generating biomechanically plausible IMU sequences for non-elite butterfly swimmers based on static anthropometric and performance parameters. The core contribution lies in integrating deep generative modeling with domain-specific physics constraints to overcome the scarcity of training data and enable individualized motion prediction. Despite the innovation in architecture and conceptual design, several critical limitations substantially affect the current reliability and applicability of the model.

Compared to related studies that utilize (1) deep learning for swimming activity classification (Delhaye, 2022); (2) machine learning for activity recognition (Chen, 2023); (3) macro analysis for swimming bout and lap detection, and swimming technique identification (Rad, 2021); musculoskeletal simulation using MoCap data (Nakashima, 2019); biomechanical profiling and functional clustering of variabilities (Bouvet, 2024), our model's generative approach provides not only classification but also simulation capabilities, allowing prescriptive feedback generation a methodological advance over prior diagnostic-focused systems.

The CVAE framework demonstrated a promising proof of concept: the system successfully reconstructed IMU time series data conditioned on individual static inputs (anthropometrics and somatics), producing output sequences that represent idealized stroke patterns. The use of TCNs and LSTM layers provided a foundation for modeling temporal dependencies without relying on large datasets. Additionally, the incorporation of physics-informed loss functions, enforcing consistency between gyroscope and accelerometer signals, joint biomechanics, and temporal smoothness, added a layer of plausibility to the generated sequences, aligning them more closely with real-world movement dynamics.

However, the limitations of this study are significant and must be acknowledged. Most notably, the extremely limited dataset (only four elite swimmers) severely restricts the generalizability and robustness of the trained model. With such a small sample size, the risk of overfitting is high, and any performance claims must be treated cautiously. Furthermore, the quality of the generated sequences remains suboptimal, with outputs occasionally appearing blurred or lacking fine-grained structure. This suggests that the latent space may not yet adequately capture the variability required for high-fidelity motion synthesis.

Another challenge is the complexity of the IMU data itself. High sampling rates introduce noise and make it difficult to model long-term dependencies or general stroke patterns without additional filtering or regularization. Moreover, the CVAE's reliance on random sampling from the latent space (z) reduces the interpretability and reproducibility of outputs. In practice, this could lead to inconsistent feedback when applied in coaching or training contexts.

These limitations highlight that, at its current stage, this research does not deliver a deployable performance prediction or correction tool. Instead, its value lies in the methodological groundwork and demonstration that such a system can, in principle, learn meaningful relationships between anthropometric features and IMU motion sequences.

Looking forward, the potential of this framework remains considerable. Increasing the dataset size to include over 100 athletes, implementing real-time and federated learning strategies, and integrating more advanced generative architectures such as GANs may dramatically improve sequence sharpness and prediction accuracy. The use of waterproof IMU sensors specifically designed for in-pool data capture, as well as improvements in the design of temporal labeling and normalization pipelines, will also be necessary to transition from conceptual demonstration to applied utility.

In conclusion, while this study presents a technically sound and novel methodological contribution, its practical impact is currently constrained by data limitations and sequence fidelity issues. These findings highlight the importance of rigorous dataset expansion and refinement to realize the full potential of AI-driven prescriptive systems in elite sport contexts.

CONCLUSION

This study proposed a Conditional Variational Autoencoder model for generating personalized, biomechanically plausible IMU motion sequences conditioned on static anthropometric and swimming performance parameters. By integrating machine learning with domain-specific physical constraints, the model demonstrates the feasibility of simulating elite-referenced movement patterns for non-elite athletes, offering a potential tool for technique optimization in butterfly swimming.

While the model architecture shows promise in terms of learning from extremely limited data and incorporating biomechanical reasoning, the current implementation is constrained by several factors. Chief among them is the insufficient dataset size, which limits generalizability, and the relatively low fidelity of generated sequences, which restricts practical application. The study, therefore, should be interpreted as a proof of concept rather than a ready-to-deploy solution.

Despite these limitations, the work lays a strong foundation for future research into AI-driven, personalized sports performance analysis. Expanding the dataset, refining the generative process, and exploring hybrid architectures with adversarial training and real-time feedback systems are logical next steps. Ultimately, the CVAE framework represents an important step toward scalable, athlete-specific modeling systems that can support coaches, athletes, and practitioners in data-driven decision-making for performance improvement.

