Asilbek Abdugabbarov
Navoi, Republic of Uzbekistan
abdugabbarovasilbek@gmail.com

DEVELOPMENT OF AN AI-SUPPORTED TRAINING METHODOLOGY FOR TRIATHLETES

Abstract. *The purpose of this study was to develop a scientifically grounded methodology for managing triathletes' training in which artificial intelligence (AI) is used as a decision-support tool for the coach. The study followed a design-oriented methodological approach. Scientific literature on endurance monitoring, training load, heart rate variability, wearable technology, and machine learning was analysed; data requirements were defined; an algorithm for daily readiness assessment and weekly load adjustment was developed; and safety and pedagogical-control criteria were formulated. The resulting methodology integrates swimming, cycling, and running data with heart rate, heart rate variability, sleep, perceived exertion, and subjective well-being. A closed loop is proposed: data acquisition, quality control, state assessment, recommendation, coach decision, and feedback. AI classifies the athlete's state, identifies atypical changes, and proposes adjustments to volume, intensity, and recovery. Final authority, however, remains with the coach and, where appropriate, a medical professional. The practical value of the methodology lies in its potential use by sports schools, clubs, and individually coached athletes. A controlled trial using real athlete data is required to determine its effectiveness.*

Keywords: *triathlon, artificial intelligence, machine learning, training load, individualisation, endurance, wearable devices, recovery.*

Introduction

Triathlon combines three cyclic disciplines—swimming, cycling, and running—and therefore places distinctive demands on load planning. A coach must coordinate the development of aerobic power, threshold performance, technique, strength endurance, and transition ability while limiting accumulated fatigue. The same training programme may elicit different responses because athletes differ in fitness, sleep, psychological stress, nutrition, environmental exposure, and injury history.

The spread of GPS devices, heart-rate monitors, power meters, and sleep sensors has enabled continuous collection of large volumes of data. Data availability alone, however, does not ensure sound decisions. Devices differ in accuracy, some metrics remain experimental, and every measurement must be interpreted in the athlete's context [1, 2]. Current research demonstrates the potential of machine learning for analysing endurance physiology, predicting training response, and identifying physiological thresholds [3–5]. Nevertheless, a complex model does not necessarily outperform an interpretable one when sample sizes are small or data quality is poor [3].

The research problem is the absence of an integrated methodology that combines the three triathlon disciplines, subjective recovery, and the coach's pedagogical judgement. Unrestricted use of AI may lead to automatic load increases, erroneous conclusions from isolated indicators, and reduced professional accountability. A human-in-the-loop model is therefore necessary: the algorithm should support rather than replace the coach [6].

The aim of the study was to develop the structure, content, and operating algorithm of an AI-supported training methodology for triathletes. The objectives were to: (1) identify the required input data; (2) develop a training analysis and adjustment cycle; (3) establish safety and interpretability rules; and (4) propose a design for subsequent experimental evaluation.

Materials and Methods

Study design. This study is a design-oriented methodological investigation consistent with design-science logic: a practical problem is defined from scientific evidence, a methodological artefact is developed, and a protocol for empirical evaluation is specified. No invented athlete outcomes are reported in this article. The Results section describes the developed methodology and measurable criteria for its future testing.

Evidence base. The methodology was informed by publications concerning machine learning in endurance physiology, wearable monitoring, heart rate variability (HRV), training-load management, and individualisation [2–10]. The following criteria guided selection of indicators: physiological relevance, feasibility of repeated measurement, interpretability for the coach, robustness to missing data, and protection of confidential information.

Recorded Variables

Domain	Core variables	Purpose
External load	Duration, distance, pace/speed, elevation gain, power output, number of accelerations	Characterises the work completed in each discipline
Internal load	Heart rate, time in zones, TRIMP, session-RPE × duration	Estimates the physiological and perceived cost of training
Recovery	Morning HRV, resting heart rate, sleep duration and quality, muscle soreness	Detects deviations from the athlete's individual baseline
Context	Temperature, humidity, illness, stress, travel, nutrition, pain	Prevents misleading interpretation of physiological signals
Performance	Standardised trials, threshold power/pace, economy, competition results	Assesses adaptation and supports goal adjustment

Note. The exact dataset depends on the available equipment. The methodology can operate with a minimum package consisting of a heart-rate monitor, GPS, an RPE scale, and a well-being diary.

Processing and Decision Algorithm

Before implementation, an individual baseline profile is established using 21–28 days of stable observations. For each indicator, the median, an acceptable range, and the direction of a meaningful individual deviation are calculated. This approach is preferable to universal thresholds because athletes may respond differently to the same load.

