

Injury Risk and Recovery: Machine Learning Applied to Biomechanical and Physiological Markers to Flag Overtraining and Injury Likelihood

Aron K. Parnell, MSc^{1*}, William A. Smith, MSc¹

¹ Department of Sports, University of Calgary, Calgary, AB, Canada

ABSTRACT

Introduction and Objective. Athlete monitoring has increasingly shifted from retrospective injury surveillance toward longitudinal, data-driven assessment of training tolerance, recovery and injury-related risk. The combination of wearable sensors, force-plate testing, movement analysis and physiological monitoring generates multimodal datasets that may contain early signals of altered recovery and changes in movement or workload patterns. This narrative review examines how machine-learning (ML) methods can integrate biomechanical and physiological markers to support injury-risk assessment, overtraining monitoring and recovery prediction. **Material and method.** The review was organized around three major predictive targets: non-contact musculoskeletal injury, overtraining or maladaptation, and short-term recovery status. Particular attention was given to ground-reaction forces, joint kinematics, movement asymmetry, heart-rate variability, resting heart rate, perceived exertion, sleep and internal- and external-load variables. The principal ML approaches considered were regularized regression, support vector machines, tree-based ensembles, and temporal models including recurrent neural networks. Model development, feature construction, class imbalance, athlete-level validation, calibration and interpretability were considered as key methodological requirements. **Results.** ML can combine multiple weak and correlated signals to identify patterns that may be difficult to detect using isolated thresholds. Wearable-derived workload and physiological data are particularly suitable for longitudinal monitoring, while force-plate and kinematic measurements provide complementary information about neuromuscular and mechanical changes. However, reported performance varies substantially between studies because of differences in sample size, injury definitions, outcome labels, sensor technology, feature engineering and validation procedures. Small athlete cohorts, repeated observations, class imbalance and limited external validation remain major sources of uncertainty. **Conclusions.** ML provides a useful decision-support framework for individualized injury and recovery monitoring, but predictive accuracy should not be interpreted as diagnostic validity. Robust implementation requires athlete-level or temporal validation, transparent preprocessing, calibrated predictions, external testing and integration of model output with clinical and coaching expertise.

Keywords: machine learning, injury prediction, athlete monitoring, overtraining, recovery

INTRODUCTION

Injury risk and recovery status are dynamic states influenced by the interaction between training exposure, physiological response, mechanical loading, sleep, psychological stress and individual characteristics. Unlike a static screening measurement, longitudinal athlete monitoring provides repeated observations that can describe how an athlete responds to training over time. This distinction is important because the same external workload can produce different internal and biomechanical responses between athletes and within the same athlete at different stages of a training cycle.

Traditional injury-prevention approaches frequently rely on periodic physical testing, clinical screening and workload thresholds. These approaches remain valuable, but they may provide only limited information about short-term changes occurring between testing sessions. Contemporary athlete-monitoring systems can collect heart rate, heart-rate variability (HRV), sleep, perceived exertion, GPS-derived workload, accelerometry and neuromuscular measurements repeatedly across training and competition. Force plates

and motion-capture systems can additionally characterize changes in ground-reaction forces, joint kinematics and movement asymmetry [1-4].

Training load is commonly divided into external and internal components. External load represents the physical work performed by an athlete, including distance, acceleration, deceleration, sprint exposure and sport-specific actions. Internal load represents the physiological and perceptual response to that work and may include heart rate, oxygen consumption, session rating of perceived exertion and derived training-load indices [5-8]. The relationship between these dimensions is individual and context dependent. Consequently, injury and recovery assessment cannot necessarily be reduced to a single universal workload threshold.

The training-injury prevention literature has emphasized the importance of considering both training exposure and athlete adaptation. Gabbett described the training-injury prevention paradox, while subsequent workload-injury models emphasized that injury emerges from interacting and changing constraints rather than from one isolated workload variable [9,10]. Complex-systems perspectives similarly conceptualize sports injury as the result of interactions among multiple intrinsic and extrinsic factors [11].

Machine learning provides a set of computational approaches for modelling such multivariable relationships. Supervised learning can estimate the probability of an injury-related outcome or predict a continuous recovery measure, while unsupervised learning can identify patterns or athlete subgroups without predefined labels. Tree-based ensembles can represent nonlinear interactions among heterogeneous variables, whereas temporal models can retain information about the order and timing of repeated observations [12,13].