REFERENCES

A., D., M., T., L., P., Mantripragada, N., P., J., Machado, L., B., F., I., A., & J., A. (2010). Modelling Hydrodynamic Drag in Swimming using Computational Fluid Dynamics. In H. Woo (Ed.), *Computational Fluid Dynamics*. InTech. https://doi.org/10.5772/7112

AI and Expert Systems: All You Need to Know. (n.d.). Retrieved February 12, 2025, from https://www.jaroeducation.com/blog/what-are-expert-systems-in/

Apostolou, K., & Tjortjis, C. (2019). Sports Analytics algorithms for performance prediction. *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, 1–4. https://doi.org/10.1109/IISA.2019.8900754

Barbosa, T. M., Barbosa, A. C., Simbaña Escobar, D., Mullen, G. J., Cossor, J. M., Hodierne, R., Arellano, R., & Mason, B. R. (2023). The role of the biomechanics analyst in swimming training and competition analysis. *Sports Biomechanics*, *22*(12), 1734–1751. https://doi.org/10.1080/14763141.2021.1960417

Barbosa, T. M., Bragada, J. A., Reis, V. M., Marinho, D. A., Carvalho, C., & Silva, A. J. (2010). Energetics and biomechanics as determining factors of swimming performance: Updating the state of the art. *Journal of Science and Medicine in Sport*, *13*(2), 262–269. https://doi.org/10.1016/j.jsams.2009.01.003

Barbosa, T. M., Fernandes, R. J., Morouco, P., & Vilas-Boas, J. P. (2008). Predicting the intra-cyclic variation of the velocity of the centre of mass from segmental velocities in butterfly stroke: A pilot study. *Journal of Sports Science & Medicine*, *7*(2), 201–209.

Barbosa, T. M., Keskinen, K. L., Fernandes, R., Cola□o, P., Lima, A. B., & Vilas-Boas, J. P. (2005). Energy cost and intracyclic variation of the velocity of the centre of mass in butterfly stroke. *European Journal of Applied Physiology*, *93*(5–6), 519–523. https://doi.org/10.1007/s00421-004-1251-x

Bartlett, R. (2006). Artificial intelligence in sports biomechanics: New dawn or false hope? *Journal of Sports Science & Medicine*, *5*(4), 474–479.

Biró, A., Cuesta-Vargas, A. I., & Szilágyi, L. (2024). AI-controlled training method for performance hardening or injury recovery in sports. *2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 000259–000264. https://doi.org/10.1109/SAMI60510.2024.10432911

Bodemer, O. (2023). *Enhancing Individual Sports Training through Artificial Intelligence: A Comprehensive Review*. https://doi.org/10.36227/techrxiv.24005916.v1

Boulesteix, L., & Chollet, D. (2001). *EVOLUTION OF THE BUTTERFLY COORDINATION IN RELATION TO VELOCITY AND SKILL LEVEL OF SWIMMERS*. 22–26.

Bouvet, A., Pla, R., Delhaye, E., Nicolas, G., & Bideau, N. (2024). Profiling biomechanical abilities during sprint front-crawl swimming using IMU and functional clustering of variabilities. *Sports Biomechanics*, 1–21. https://doi.org/10.1080/14763141.2024.2368064

C. Cao, D., & Sun, Y. (2024). An Intelligent Program to Assist in Swim Training and Technique Optimization Utilizing Machine Learning and Motion Tracking. *Networks, Blockchain and Internet of Things*, 187–196. https://doi.org/10.5121/csit.2024.1405116

Cabarkapa, D., Whetstone, J. M., Patterson, A. M., Mosier, E. M., Cabarkapa, D. V., & Fry, A. C. (2022). Relationship between Health-Related Physical Fitness Parameters and Functional Movement Screening Scores Acquired from a Three-Dimensional Markerless

Motion Capture System. *International Journal of Environmental Research and Public Health*, *19*(8), 4551. https://doi.org/10.3390/ijerph19084551

Ceseracciu, E., Sawacha, Z., Fantozzi, S., Cortesi, M., Gatta, G., Corazza, S., & Cobelli, C. (2011). Markerless analysis of front crawl swimming. *Journal of Biomechanics*, *44*(12), 2236–2242. https://doi.org/10.1016/j.jbiomech.2011.06.003