The daily module performs five operations: it checks data completeness and plausibility; compares current values with the individual baseline; calculates an integrated readiness index; determines a risk level; and produces an explainable recommendation. Logistic regression, a decision tree, or gradient boosting with a restricted feature set may be used as the initial model. Every recommendation must display its main determinants, for example: "HRV is 14% below the individual median, sleep was 75 minutes shorter, and yesterday's session-RPE was high."

The integrated score is not a medical diagnosis. It classifies the athlete into three zones: green—the planned session may proceed; amber—the load should be reduced or replaced with recovery work; red—high-intensity training should be cancelled and the athlete should undergo additional assessment by the coach and, when symptoms are present, a medical professional.

Structure of the Training Intervention

Component	AI function	Coach's decision
Microcycle planning	Compares goals, load history, and available time; proposes the distribution of swimming, cycling, running, strength work, and recovery	Approves goals, sequence, and compatibility of sessions
Daily adaptation	Assesses readiness and recommends maintaining, reducing, or replacing the planned load	Considers technique, symptoms, motivation, and context
Post-session analysis	Compares planned and completed work and flags anomalies	Interprets causes and provides pedagogical feedback
Weekly review	Displays load, monotony, recovery, and data-quality trends	Adjusts the next microcycle
Pre-competition phase	Supports tapering and freshness monitoring without abrupt changes	Determines the individual taper and race strategy

Proposed Evaluation Protocol

A 12-week controlled study involving at least 40 trained triathletes is recommended. Participants should be allocated to experimental and control groups while balancing sex, age, experience, and baseline performance. The control group would train under the coach's usual plan. The experimental group would follow the same periodisation principles with additional daily AI-supported decision-making. The primary outcome should be change in a standardised combined test or race performance. Secondary outcomes should include threshold pace or power, sessions missed because of fatigue, pain episodes, adherence, session-RPE, and recovery quality. Analysis should follow the intention-to-treat principle; between-group comparisons should adjust for baseline values and report confidence intervals and effect sizes.

Ethical requirements include informed consent, data minimisation, encryption, role-based access, the right to data deletion, and a prohibition on automated medical conclusions. Chest pain, syncope, marked dyspnoea, fever, acute injury, or persistent deterioration in well-being must always take priority over an algorithmic recommendation.

Results

The principal result of the study is an AI-supported triathlon training methodology comprising five interrelated modules: diagnostic, planning, monitoring, adjustment, and reflective-analytical.

Module	Input	Output
1. Diagnostics	Sport history, testing, intensity zones, limitations	Individual profile and initial objectives
2. Planning	Season goals, calendar, available time, athlete profile	Meso- and microcycles integrating the three disciplines
3. Monitoring	Device data, RPE, sleep, HRV, well-being	Validated daily dataset
4. Adjustment	Detected deviations and state prediction	Explainable recommendation and session options

5. Analysis	Plan completion, test results, and feedback	Updated profile and next training cycle

The cycle operates as follows. After each session, data are imported automatically and supplemented by a short athlete questionnaire. The system identifies missing or anomalous values; the model compares the current response with the athlete's history; the coach receives a readiness category, its reasons, and up to three action options; and the coach's decision together with the athlete's subsequent response is returned to the database. This creates a learning loop, although automatic model retraining is permitted only after expert validation.

The originality of the proposed methodology lies not in a single sensor or algorithm but in integrating triathlon's three modes of loading with subjective recovery and mandatory pedagogical oversight. Each discipline is analysed separately and in relation to total accumulated stress. For example, a demanding cycling session may require an adjustment to the running plan even when no running was performed on the previous day.

Implementation quality criteria were defined as follows: at least 90% of sessions should be imported correctly; each recommendation must be accompanied by an explanation; the coach must be able to override a recommendation and record the reason; the system must not automatically increase weekly volume beyond a predefined limit; and the model must be reviewed when predicted and observed states diverge systematically. These are implementation requirements, not claims of demonstrated clinical effectiveness.

Discussion

The proposed methodology reflects the contemporary shift from passive data accumulation to decision support. Research indicates that machine-learning models can classify training response and extract useful patterns from physiological indicators [3–5, 9]. Multisensor assessment may be more robust than any single indicator because fatigue is multidimensional [8]. Predictive accuracy and practical value nevertheless remain dependent on data quality, sample size, device compatibility, and stable measurement conditions.