The purpose of this review is therefore to examine how ML can integrate biomechanical and physiological markers for three related applications: injury-risk assessment, overtraining or maladaptation monitoring, and short-term recovery prediction. Particular attention is given to data preparation, model selection, validation, interpretability and the translation of predictive output into practical athlete-monitoring decisions.

MATERIAL AND METHOD

Review design. A narrative review design was used to integrate literature addressing machine learning, athlete monitoring, sports injury prediction, training load, recovery and biomechanical assessment. The review was structured around an applied analytical workflow: data acquisition, outcome definition, preprocessing, feature construction, model development, validation, interpretation and practitioner decision-making. Emphasis was placed on studies that used measurable biomechanical, physiological or training-load variables to address a defined sport-science outcome.

Predictive domains. Three predictive domains were considered. The first was injury risk, primarily non-contact musculoskeletal injury and related time-loss outcomes. The second was overtraining, maladaptation or persistent fatigue, recognizing that these conditions are more difficult to define than discrete injury events. The third was short-term recovery or readiness, including next-day perceived recovery, HRV response, neuromuscular status or other measurable indicators of recovery [12-16].

Biomechanical data. Biomechanical variables may be obtained from force plates, motion-capture systems, inertial measurement units and other movement-analysis technologies. Force-plate measurements can include peak force, impulse, rate of force development, braking and propulsive characteristics, jump height and inter-limb asymmetry. Motion-capture and inertial systems can provide joint angles, segment velocities, accelerations and movement trajectories. These measurements can be reduced to summary features or retained as time-dependent sequences depending on the analytical objective [2,3,17].

Physiological and training-load data. Physiological markers considered in this review included resting heart rate, HRV, heart-rate response to exercise, perceived exertion and sleep-related variables. External-load variables included distance, high-speed running, accelerations, decelerations, sprint exposure and sport-specific workload. Derived load measures, including acute and longer-term exposure summaries and established internal-load indices, may also be incorporated [5-8,18].

Machine-learning approaches. The principal model families were regularized regression, support vector

machines, random forests, gradient-boosting methods and temporal sequence models. Logistic or penalized regression provides an interpretable baseline for binary injury outcomes. Support vector machines can perform well with nonlinear decision boundaries in small datasets. Tree-based ensembles can capture nonlinear interactions among heterogeneous features, while recurrent or other sequence models may be appropriate when the temporal ordering of training and physiological observations is central to the prediction problem [12,13].

Outcome definition. Outcome definition is a critical component of model development. Injury prediction should specify injury type, severity, time horizon and exposure denominator. Recovery models should define whether the target is subjective readiness, HRV change, neuromuscular performance, return-to-training status or another measurable outcome. Overtraining models require particular caution because fatigue, functional overreaching, non-functional overreaching and overtraining syndrome are not interchangeable constructs. Ambiguous labels can reduce the validity of otherwise sophisticated models.

Data preprocessing. Raw sensor measurements require quality control before model development. Missing heart-rate observations, implausible GPS values, force-plate recording errors, marker loss and synchronization problems can introduce systematic noise. Filtering, normalization, imputation and feature scaling should be performed within the training portion of each validation procedure. Performing these operations on the entire dataset before splitting can allow information from the test set to influence the model and produce optimistic estimates of performance [12,13].

Feature construction. Raw measurements were conceptually grouped into acute, chronic, baseline-relative and temporal features. Examples include recent training load, rolling workload averages, change from individual baseline, HRV trends, resting heart-rate deviations, sleep duration, perceived exertion and changes in force-time characteristics. Biomechanical features may include peak force, impulse, rate of force development, jump height, contact time and asymmetry. Feature construction should remain physiologically interpretable whenever possible because transparent variables facilitate quality control and practitioner interpretation.

Validation principles. Validation should correspond to the intended application. If the goal is prediction in a new athlete, athlete-level splitting or leave-one-athlete-out validation is more appropriate than random row-level splitting. If the goal is forecasting a future state within an existing athlete, temporally ordered validation is more appropriate. Randomly distributing repeated observations from the same athlete across training and testing sets can produce leakage because the model may learn athlete-specific characteristics rather than a generalizable relationship [12,13].

