

Original research article DOI:

INTEGRATION OF PHYSIOLOGY, DIGITAL TWIN AND ARTIFICIAL INTELLIGENCE IN SPORT: A SYSTEMATIC REVIEW, CONCEPTUAL MODEL AND THE ROLE OF PHYSIOLOGISTS

Vojkan Nestorović[1], Anja Nestorović[2]

[1]Faculty of Medicine, University of Priština in Kosovska Mitrovica
[2]Health Center "Medicus Universalis" Kruševac

ABSTRACT

Digital twin (DT) and artificial intelligence (AI) are becoming the foundation of modern sports engineering and medicine. The aim of this paper is to systematically review the existing literature (2021–2026) on the application of DT and AI in sport, map the role of physiologists in the development of these models, and propose a conceptual framework integrating DNA, microbiome, nutrition, psychophysiological data and the digital patient–athlete. According to the PRISMA methodology, 120 papers were identified, of which 46 were analyzed in detail after risk of bias assessment using the ROBINS-I tool. The paper highlights serious methodological shortcomings in the literature, quantifies the distribution of effects, and concludes that the level of evidence is low. The role of the physiologist is crucial but systemically neglected.

Key words: digital twin, artificial intelligence, sports physiology, athlete monitoring, personalized sports medicine

Corresponding author
Vojkan Nestorović
vojkan.nestorovic@pr.ac.rs

INTRODUCTION

Digital twins and artificial intelligence are increasingly being explored as tools for athlete monitoring, performance enhancement, injury prevention, and rehabilitation (Hliš et al., 2024). Previous reviews have examined digital twins in sport and selected applications of artificial intelligence, including real-time sensor-based analytics, injury-risk assessment, and performance prediction (Vec et al., 2024). However, the targeted literature search did not identify a systematic review specifically mapping the involvement of sport physiologists in the development of these technologies or critically appraising the methodological quality and validation of the underlying evidence. Given the rapid expansion of this field and the documented need for prospective evaluation of AI methods in sport, a systematic and critical synthesis is timely (Bjelica et al., 2025).

A digital twin is a dynamic virtual replica of a physical system that is continuously updated with data to predict real-world behavior. The concept of DT is used in sport for performance modelling, training optimization and injury risk assessment (Hliš et al., 2024). These technologies are inspired by industrial applications but are increasingly being introduced into sports science and medicine to enable personalized interventions for athletes (Lukač et al., 2022).

Artificial intelligence (AI) enables the analysis of large datasets, pattern recognition and predictive models that can be used in real time for decision making (Zhou et al., 2025). AI integrates data on athletes' physiological parameters, including heart rate, energy metabolism, movement biomechanics and other variables.

The introduction of DT in sport reveals that it is used not only for performance enhancement but also for strategic planning, injury prevention and personalized rehabilitation, thus creating a platform for data-driven sports science and medicine (Hliš et al., 2024).

Accordingly, the objective of this review was to systematically map the existing literature (2021–2026) on the application of digital twin and artificial intelligence technologies in sport, in order to define the role of physiologists in the development of these models and to propose a conceptual framework integrating genomic, microbiome, nutritional and psychophysiological data through AI within a digital patient–athlete model.

MATERIAL AND METHODS

Study design and protocol

This systematic review was not pre-registered in the PROSPERO database; the complete review protocol is available from the authors upon request. This review followed the PRISMA 2020 reporting guideline (Page et al., 2021; the completed checklist is provided in Appendix A.

Data sources and search strategy

A systematic search was performed in PubMed, Web of Science, ScienceDirect, SpringerLink and arXiv. The search period covered January 1, 2020 to March 31, 2026. The following combinations of keywords and Boolean operators were used:

("digital twin" AND sport) OR ("digital twin" AND athlete AND biomechanics) OR ("artificial intelligence" AND sport performance) OR ("digital twin" AND "artificial intelligence" AND sport).

Limits: English language, original research articles and systematic reviews. Manual searching of references (snowballing) was added.

Inclusion and exclusion criteria

Inclusion: focus on DT/AI in sport, human subjects, performance/injury/training/physiology, peer-reviewed journals or relevant preprints, full text available.

Exclusion: industrial applications without sport, no original data, duplicates, insufficient methodological transparency, animal studies.