Chen, Z., & Dai, X. (2024). Utilizing AI and IoT technologies for identifying risk factors in sports. *Heliyon*, *10*(11), e32477. https://doi.org/10.1016/j.heliyon.2024.e32477

Chmait, N., & Westerbeek, H. (2021). Artificial Intelligence and Machine Learning in Sport Research: An Introduction for Non-data Scientists. *Frontiers in Sports and Active Living*, *3*, 682287. https://doi.org/10.3389/fspor.2021.682287

Cortesi, M., Giovanardi, A., Gatta, G., Mangia, A. L., Bartolomei, S., & Fantozzi, S. (2019). Inertial Sensors in Swimming: Detection of Stroke Phases through 3D Wrist Trajectory. *Journal of Sports Science & Medicine*, *18*(3), 438–447.

Dalamitros, A. A., Manou, V., & Pelarigo, J. G. (2014). Laboratory-based tests for swimmers: Methodology, reliability, considerations and relationship with front-crawl performance. *Journal of Human Sport and Exercise*, *9*(1), 172–187. https://doi.org/10.4100/jhse.2014.91.17

Daukantas, S., Marozas, S., & Lukosevicius, A. (2008). *Inertial sensor for objective evaluation of swimmer performance*.

Dedhia, U., Bhoir, P., Ranka, P., & Kanani, P. (2023). Pose Estimation and Virtual Gym Assistant Using MediaPipe and Machine Learning. *2023 International Conference on Network, Multimedia and Information Technology (NMITCON)*, 1–7. https://doi.org/10.1109/NMITCON58196.2023.10275938

Dees, W., Walsh, P., McEvoy, C. D., McKelvey, S., Mullin, B. J., Hardy, S., & Sutton, W. A. (2022). *Sport marketing* (Fifth edition). Human Kinetics.

Delhaye, E., Bouvet, A., Nicolas, G., Vilas-Boas, J. P., Bideau, B., & Bideau, N. (2022). Automatic Swimming Activity Recognition and Lap Time Assessment Based on a Single IMU: A Deep Learning Approach. *Sensors*, *22*(15), 5786. https://doi.org/10.3390/s22155786

Derie, R., Robberechts, P., Van Den Berghe, P., Gerlo, J., De Clercq, D., Segers, V., & Davis, J. (2020). Tibial Acceleration-Based Prediction of Maximal Vertical Loading Rate During Overground Running: A Machine Learning Approach. *Frontiers in Bioengineering and Biotechnology*, *8*, 33. https://doi.org/10.3389/fbioe.2020.00033

Engel, A., Schaffert, N., Ploigt, R., & Mattes, K. (2021). Intra-cyclic analysis of the front crawl swimming technique with an inertial measurement unit. *Journal of Human Sport and Exercise*, *17*(3). https://doi.org/10.14198/jhse.2022.173.17

Exel, J., & Dabnichki, P. (2024). Precision Sports Science: What Is Next for Data Analytics for Athlete Performance and Well-Being Optimization? *Applied Sciences*, *14*(8), 3361. https://doi.org/10.3390/app14083361

Fantozzi, S., Giovanardi, A., Magalhães, F. A., Di Michele, R., Cortesi, M., & Gatta, G. (2016). Assessment of three-dimensional joint kinematics of the upper limb during simulated swimming using wearable inertial-magnetic measurement units. *Journal of Sports Sciences*, *34*(11), 1073–1080. https://doi.org/10.1080/02640414.2015.1088659

Fernandes, A., Mezêncio, B., Soares, S., Duarte Carvalho, D., Silva, A., Vilas-Boas, J. P., & Fernandes, R. J. (2024). Intra- and inter-cycle velocity variations in sprint front crawl swimming. *Sports Biomechanics*, *23*(12), 2960–2973. https://doi.org/10.1080/14763141.2022.2077815

Gifford, M., & Bayrak, T. (2023). A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression. *Decision Analytics Journal*, *8*, 100296. https://doi.org/10.1016/j.dajour.2023.100296

Gonjo, T., & Olstad, B. H. (2020). Race Analysis in Competitive Swimming: A Narrative Review. *International Journal of Environmental Research and Public Health*, *18*(1), 69. https://doi.org/10.3390/ijerph18010069