Performance evaluation. Injury-classification models should not be evaluated using accuracy alone. Sensitivity, specificity, area under the receiver-operating-characteristic curve, precision-recall measures and calibration provide complementary information. Recovery and continuous-outcome models should report mean absolute error, root mean squared error and explained variance where appropriate. Confidence intervals or repeated-validation distributions should be reported to characterize uncertainty in model performance.

Class imbalance. Injury events are typically much less frequent than non-injury observations. This creates an imbalance problem in which a model can achieve apparently high accuracy by predicting that most observations will not be followed by injury. Class weighting, resampling and threshold adjustment can be considered, but these procedures should be performed only within the training data. Precision-recall measures and sensitivity to clinically meaningful injury events should be emphasized rather than accuracy alone.

RESULTS

Injury-risk assessment. Injury prediction is one of the most extensively studied applications of artificial intelligence in sports. Models commonly combine external workload, internal physiological response, previous injury history and athlete-specific characteristics. GPS-derived distance, sprint exposure, accelerations and decelerations can be integrated with heart rate and session-RPE to represent both the physical work performed and the athlete's response to that work [5-10].

Workload-derived features have frequently been investigated as predictors of injury. Acute and longer-term workload relationships can provide useful information about exposure, but they should not be interpreted as universal injury thresholds. The relationship between training load and injury is influenced by sport, training history, competition schedule, athlete characteristics and the quality of load measurement [9,10].

ML can extend conventional workload analysis by incorporating several variables simultaneously. For example, a model may combine recent external workload, HRV deviation, sleep reduction, previous injury, force-plate asymmetry and subjective fatigue. The resulting output can represent an estimated probability or classification rather than a single threshold. Such a model may identify combinations of moderate deviations that would not individually exceed a conventional screening threshold.

Biomechanical injury markers. Biomechanical measurements provide information about how an athlete produces and absorbs force. Force plates can characterize changes in ground-reaction-force profiles, impulse, rate of force development and asymmetry. Motion-capture and inertial systems can quantify changes in joint motion, coordination and movement strategy [2,17].

ML approaches can classify movement patterns, predict continuous biomechanical outcomes or identify deviations from an individual's baseline. A force-time curve can therefore be treated not only as a collection of isolated peaks but also as a temporal signal containing information about the entire movement. This may be particularly useful for monitoring changes in landing strategy, jump performance or neuromuscular function.

However, biomechanical predictions remain dependent on measurement quality. Changes in force-plate output may result from genuine alterations in movement, fatigue, technical factors, testing instructions or sensor-processing procedures. Model predictions should therefore be interpreted alongside the underlying force-time or kinematic data.

Physiological markers and recovery. Physiological monitoring provides a complementary perspective because it captures the athlete's response to training. HRV and resting heart rate are frequently used as indicators of autonomic state, while sleep, perceived exertion and wellness questionnaires provide additional information about recovery and subjective readiness.

ML can combine these measurements longitudinally rather than interpreting each observation independently. A decrease in HRV, increased resting heart rate, reduced sleep and elevated perceived exertion may each represent modest deviations from baseline, but their combination may provide stronger evidence of an altered recovery state. Individualized models can also determine whether a particular change is unusual for the athlete rather than simply unusual relative to a population average.

Recent explainable-ML work has demonstrated the potential for models to examine how training and match load relate to short-term HRV responses. Such approaches are relevant because they allow the model to describe which features contributed to a prediction rather than providing only a final numerical output.

Overtraining and maladaptation. Overtraining presents a different modelling problem from acute injury. An injury is usually associated with a defined event or clinical diagnosis, whereas overtraining-related conditions exist along a continuum of training stress, recovery and maladaptation. Consequently, model labels may be less reliable.

Potential features include persistent changes in HRV, resting heart rate, sleep, perceived fatigue, mood, training load, performance and endocrine markers. Sequence-based approaches may be appropriate because

the relevant signal may emerge from sustained changes over weeks rather than from a single abnormal measurement.

The distinction between transient fatigue and clinically meaningful maladaptation is critical. A short-term decrease in performance after a heavy training session should not automatically be classified as overtraining. Models should therefore be linked to clearly defined observable outcomes and should avoid treating a single physiological deviation as evidence of a chronic syndrome.