Study selection process (PRISMA flow)

A total of 312 records were identified through the database search. Following the removal of 78 duplicates, 234 unique records remained for title and abstract screening, of which 142 were excluded. The remaining 92 articles underwent full-text assessment, resulting in the exclusion of 44 articles due to insufficient methodological transparency, an inappropriate study population, duplicate or secondary publications, or the absence of original data. Consequently, 48 studies were included in the risk-of-bias assessment using ROBINS-I. Following this assessment, two studies were excluded due to critical risk of bias, leaving a total of 46 studies for inclusion in the qualitative synthesis (Table 1).

Table 1. PRISMA 2020 Flow Diagram (tabular format)

Phase	Description	Number of records
Identification	Records from databases (PubMed, WoS, ScienceDirect, SpringerLink, arXiv)	312
	Additional records from snowballing	0
Deduplication	Duplicates removed	78
	Unique records for screening	234
Screening	Records screened (title/abstract)	234
	Excluded	142
Eligibility	Full-text articles assessed	92
	Full-text excluded	44
	– Insufficient methodological transparency	18
	– Inappropriate population	12
	– Duplicates/secondary publications	8
	– No original data	6

Phase	Description	Number of records
Inclusion	Studies for risk assessment (ROBINS-I)	48
	Excluded after ROBINS-I (critical risk)	2
	Studies included in synthesis	**46**

Data extraction

The following data were extracted: author, year, sport type, sample size, method (DT, AI, ML), main outcomes, key findings, reported effect (positive/neutral/negative).

Risk of bias assessment (ROBINS-I

The ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool was used to assess the risk of bias in the included observational studies (Sterne et al., 2016). Seven domains were assessed: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selective reporting. Each domain was rated as having low, moderate, serious, or critical risk of bias. Studies judged to have a critical risk of bias in any domain were excluded from the synthesis.

Certainty of evidence assessment (GRADE)

The certainty of evidence across the identified domains (digital twin, artificial intelligence, DT+AI integration, physiological models, genomics/microbiome, personalised nutrition) was assessed using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach. Five domains were evaluated for each outcome: risk of bias, inconsistency of results, indirectness of evidence, imprecision of estimates, and publication bias. As none of the included studies was a randomised controlled trial, the initial certainty level for all domains was set as low, in accordance with GRADE methodology for observational evidence, and was further downgraded where appropriate based on the findings from the domain-level assessment. Full details of the assessment and the Summary of Findings table are provided in Appendix C and C1.

Data synthesis

Due to heterogeneity, a qualitative synthesis was performed together with quantification of the distribution of effects (positive/neutral/negative). Meta-analysis was not performed.

RESULTS

After ROBINS-I assessment, 46 studies were included in the synthesis (two excluded due to critical risk). Based on the risk-of-bias assessment (ROBINS-I, Table

2 and the Appendix B), the majority of the final sample of 46 studies carried low to moderate risk of bias (12/46, 26.1% low; 24/46, 52.2% moderate), while a smaller proportion carried serious risk (10/46, 21.7%); the two studies rated as critical risk were excluded prior to synthesis.

Table 2. Summary of ROBINS-I risk-of-bias assessment in 48 studies before exclusion.

Risk	Number of studies	Percentage
Low	12	25.0%
Moderate	24	50.0%
Serious	10	20.8%
Critical	2	4.2%

Note: Two studies with critical risk (mainly due to uncontrolled confounding) were excluded from the synthesis. The remaining 46 studies were used for qualitative synthesis and quantification of effects.

Based on the GRADE assessment across the six identified domains, the overall certainty of evidence was rated as low to very low. Digital twin, artificial intelligence and physiological model domains were rated as low certainty, while DT+AI integration, genomics/microbiome and personalised nutrition domains were rated as very low certainty, mainly due to the absence of randomised controlled trials, substantial methodological heterogeneity, and limited external validation. Full domain-level ratings and the Summary of Findings table are presented in Appendix C and C1.

The studies covered cyclic, acyclic, and team sports and were published between 2020 and 2026, with sample sizes ranging from 30 to 210 participants. As summarised in Table 3, digital twin (DT) approaches were addressed in 30 studies, artificial intelligence (AI) approaches in 35 studies, combined DT+AI integration in 18 studies, physiological models in 27 studies, genomics/microbiome approaches in 6 studies, and personalised nutrition in 12 studies. Because several studies addressed more than one domain, these totals exceed the overall number of included studies (n = 46). The qualitative synthesis was based on categorical distributions of effect direction rather than pooled continuous outcomes; therefore, means and standard deviations were not applicable to the synthesis as a whole. Among the 10 representative studies, eight reported sample sizes, yielding a mean sample size of 104.6 ± 62.9 participants (range, 38–210), which illustrates the substantial variability in study scale across the literature. The distribution of reported effects across the six domains, categorised as positive, neutral or inconsistent, and negative or potentially harmful, is presented in Table 4.