Grimston, S. K., & Hay, J. G. (1986). Relationships among anthropometric and stroking characteristics of college swimmers: *Medicine & Science in Sports & Exercise*, *18*(1), 60???68. https://doi.org/10.1249/00005768-198602000-00011

Guignard, B., Rouard, A., Chollet, D., & Seifert, L. (2017). Behavioral Dynamics in Swimming: The Appropriate Use of Inertial Measurement Units. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00383

Hall, A., Victor, B., He, Z., Langer, M., Elipot, M., Nibali, A., & Morgan, S. (2021). The detection, tracking, and temporal action localisation of swimmers for automated analysis. *Neural Computing and Applications*, *33*(12), 7205–7223. https://doi.org/10.1007/s00521-020-05485-3

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*, 793302. https://doi.org/10.3389/fbioe.2021.793302

Hamidi Rad, M., Gremeaux, V., Dadashi, F., & Aminian, K. (2021). A Novel Macro-Micro Approach for Swimming Analysis in Main Swimming Techniques Using IMU Sensors. *Frontiers in Bioengineering and Biotechnology*, *8*, 597738. https://doi.org/10.3389/fbioe.2020.597738

Houtmeyers, K. C., Jaspers, A., & Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, *16*(11), 1719–1723. https://doi.org/10.1123/ijspp.2020-0958

Hughes, M., & Franks, I. M. (Eds.). (2015). *Essentials of performance analysis in sport* (Second edition). New York, NY : Routledge.

Kamarudin, N. A., Abdullah, M. R., Musa, R. M., Eswaramoorthi, V., Maliki, A. B. H. M., Ab Rasid, A. M., & Nadzmi, A. (2022). A Study of Sports Performance Monitoring on Individual Sports and Team Sports Physical Fitness Performance Using Multivariate Approach. *International Journal of Academic Research in Progressive Education and Development*, *11*(1), Pages 331-341. https://doi.org/10.6007/IJARPED/v11-i1/12065

Lam, W. W. T., Tang, Y. M., & Fong, K. N. K. (2023). A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, *20*(1), 57. https://doi.org/10.1186/s12984-023-01186-9

Lapham, A. C., & Bartlett, R. M. (1995). The use of artificial intelligence in the analysis of sports performance: A review of applications in human gait analysis and future directions for sports biomechanics. *Journal of Sports Sciences*, *13*(3), 229–237. https://doi.org/10.1080/02640419508732232

Liu, S. (2021). Application Research of Artificial Intelligence in Swimming. In M. Atiquzzaman, N. Yen, & Z. Xu (Eds.), *Big Data Analytics for Cyber-Physical System in Smart City* (Vol. 1303, pp. 202–207). Springer Singapore. https://doi.org/10.1007/978-981-33-4572-0_29

Liu, T. (2024). Assessing the effectiveness of fuzzy logic-based models for predicting sports event outcomes: A CRITIC-VIKOR approach. *PLOS ONE*, *19*(12), e0313913. https://doi.org/10.1371/journal.pone.0313913

M., T., A., D., J., M., & J., A. (2011). Biomechanics of Competitive Swimming Strokes. In V. Klika (Ed.), *Biomechanics in Applications*. InTech. https://doi.org/10.5772/19553

Magalhaes, F. A. D., Vannozzi, G., Gatta, G., & Fantozzi, S. (2015). Wearable inertial sensors in swimming motion analysis: A systematic review. *Journal of Sports Sciences*, *33*(7), 732–745. https://doi.org/10.1080/02640414.2014.962574

Maglischo, E. W. (2003). *Swimming fastest: The essential reference on technique, training, and program design*. Human Kinetics.

Marinho, D. A., Barbosa, T. M., Neiva, H. P., Silva, A. J., & Morais, J. E. (2020). Comparison of the Start, Turn and Finish Performance of Elite Swimmers in 100 m and 200 m Races. *Journal of Sports Science & Medicine*, *19*(2), 397–407.

Mason, B. R. (1991). A Biomechanical Analysis of the Butterfly Stroke. In *EXCEL A Publication of the Australian Institute of Sport* (Peter Fricker, Richard Telford, Vol. 7). Australian Institute of Sport.