Recovery prediction. Recovery prediction generally represents a shorter prediction horizon than injury forecasting. The model may estimate next-day readiness, perceived recovery, HRV response or neuromuscular performance based on the previous training session and recent history.

Potential predictors include training duration, intensity, external workload, sleep duration, sleep quality, nutrition, subjective fatigue, HRV and previous recovery responses. Regression models can predict a continuous recovery score, whereas classification models can identify categories such as expected, reduced or substantially reduced recovery.

Individual variability is particularly important in this application. The same workload may produce very different responses between athletes. Consequently, models that include within-athlete change-from-baseline features may provide more useful information than models relying entirely on population-level norms.

Temporal modelling. Recovery and injury risk are inherently longitudinal. Sequence models such as LSTM networks can theoretically capture dependencies across previous training sessions and physiological observations. This allows the model to distinguish an isolated deviation from a persistent trend.

However, temporal models require sufficient data to estimate sequence relationships reliably. In small elite-sport cohorts, the number of athletes may be low even when the number of observations is large. A dataset containing thousands of sessions may therefore still provide limited independent information. Temporal deep learning should consequently be justified by the structure and quantity of the data rather than selected solely because it is more complex.

Tree-based and kernel methods. Tree-based ensembles remain attractive for many applied athlete-monitoring datasets because they can accommodate nonlinear relationships and interactions among heterogeneous variables. Random forests and gradient-boosting algorithms can combine workload, physiological and biomechanical features without requiring all variables to be on the same scale.

Support vector machines may also be useful in relatively small datasets, particularly when nonlinear boundaries are expected. However, reported performance from individual studies should not be generalized across sports because cohort size, outcome definition and validation procedures can differ substantially.

Individualized modelling. One of the principal opportunities for ML is the transition from population-level prediction toward individualized monitoring. A population model can establish broad relationships, but individual models can evaluate whether a new observation is unusual relative to the athlete's own historical distribution.

For example, a resting heart rate that is normal for one athlete may represent a substantial deviation for another. Similarly, a force-plate asymmetry that is common during one athlete's baseline testing may not carry the same meaning as the same numerical asymmetry in another athlete. Individualized baselines can therefore complement population-level reference values.

Multimodal data fusion. The combination of biomechanical and physiological variables is potentially more informative than either domain alone. Physiological data describe how the athlete responds to training, while biomechanical data describe how force and movement are produced.

A model may therefore distinguish between two athletes who completed similar external workloads but demonstrated different physiological and mechanical responses. One may show normal HRV and stable jump mechanics, while another may show reduced HRV, elevated perceived exertion and altered force-time characteristics. The purpose of ML is not to replace these individual measures but to model their interactions.

Explainable machine learning. Explainability is particularly important when model output influences training or medical decisions. SHAP and related approaches can estimate the contribution of individual

features to a prediction. For example, a model may indicate that recent high-speed running exposure, reduced sleep and a substantial deviation in HRV contributed to an elevated predicted risk. Such explanations should be interpreted carefully. A feature that contributes strongly to prediction is not necessarily a causal risk factor. Explanations are therefore most useful as prompts for further assessment rather than as definitive physiological conclusions.

Model calibration and decision thresholds. Discrimination and calibration represent different properties of a predictive model. A model can rank athletes correctly according to relative risk while producing poorly calibrated absolute probabilities. Calibration assessment is therefore particularly important when a probability is presented directly to coaches or clinicians.

Decision thresholds should also reflect the consequences of false-positive and false-negative classifications. A lower threshold may identify more athletes requiring review but can also increase unnecessary follow-up. The appropriate threshold depends on the decision context, available clinical resources and the consequences of missed cases.

Data leakage and repeated observations. Athlete-monitoring datasets frequently contain repeated measurements from the same individuals. If observations from the same athlete are randomly distributed between training and testing sets, the model can partially recognize the athlete's individual signature. This can lead to substantially inflated estimates of generalization performance.

Athlete-level splitting is therefore essential when the intended target is a new athlete. Temporal splitting is preferable when the objective is future prediction within the same athlete. The validation strategy should always match the intended deployment scenario.