A central strength of the methodology is interpretability: the coach receives both a category and the reasons behind it. This reduces the risk of blind reliance on an algorithm. Individual baselines also limit the error created by comparison with population averages. Recording occasions when the coach rejects the system's advice provides an additional benefit, as these decisions become valuable evidence about expert pedagogical judgement.

The methodology has several limitations. Wearable devices differ in accuracy, especially during swimming, rapid intensity changes, or poor sensor contact. Observed associations between load and deterioration do not prove causality. Small samples increase the risk of overfitting. Injury risk should not be inferred from a single biomechanical or physiological indicator because sports injuries are multifactorial [11]. Finally, the algorithm may reproduce systematic bias in its source data and therefore requires regular auditing.

Practical implementation should begin with a high-quality measurement protocol and transparent rules rather than a complex neural network. Once a representative dataset has accumulated, simple and advanced models may be compared in terms of discrimination, calibration, interpretability, and practical utility. The principal criterion of success should not be maximal algorithmic accuracy but improved training decisions without increased risk or burden for the athlete.

Conclusion

An AI-supported methodology for training triathletes was developed on the basis of integrated monitoring of external load, internal load, recovery, and contextual factors. Its defining principle is that the coach remains at the centre of decision-making. AI provides analysis, early detection of deviations, explainable recommendations, and personalisation, but it does not replace pedagogical or medical expertise.

The methodology may serve as the foundation of a digital platform for triathlon preparation. The next stage should be a 12-week controlled evaluation examining performance, load tolerance, adherence, and safety. Until empirical data are obtained, effectiveness must be treated as a testable hypothesis rather than an established outcome.

Practical Recommendations

1. Begin with a minimal reliable dataset rather than the maximum possible number of sensors.
2. Use individual baseline values and standardised conditions for morning measurements.
3. Show the coach which variables influenced every recommendation.
4. Do not change training volume, intensity, and frequency simultaneously without justification.
5. Introduce medical red flags that block automated training recommendations.
6. Review data quality weekly and audit the model monthly.

References

1. American College of Sports Medicine. ACSM Fitness Trends: Wearable Technology. Available at: https://acsm.org/education-resources/trending-topics-resources/acsm-fitness-trends/ (accessed 23 July 2026).

2. Sperlich B, Holmberg H-C, Aminian K. Wearable Sensor Technology for Monitoring Training Load and Health in the Athletic Population. Frontiers in Physiology Research Topic. Available at: https://www.frontiersin.org/research-topics/5516/

3. Boudry F, et al. The role of machine learning methods in physiological data analysis in endurance exercise. Front Sports Act Living. 2024;6:1440652. doi:10.3389/fspor.2024.1440652.

4. Tomaszewski M, et al. From data to decision: machine learning determination of aerobic and anaerobic thresholds in athletes. PLoS One. 2024;19:e0309427. doi:10.1371/journal.pone.0309427.

5. Barsumyan A, et al. Quantifying training response in cycling based on machine learning. Front Artif Intell. 2025;8:1623384. doi:10.3389/frai.2025.1623384.

6. Alkasasbeh WJ, et al. Artificial intelligence and wearables in sport: performance, injury risk and athlete wellbeing. Front Artif Intell. 2026;9:1838507. doi:10.3389/frai.2026.1838507.

7. Düking P, et al. Monitoring and adapting endurance training on the basis of heart rate variability monitored by wearable technologies: a systematic review with meta-analysis. J Sci Med Sport. 2021;24:1180–1192. doi:10.1016/j.jsams.2021.04.012.

8. Xu X, et al. Multi-sensor fusion outperforms single indicators for fatigue prediction in collegiate athletes. Front Physiol. 2026;17:1775906. doi:10.3389/fphys.2026.1775906.

9. Li M, et al. Physiological load estimation in athletes using ECG- and HRV-derived features and machine learning. Front Public Health. 2026;14:1843963. doi:10.3389/fpubh.2026.1843963.

10. Foucaud A, et al. Using unsupervised machine learning to characterize recovery states in Olympic-level athletes. Front Sports Act Living. 2025;7:1629924. doi:10.3389/fspor.2025.1629924.

11. American College of Sports Medicine. Vertical Loading Rate Should Not Be Used as a Predictor of Running Injury. 2022. Available at: https://acsm.org/vertical-loading-running-injury-predictor/

12. Cattle A, et al. Machine learning applications in sport: a scoping review. Front Psychol. 2026;17:1802549. doi:10.3389/fpsyg.2026.1802549.