Matsuda, Y., Kaneko, M., Sakurai, Y., Akashi, K., & Yasuo, S. (2024). Three-dimensional lower-limb kinematics during undulatory underwater swimming. *Sports Biomechanics*, *23*(11), 2093–2107. https://doi.org/10.1080/14763141.2021.1995475

Mooney, R., Corley, G., Godfrey, A., Quinlan, L. R., & ÓLaighin, G. (2015). Inertial Sensor Technology for Elite Swimming Performance Analysis: A Systematic Review. *Sensors (Basel, Switzerland)*, *16*(1), 18. https://doi.org/10.3390/s16010018

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*, 691919. https://doi.org/10.3389/fphys.2021.691919

Morais, J. E., Barbosa, T. M., Nevill, A. M., Cobley, S., & Marinho, D. A. (2022). Understanding the Role of Propulsion in the Prediction of Front-Crawl Swimming Velocity and in the Relationship Between Stroke Frequency and Stroke Length. *Frontiers in Physiology*, *13*, 876838. https://doi.org/10.3389/fphys.2022.876838

Morais, J. E., Marinho, D. A., Arellano, R., & Barbosa, T. M. (2019). Start and turn performances of elite sprinters at the 2016 European Championships in swimming. *Sports Biomechanics*, *18*(1), 100–114. https://doi.org/10.1080/14763141.2018.1435713

Nevalainen, P., Kauhanen, A., Raduly-Baka, C., Laakso, M.-J., & Heikkonen, J. (2016). Video based Swimming Analysis for Fast Feedback: *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods*, 457–466. https://doi.org/10.5220/0005753704570466

Nevill, A. M., Oxford, S. W., & Duncan, M. J. (2015). Optimal Body Size and Limb Length Ratios Associated with 100-m Personal-Best Swim Speeds. *Medicine & Science in Sports & Exercise*, *47*(8), 1714–1718. https://doi.org/10.1249/MSS.0000000000000586

Nicholls, S. B., James, N., Bryant, E., & Wells, J. (2019). The implementation of performance analysis and feedback within Olympic sport: The performance analyst's perspective.

International Journal of Sports Science & Coaching, *14*(1), 63–71. https://doi.org/10.1177/1747954118808081

Pantzalis, V. C., & Tjortjis, C. (2020). Sports Analytics for Football League Table and Player Performance Prediction. *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA*, 1–8. https://doi.org/10.1109/IISA50023.2020.9284352

Peulić, J., Obradović, A., Vukadinović Jurišić, M., & Obradović, J. (2023). The influence of anthropometric characteristics on swimming speed in adolescent swimmers. *Exercise and Quality of Life*, *15*(2), 33–40. https://doi.org/10.31382/eqol.231204

Pourrahim Ghouroghchi, A., & Pahlevani, M. (2020). The investigation of relationship between anthropometrical and physiological parameters of elite young boys in breaststroke and butterfly swimming. *International Journal of Pediatrics*, *Online First*. https://doi.org/10.22038/ijp.2020.53077.4207

Powell, C., Pyne, D. B., Crowley, E., & Mujika, I. (2025). What It Takes to Win: Examining Predicted Versus Actual Swimming Performances at the Paris 2024 Olympic Games, and What Comes Next. *International Journal of Sports Physiology and Performance*, *20*(4), 504–514. https://doi.org/10.1123/ijspp.2024-0409

Puel, F., Hellard, P., Pyne, D. B., & Morlier, J. (2023). Technical Skills Influences on Front Crawl Tumble Turn Performance in Elite Female Swimmers. *Journal of Sports Science and Medicine*, 571–581. https://doi.org/10.52082/jssm.2023.571

Qi, Y., Sun, K., Zhao, D., Liu, L., & Zhang, S. (2023). Kinematic and electromyography characteristics of performing butterfly stroke with different swimming speeds in flow environment. *Heliyon*, *9*(9), e20122. https://doi.org/10.1016/j.heliyon.2023.e20122

Sarcevic, A., Vranic, M., Pintar, D., & Krajna, A. (2022). Predictive modeling of tennis matches: A review. *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1099–1104. https://doi.org/10.23919/MIPRO55190.2022.9803645