External validation. External validation remains one of the principal weaknesses of sports ML research. A model trained using one professional team may encounter substantially different training methods, competition schedules, athlete characteristics and sensor technologies when applied to another team.

Validation in a separate team, season or competition environment provides stronger evidence of transportability than a random split of observations from the original dataset. Where independent data are unavailable, temporal validation can provide a more realistic test than random row-level cross-validation.

Practical deployment. The practical value of a predictive model depends on whether its output can be incorporated into the athlete-monitoring workflow. A model that requires several days of manual processing may be unsuitable for decisions about the following day's training. Conversely, a return-to-sport model may tolerate slower processing if it integrates detailed laboratory measurements.

Automated systems should include data-quality checks for missing sensors, implausible values, synchronization errors and unusual measurement conditions. A model should not automatically generate a high-risk alert when the underlying data are incomplete or technically unreliable.

Decision support rather than diagnosis. Injury-risk prediction should be interpreted as a screening or triage process rather than as diagnosis. A high-risk classification does not demonstrate that an injury will occur, and a low-risk classification cannot guarantee that an athlete is free from injury risk.

The most appropriate practical response to an elevated prediction is therefore additional assessment. Coaches and clinicians can review recent training exposure, athlete-reported symptoms, movement characteristics, recovery status and relevant medical history before determining an appropriate action.

Table 1. Common biomechanical and physiological markers used in ML-supported injury and recovery monitoring

Domain	Example variables	Potential analytical role
--------	-------------------	---------------------------

Force plates	Peak force, impulse, rate of force development, jump height	Neuromuscular status and movement changes
Motion capture	Joint angles, segment velocity, coordination	Technique and movement-pattern classification
Inertial sensors	Acceleration, angular velocity, movement frequency	Field-based movement monitoring
GPS	Distance, high-speed running, accelerations, decelerations	External workload
Heart rate	Mean HR, peak HR, HR zones	Internal workload
HRV	RMSSD and related indices	Recovery and autonomic monitoring
Resting HR	Morning or pre-session HR	Baseline deviation and recovery
sRPE	Session duration $\times$ perceived exertion	Internal training load
Sleep	Duration, quality, sleep disruption	Recovery status
Wellness	Fatigue, soreness, stress, mood	Subjective recovery
Injury history	Previous injury, location, time since injury	Individual risk context

Table 2. Alignment between ML approaches and athlete-monitoring applications

ML approach	Typical target	Main advantage	Principal limitation
Logistic/regularized regression	Injury or recovery classification	Interpretability and simple baseline	Limited nonlinear representation
Support vector machine	Injury classification	Effective in some small datasets	Parameter sensitivity and limited transparency
Random forest	Risk classification	Nonlinear interactions and heterogeneous features	Probability calibration may require additional assessment
Gradient boosting	Risk and recovery prediction	Strong tabular prediction capability	Greater tuning and interpretability requirements
LSTM/recurrent models	Longitudinal recovery or fatigue	Temporal representation	High data requirements and complexity
Unsupervised learning	Athlete/movement subgrouping	Detects latent patterns	Clinical meaning may be uncertain
Explainable ML	Interpretation of predictions	Identifies influential features	Attribution does not establish causality

DISCUSSION

The central finding of this review is that ML can extend conventional athlete monitoring by combining biomechanical, physiological and workload variables across time. Its principal contribution is not the replacement of established sports-science measures but the ability to integrate multiple measurements and identify patterns that may be difficult to detect using isolated thresholds.

The strongest methodological limitation is the mismatch between the apparent size of athlete-monitoring datasets and the number of independent athletes. Thousands of training sessions can create a large table of observations, but the number of independent individuals may remain small. Treating every session as an independent observation can therefore exaggerate model performance and underestimate uncertainty.

Validation must consequently reflect the intended deployment scenario. When the target is a new athlete, all observations from that athlete should remain outside the training set. When the target is future prediction in an existing athlete, temporal validation provides a more appropriate test. This distinction is particularly important for individualized injury and recovery systems.

A second challenge concerns the definition of injury and recovery outcomes. Injury is not a single homogeneous event. Models may predict any injury, time-loss injury, muscle injury, lower-limb injury or injury within a specified number of days. These outcomes should not be combined without considering their different mechanisms and base rates.