Table 3. Characteristics of included studies (representative sample, n=10 of 46)

ID	Author, year	Sport type	N	Method	Main outcome	Effect	ROBINS-I risk
1	Hliš et al., 2024	Multiple sports	120	Digital twin	Performance	Positive	Low
2	Lukač et al., 2022	Cycling	45	Digital twin	Training efficiency	Positive	Moderate
3	Jiang et al., 2025	Sport dance	38	DT + AI	Movement biomechanics	Positive	Moderate

ID	Author, year	Sport type	N	Method	Main outcome	Effect	ROBINS-I risk
4	Zhou et al., 2025	Multiple sports	210	AI (ML/DL)	Injuries	Positive	Low
5	Su & Zhu, 2026	Swimming	89	ML + explainable AI	Performance	Neutral	Moderate
6	Kakavas et al., 2022	Sports medicine	60	Virtual twin	Clinical application	Neutral	Serious
7	Van Eetvelde et al., 2021	Multiple sports	180	Machine learning	Injuries	Positive	Low
8	Bartlett et al., 2020	Multiple sports	95	Physiological monitoring	Fatigue	Positive	Moderate
9	Glebova et al., 2023	Sport events	N/A	Digital venue twin	System performance	Neutral	Moderate
10	Li et al., 2025	Multiple sports	N/A	DT + AI (conceptual)	Personalisation	Neutral	Moderate

Note: The full table of all 46 studies is available upon request. Sample size was not reported for two of the ten representative studies (Glebova et al., 2023; Li et al., 2025); the mean ± SD sample size reported in Section 3 (104.6 ± 62.5) is based on the remaining eight studies with reported N and should be interpreted as descriptive of this representative subset rather than the full sample of 46 studies.

Table 4. Distribution of effects by domain

Domain	Number of studies	Positive effect n (%)	Neutral/inconsistent n (%)	Negative/harmful n (%)
Digital twin (DT)	30	18 (60.0%)	10 (33.3%)	2 (6.7%)
Artificial intelligence (AI)	35	22 (62.9%)	11 (31.4%)	2 (5.7%)
DT + AI integration	18	8 (44.4%)	9 (50.0%)	1 (5.6%)
Physiological models	27	21 (77.8%)	5 (18.5%)	1 (3.7%)
Genomics / microbiome	6	1 (16.7%)	5 (83.3%)	0
Personalised nutrition	12	5 (41.7%)	7 (58.3%)	0

Critical comment: Although most studies report positive effects, as many as 50% of studies on DT+AI integration show neutral or inconsistent results, indicating premature optimism in the literature. Genomics and microbiome are almost without empirical confirmation (only 1/6 studies shows a clear benefit). Physiological models have the highest rate of positive effects (78%), but all studies are short-term and lack external validation.

Digital twin in sport

DT was analysed in 30/46 studies. Only 60% show a clear benefit; 33% give inconsistent results, and 7% even negative (e.g., incorrect injury predictions). Most studies are pilot or experimental, without long-term validation.

Artificial intelligence in sport

AI was analysed in 35/46 studies. The most common models: machine learning (ML), deep learning (DL), SVM. Positive effect in 63% of studies, but variability is high, and studies with serious risk often overestimate effects.

Integration of DT and AI

In 18/46 studies. Only 44% positive effects – the lowest rate among domains. Half of the studies failed to show an advantage of integration over individual approaches.

Physiological models

In 27/46 studies. The highest rate of positive effects (78%). However, all studies were conducted in the laboratory, without testing in real competitive conditions.

Genomics and microbiome

Only 6 studies. One shows benefit (personalisation of nutrition), the others are conceptual or without clear conclusion. The level of evidence is very low.

Personalised nutrition

12 studies, 42% positive. Standardised protocols are lacking, results are heterogeneous.

Digital patient–athlete

The concept was identified in 14 studies, but none offers a validated system. All are at the level of conceptual framework or preliminary simulations.

Interdisciplinary authorship in studies using physiological models

Of the 46 included studies, 27 (58.7%) employed physiological models (e.g., HRV, VO_2max, lactate) as a basis for prediction or interpretation of outcomes. Table 5 summarises the disciplinary composition of authorship in these 27 studies.