Sarlis, V., & Tjortjis, C. (2020). Sports analytics—Evaluation of basketball players and team performance. *Information Systems*, *93*, 101562. https://doi.org/10.1016/j.is.2020.101562

Seifert, L., Boulesteix, L., Chollet, D., & Vilas-Boas, J. P. (2008). Differences in spatial-temporal parameters and arm–leg coordination in butterfly stroke as a function of race pace, skill and gender. *Human Movement Science*, *27*(1), 96–111. https://doi.org/10.1016/j.humov.2007.08.001

Soligard, T., Schwellnus, M., Alonso, J.-M., Bahr, R., Clarsen, B., Dijkstra, H. P., Gabbett, T., Gleeson, M., Hägglund, M., Hutchinson, M. R., Janse Van Rensburg, C., Khan, K. M., Meeusen, R., Orchard, J. W., Pluim, B. M., Raftery, M., Budgett, R., & Engebretsen, L. (2016). How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British Journal of Sports Medicine*, *50*(17), 1030–1041. https://doi.org/10.1136/bjsports-2016-096581

Staunton, C. A., Romann, M., Björklund, G., & Born, D.-P. (2024). 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

Stosic, J., Veiga, S., Trinidad, A., & Navarro, E. (2020). How Should the Transition from Underwater to Surface Swimming Be Performed by Competitive Swimmers? *Applied Sciences*, *11*(1), 122. https://doi.org/10.3390/app11010122

Strzała, M., Stanula, A., Krężałek, P., Ostrowski, A., Kaca, M., & Głąb, G. (2017). Butterfly Sprint Swimming Technique, Analysis of Somatic and Spatial-Temporal Coordination Variables. *Journal of Human Kinetics*, *60*, 51–62. https://doi.org/10.1515/hukin-2017-0089

Tharatipyakul, A., Srikaewsiew, T., & Pongnumkul, S. (2024). Deep learning-based human body pose estimation in providing feedback for physical movement: A review. *Heliyon*, *10*(17), e36589. https://doi.org/10.1016/j.heliyon.2024.e36589

Toussaint, H. M. (2011). Biomechanics of drag and propulsion in front crawl swimming. In *World book of swimming: From science to performance* (pp. 3–20). Nova Science Publishers.

Veiga, S., Lorenzo, J., Trinidad, A., Pla, R., Fallas-Campos, A., & De La Rubia, A. (2022). Kinematic Analysis of the Underwater Undulatory Swimming Cycle: A Systematic and Synthetic Review. *International Journal of Environmental Research and Public Health*, *19*(19), 12196. https://doi.org/10.3390/ijerph191912196

Wanivenhaus, F., Fox, A. J. S., Chaudhury, S., & Rodeo, S. A. (2012). Epidemiology of injuries and prevention strategies in competitive swimmers. *Sports Health*, *4*(3), 246–251. https://doi.org/10.1177/1941738112442132

Yu, Z. (2024). Analysis of the prospective application of artificial intelligence in swimming. *Applied and Computational Engineering*, *37*(1), 68–73. https://doi.org/10.54254/2755-2721/37/20230472

Yu-Hung Hsu, Y.-H. H., & Yu-Hung Hsu, C.-H. L. (2024). Using Machine-Learning Technology to Implement a Nonhardware and Inexpensive Posture Detection System for Analyzing Body Posture Angles in Front Crawl Swimming. 電腦學刊, *35*(1), 131–143. https://doi.org/10.53106/199115992024023501009

License Agreement on accession of
Open access archives (repository) to
the autonomous organization of education Nazarbayev University

1. General provisions

1.1. Autonomous Organization of education Nazarbayev University (hereinafter - University) offers to an individual (hereinafter – the Licensor) to use Open access archives (repository) of the University (hereinafter - Nazarbayev University Repository) on the terms set out in this license agreement on accession (hereinafter – the Agreement). The Agreement shall enter into force from the moment of expression of consent by Licensor to its terms in the manner provided in § 1.4 hereof.

1.2. The University offers to Licensor access to Nazarbayev University Repository with the aim of placement (self-archiving) of storage facilities. All Nazarbayev University services associated with Nazarbayev University Repository are the subject of herein.