Recovery presents an even greater measurement problem. Perceived recovery, HRV, sleep quality and neuromuscular performance represent different dimensions of the athlete's state. A model predicting one of these variables should not automatically be described as predicting “recovery” in its entirety. Clear outcome definitions are therefore essential for scientific interpretation.

The third major challenge is the use of workload metrics. Measures such as acute-to-chronic workload ratios have received considerable attention, but workload-injury relationships are complex and context dependent [9,10]. A predictive model that includes workload variables should therefore be evaluated against simpler baselines and should not be interpreted as establishing a universal causal threshold.

Biomechanical variables may provide an important complementary signal. Force-time curves and movement trajectories can capture mechanical changes that are not visible in workload summaries. However, these measurements are also sensitive to testing conditions, equipment and processing decisions. Standardized acquisition and preprocessing are therefore necessary before ML can reliably distinguish physiological change from measurement noise.

Physiological markers provide another complementary domain. HRV, resting heart rate, sleep and perceived exertion can describe the athlete's internal response to training. Their interpretation is strengthened when changes are evaluated longitudinally and relative to the athlete's own baseline. A single abnormal value may be less informative than a persistent multivariable pattern.

The use of explainable ML is particularly relevant to sports medicine. A prediction without an explanation may be difficult for practitioners to trust or act upon. However, explainability should not be confused with causality. SHAP values or feature-importance measures indicate how variables contributed to a model prediction; they do not prove that changing the variable would change the outcome.

Individual versus population models. Population-level models remain useful when individual data are limited, but elite sport increasingly provides the opportunity to construct personalized baselines. A practical framework may therefore combine both approaches: a population model provides initial expectations, while accumulating athlete-specific observations allow predictions to become increasingly individualized.

Recovery as a multidimensional state. Recovery should not be represented as a single physiological variable whenever the available data allow a multidimensional approach. Sleep, HRV, perceived fatigue, neuromuscular performance and recent workload may each provide different information. ML can combine these signals, but the target must still be defined clearly. A model predicting next-day HRV should not be interpreted as a comprehensive model of psychological or physiological recovery.

Overtraining versus short-term fatigue. A major conceptual issue is the distinction between normal

responses to demanding training and pathological or maladaptive states. A temporary decline in performance following a heavy training block may represent an expected training response. Persistent performance deterioration accompanied by altered physiological and psychological markers represents a different problem. ML models should therefore use longitudinal outcomes and clinically meaningful definitions rather than treating any deviation from baseline as overtraining.

Injury forecasting and class imbalance. Injury events are generally rare compared with non-injury observations. This means that conventional accuracy can be misleading. A model that predicts “no injury” for nearly every observation may obtain a high accuracy despite providing little practical value. Precision, recall, sensitivity, specificity, calibration and decision-curve or utility-based analyses may provide a more informative assessment.

Model complexity and sample size. Deep learning and sequence models can be attractive because athlete-monitoring data are inherently temporal. However, model complexity must be justified by sample size and outcome availability. A simpler regularized model may be more reliable when only a small number of athletes and injury events are available. Complexity should therefore be added when it provides reproducible out-of-sample improvement rather than because it is technologically sophisticated.

Comparison with conventional statistics. ML and conventional statistical methods should be considered complementary. Classical models remain valuable for estimating associations, testing hypotheses and describing effect sizes. ML is particularly useful when the primary objective is prediction, nonlinear interaction modelling or individualized classification. A strong study can therefore include both: conventional statistical analysis for inference and ML for out-of-sample prediction.

External validity and transportability. A model trained in one sport or team should not automatically be transferred to another. Differences in athlete level, competition schedule, tactical role, training philosophy, sensor hardware and data-processing pipelines can alter feature distributions. External validation is therefore essential before a model is deployed outside its development environment.

Model drift. Athlete-monitoring environments change over time. Sensor replacement, changes in GPS firmware, modifications to training methodology, changes in competition density and athlete rehabilitation can all alter the statistical distribution of input variables. Models should therefore be monitored for performance drift and recalibrated when appropriate.

Clinical and coaching integration. The most defensible implementation is a decision-support workflow. The model identifies an unusual pattern; the practitioner reviews the evidence; the athlete's context is considered; and an appropriate assessment or training modification is then considered. This preserves the advantages of computational pattern recognition while retaining human interpretation.

Measurement quality as a prerequisite. ML cannot compensate for systematically poor measurements. If force-plate trials are inconsistent, HRV measurements are collected under different conditions, or GPS data contain substantial missingness, the model may learn technical artefacts rather than athlete characteristics. Data-quality control should therefore be treated as part of the model itself rather than as a separate technical issue.

Actionability of predictions. Predictive performance should ultimately be connected to a decision. A model may improve an AUC statistic without changing how coaches or clinicians act. Conversely, a moderately discriminative model may have substantial practical value if it reliably identifies athletes who warrant additional assessment. Evaluation should therefore include not only statistical performance but also stability, calibration, interpretability and potential effect on decision-making.

Recommended reporting structure. A reproducible sports ML study should report the number of athletes, number of observations, sport and competition level, sensor technology, sampling frequency, preprocessing procedures, feature definitions, missing-data strategy, outcome definitions, model architecture, hyperparameters, data-splitting strategy and evaluation metrics. The paper should also specify whether the intended generalization is to a new athlete, future sessions within the same athlete, another team or another sport.

Future research directions. Future research should prioritize multi-team datasets, standardized injury

definitions, longitudinal individualized modelling and independent external validation. Studies should also compare complex ML approaches against transparent statistical baselines and report whether improvements in predictive performance translate into meaningful changes in athlete-management decisions. Multimodal systems combining workload, physiology and biomechanics represent a particularly relevant direction because they more closely reflect the multidimensional nature of training tolerance and injury risk.

CONCLUSIONS

Machine learning can extend injury-risk and recovery monitoring by integrating biomechanical, physiological and workload information across repeated observations. Force-plate measurements, movement kinematics, wearable sensors, heart-rate variability, resting heart rate, sleep, perceived exertion and training-load variables provide complementary information about how an athlete responds to training.

The principal advantage of ML is its ability to model nonlinear interactions, longitudinal patterns and individualized deviations from baseline. However, predictive accuracy alone does not establish clinical validity or causal relationships. Small athlete cohorts, class imbalance, repeated observations, inconsistent outcome definitions, data leakage and limited external validation remain major methodological challenges.

The most credible applications therefore use athlete-level or temporal validation, transparent preprocessing, appropriate calibration, interpretable outputs and clearly defined decision contexts. Injury-risk models should function as triage and decision-support tools rather than autonomous diagnostic systems, while recovery and overtraining models should distinguish normal short-term fatigue from persistent maladaptation.

Future progress will depend not only on more sophisticated algorithms but also on better measurement, larger multi-team datasets, standardized definitions and stronger validation. The practical value of ML in sports medicine will ultimately be determined by whether it improves the consistency, transparency and usefulness of decisions concerning training load, recovery, injury prevention and return to sport.

REFERENCES

1. Bartlett R, Wheat J, Robins M. Is movement variability important for sports biomechanists? *Sports Biomech.* 2007;6(2):224-243. <https://doi.org/10.1080/14763140701322994>
2. Alahaidib R, et al. Biomechanical assessment tools for injury risk prediction and return-to-sport evaluation in athletes: a systematic review. *Cureus.* 2025;17(9):e93210. <https://doi.org/10.7759/cureus.93210>
3. Seshadri DR, Li RT, Voos JE, et al. Wearable sensors for monitoring the physiological and biochemical profile of the athlete. *NPJ Digit Med.* 2019;2:72. <https://doi.org/10.1038/s41746-019-0150-9>
4. Halson SL. Monitoring training load to understand fatigue in athletes. *Sports Med.* 2014;44(Suppl 2):S139-S147. <https://doi.org/10.1007/s40279-014-0253-z>
5. Impellizzeri FM, Marcora SM, Coutts AJ. Internal and external training load: 15 years on. *Int J Sports Physiol Perform.* 2019;14(2):270-273. <https://doi.org/10.1123/ijsp.2018-0935>
6. Bourdon PC, Cardinale M, Murray A, et al. Monitoring athlete training loads: consensus statement. *Int J Sports Physiol Perform.* 2017;12(Suppl 2):S2-161-S2-170. <https://doi.org/10.1123/IJSP.2017-0208>
7. Djaoui L, Haddad M, Chamari K, et al. Monitoring training load and fatigue in soccer players with physiological markers. *Physiol Behav.* 2017;181:86-94. <https://doi.org/10.1016/j.physbeh.2017.09.004>
8. Claudino JG, Capanema DO, de Souza TV, Serrão JC, Machado Pereira AC, Nassis GP. Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Med Open.* 2019;5:28. <https://doi.org/10.1186/s40798-019-0202-3>
9. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med.* 2016;50(5):273-280. <https://doi.org/10.1136/bjsports-2015-095788>
10. Windt J, Gabbett TJ. How do training and competition workloads relate to injury? The workload-injury aetiology model. *Br J Sports Med.* 2017;51(5):428-435. <https://doi.org/10.1136/bjsports-2016-096040>
11. Bittencourt NFN, Meeuwisse WH, Mendonca LD, et al. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition. *Br J Sports Med.* 2016;50(21):1309-1314. <https://doi.org/10.1136/bjsports-2015-095850>