1.3. The use of the Nazarbayev University Repository functions is governed by the Regulations on Nazarbayev University Repository, which is available on the Repository web site (http://nur.nu.edu.kz), and this Agreement.

1.4. Starting to use any function of Nazarbayev University Repository on placement (self-archiving) of storage facilities, Licensor is deemed to have accepted the terms of herein in full, without reservations and exceptions. In case of disagreement of the Licensor with provision hereof, he has a right not to place his publication in the Nazarbayev University Repository. If the University made changes hereto in the manner prescribed by section 4.1 hereof, with which Licensor does not agree, he has the right to terminate the Agreement with Nazarbayev University.

2. Licensor Account

2.1. In order to use the functions of Nazarbayev University Repository for placement (self-archiving), storage facilities, the Licensor shall be logged by ID card or login issued by the responsible subdivision of the University.

2.2. All actions upon usage of Nazarbayev University Repository under the account of the Licensor are considered as made by the Licensor himself, except for when the Licensor, in the manner prescribed by section 2.3 hereof, has notified the University about unauthorized access to his account, and (or) a privacy password of his account.

2.3. The Licensor shall immediately notify the University about any unauthorized, that is not permitted by the Licensor, access to the functions of Nazarbayev University Repository with the use of User's account and (or) a privacy password of his account.

2.4. The Licensor may not reproduce, copy, sell and use for any commercial purposes any material of Nazarbayev University Repository, including information about storage facilities available to the Licensor, except when the Licensor has received such permission from the University.

3. Terms of usage of Nazarbayev University Repository

3.1. The Licensor shall be solely liable to third parties for his actions related to the use of functions of Nazarbayev University Repository including, if such action would lead to a violation of the rights and legitimate interests of third parties, as well as compliance with the law when using the Nazarbayev University Repository.

3.2. The Licensor shall post publication in Nazarbayev University Repository, while reserving the right of independent use of full text of electronic copies of publications.

3.3. The Licensor shall provide to the University a non-exclusive right to placement, backup and storage of scientific publications for non-commercial purposes with the following levels of access for the users of Nazarbayev University Repository (tick the necessary one):

☑ in open access, with the ability to download full text or media file.

☐ in gated access, with the ability to download full text or media file only for Nazarbayev University community.

☐ in open access from **Year** ______ **Month** ______ **Day** ______

with the ability to download a full-text or multimedia file (after the specified date the research material will be automatically placed in open access for all users).

3.4. The Licensor shall have the right:

3.4.1. to maintain his/her personal page;