12. Halilaj E, Rajagopal A, Fiterau M, Hicks JL, Hastie TJ, Delp SL. Machine learning in human movement biomechanics: best practices, common pitfalls, and new opportunities. *J Biomech*. 2018;81:1-11. <https://doi.org/10.1016/j.jbiomech.2018.09.009>
13. Van Eetvelde H, Mendonça LD, Ley C, Seil R, Tischer T. Machine learning methods in sport injury prediction and prevention: a systematic review. *J Exp Orthop*. 2021;8:27.
14. Bowen L, Gross AS, Gimpel M, Bruce-Low S, Li FX. Spikes in acute:chronic workload ratio associated with a 5-7 times greater injury rate in English Premier League football players: a comprehensive 3-year study. *Br J Sports Med*. 2020;54:731-738. <https://doi.org/10.1136/bjsports-2018-099422>
15. Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernandez J, Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One*. 2018;13(7):e0201264. <https://doi.org/10.1371/journal.pone.0201264>
16. Abruñedo-Lombardero J, Padrón-Cabo A, Vélez-Serrano D, Álvaro-Meca A, Iglesias-Soler E. An explainable machine learning approach to explain the effects of training and match load on ultra-short-term heart rate variability in semi-professional basketball players. *Sensors*. 2025;25(22):6928. <https://doi.org/10.3390/s25226928>
17. Bartlett R, Wheat J, Robins M. Is movement variability important for sports biomechanists? *Sports Biomech*. 2007;6(2):224-243. <https://doi.org/10.1080/14763140701322994>
18. Silva P, Santos ED, Grishin M, et al. Validity of heart rate-based indices to measure training load and intensity in elite football players. *J Strength Cond Res*. 2018;32(8):2340-2347. <https://doi.org/10.1519/JSC.0000000000002057>
19. Ellis M, Penny R, Wright B, et al. The dose-response relationship between training-load measures and aerobic fitness in elite academy soccer players. *Sci Med Footb*. 2021;5(2):128-136. <https://doi.org/10.1080/24733938.2020.1817536>
20. Alexiou H, Coutts AJ. A comparison of methods used for quantifying internal training load in women soccer players. *Int J Sports Physiol Perform*. 2008;3(3):320-330. <https://doi.org/10.1123/ijspp.3.3.320>
21. Casamichana D, Castellano J, Calleja-Gonzalez J, et al. Relationship between indicators of training load in soccer players. *J Strength Cond Res*. 2013;27:369-374.
22. Rothschild JA, et al. Machine learning approaches for individualized prediction of recovery and physiological responses in athletes. 2024.
23. Iturricastillo A, Yanci J, Granados C, et al. Quantifying wheelchair basketball match load: a comparison of heart-rate and perceived-exertion methods. *Int J Sports Physiol Perform*. 2016;11(4):508-514. <https://doi.org/10.1123/ijspp.2015-0257>
24. Aoki M, Ronda L, Marcelino P, et al. Monitoring training loads in professional basketball players engaged in a periodized training program. *J Strength Cond Res*. 2017;31(2):348-358. <https://doi.org/10.1519/JSC.0000000000001507>