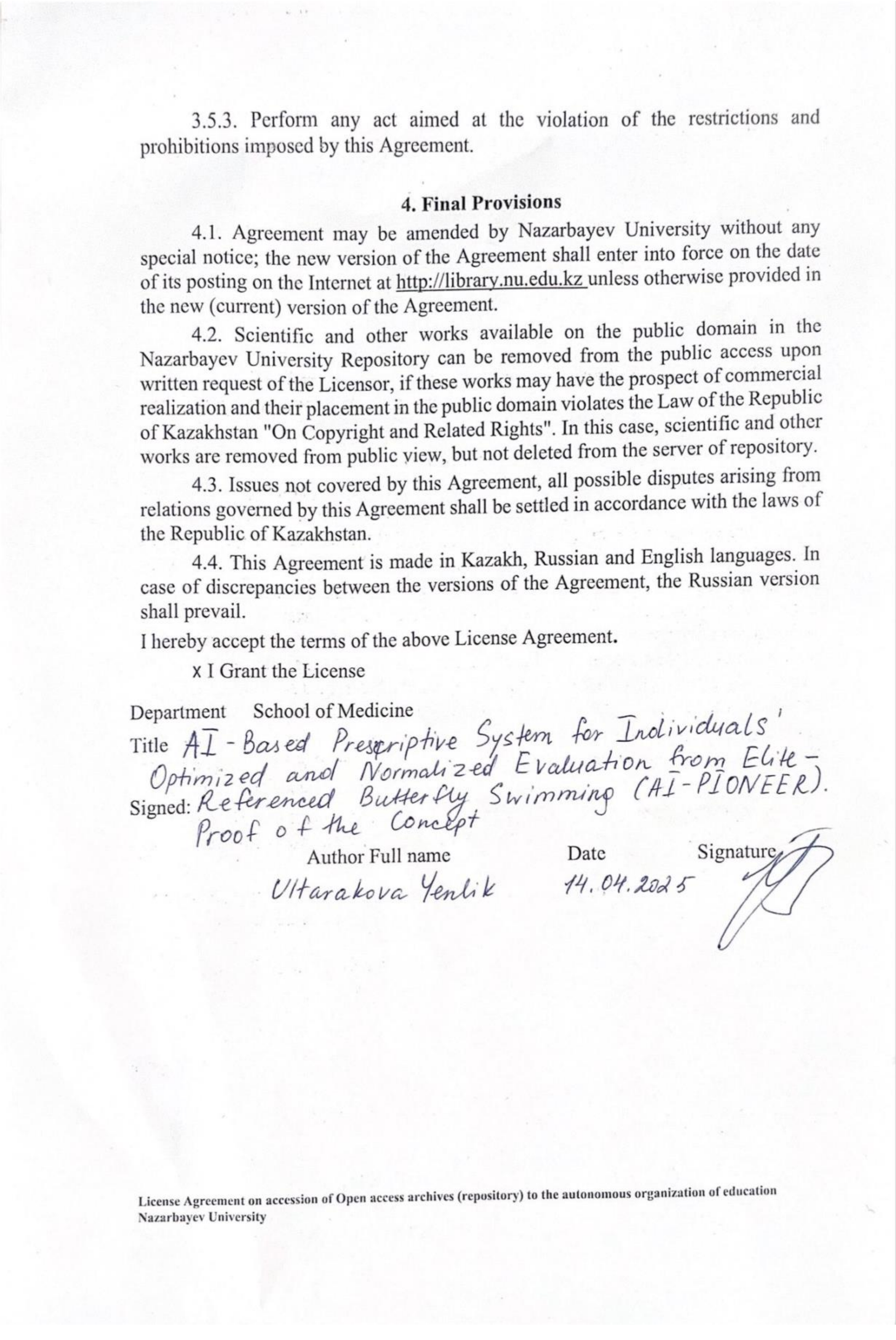
3.4.2. to add documents within his/her personal pages by self-archiving;

3.4.3. to use other rights in accordance with this Agreement.

3.5. When using Nazarbayev University Repository, the Licensor shall be prohibited to:

3.5.1. Upload, post, and (or) distribute documents and materials that are unlawful, harmful, defamatory, promoting violence and cruelty, hatred or discrimination against people on racial, ethnic, sexual, religious or social sign, containing insults against any individuals or entities, containing elements or are propaganda of pornography, child erotica, constitutes advertising of goods and services, explain the procedure of manufacture, application or other use of narcotic and psychotropic substances or their analogues, explosives or other weapons;

3.5.2. Upload, post, and (or) distribute works of science in the absence of rights on such actions in accordance with the current legislation of the Republic of Kazakhstan;



3.5.3. Perform any act aimed at the violation of the restrictions and prohibitions imposed by this Agreement.

4. Final Provisions

4.1. Agreement may be amended by Nazarbayev University without any special notice; the new version of the Agreement shall enter into force on the date of its posting on the Internet at http://library.nu.edu.kz unless otherwise provided in the new (current) version of the Agreement.

4.2. Scientific and other works available on the public domain in the Nazarbayev University Repository can be removed from the public access upon written request of the Licensor, if these works may have the prospect of commercial realization and their placement in the public domain violates the Law of the Republic of Kazakhstan "On Copyright and Related Rights". In this case, scientific and other works are removed from public view, but not deleted from the server of repository.

4.3. Issues not covered by this Agreement, all possible disputes arising from relations governed by this Agreement shall be settled in accordance with the laws of the Republic of Kazakhstan.

4.4. This Agreement is made in Kazakh, Russian and English languages. In case of discrepancies between the versions of the Agreement, the Russian version shall prevail.

I hereby accept the terms of the above License Agreement.

x I Grant the License

Department School of Medicine

Title AI - Based Prescriptive System for Individuals' Optimized and Normalized Evaluation from Elite-Referenced Butterfly Swimming (AI-PIONEER). Proof of the Concept

Signed:

Author Full name	Date	Signature
Ultarakova Yenlik	14.04.2025	