Table 5. Physiologist involvement in studies using physiological models (n=27)

Category	n	%
Physiologist listed as author/co-author	4	14.8%
No physiologist listed as author	23	85.2%
Total	27	100%

Note: Authorship was determined from author affiliations and declared expertise as reported in each publication. This table provides the underlying data for the authorship analysis discussed in Section 4.2.

DISCUSSION

Synthesis of main findings

The primary purpose of this review was to map the literature on DT and AI in sport, define the role of physiologists in these models, and propose a conceptual digital patient–athlete framework. The results of this systematic review indicate a significant gap between the declared potential of digital twins (DT) and artificial intelligence (AI) in sport and the current level of empirical evidence. This gap echoes recent reviews of digital twin technology in sport, which similarly describe the field as conceptually rich but still at an early, largely exploratory stage of empirical development (Hliš et al., 2024; Lukač et al., 2022). The most consistent findings were identified in studies that included physiological parameters, where a higher percentage of positive effects was observed compared to other domains. However, these studies are predominantly short-term, conducted under controlled conditions and without external validation, a limitation consistent with recent systematic reviews of machine learning applications in sport injury prediction, which likewise report that external validation remains the exception rather than the rule (Van Eetvelde et al., 2021; Lee et al., 2026).

In contrast, integrated approaches combining DT and AI – although conceptually the most promising – show high variability of results, with a substantial proportion of studies not demonstrating a clear advantage over individual methodological approaches. Similarly, fields such as genomics and microbiome remain in early development, with a limited number of studies and a low level of empirical confirmation.

Role of the physiologist – a systemic failure

Although 27 out of 46 included studies (58.7%) use physiological models as a basis for prediction or interpretation of outcomes, the role of the physiologist as co-author or methodological consultant remains extremely limited. Specifically, a physiologist is explicitly listed as author or co-author in only 4 out of 27 studies that include physiological models (Table 5).

This imbalance indicates a structural problem in the development of digital twins and AI models in sport: physiological interpretation of data is often separated from its mathematical modelling. As a result, a large number of models achieve statistical precision but remain without adequate biological validation.

This is not a marginal methodological shortcoming but a systemic weakness in the interdisciplinary design of research in this field. Models without active participation of physiologists show a tendency towards "signal-only" optimisation, without a clear link to the athlete's adaptive physiological mechanisms. This finding aligns with a broader consensus in the sport and health science literature that meaningful progress increasingly depends on genuine interdisciplinary collaboration among physiologists, data scientists and clinicians, rather than on computational sophistication alone (Padulo, 2025).

Methodological limitations of included studies

The identified methodological limitations significantly affect the interpretation of results. Most studies are characterised by relatively small samples, lack of a priori power analysis, and limited use of independent test sets or external validation. These factors increase the risk of overfitting and reduce the generalisability of findings. These issues are not unique to this review: they mirror those identified in prior systematic reviews of machine learning in sport injury prediction, where small, single-cohort samples and the near-absence of external validation have repeatedly been flagged as central threats to model generalisability (Van Eetvelde et al., 2021; Lee et al., 2026).

In addition, model transparency varies across studies, with details on architecture, hyperparameters and validation procedures often not fully available. Combined with the absence of replication studies, these shortcomings point to the need for higher methodological standards in future research, consistent with recent calls for explainable AI (XAI) and more standardised reporting in the sports AI literature (Lee et al., 2026).

Technical and ethical challenges

Besides methodological issues, the implementation of DT and AI systems in sport faces a number of technical and ethical challenges. Data standardisation, system interoperability and data privacy are key obstacles to wider application, echoing concerns raised in reviews of AI-based injury prediction regarding inconsistent reporting practices and the need for standardised data protocols (Lee et al., 2026).

However, the results of this review suggest that improving methodological rigour is equally as important as solving technical challenges, because the quality of the evidence directly determines the potential for clinical and practical application.

Clinical and practical implications

Although certain aspects of AI and physiological model applications show encouraging results, the current level of evidence does not support widespread implementation of DT and integrated AI systems in routine sports practice.

Application is currently limited to research prototypes and specific contexts of high-performance sport. Transition to routine use requires validation under real-world conditions, including longitudinal studies and evaluation of clinically relevant outcomes, together with the kind of sustained interdisciplinary collaboration increasingly advocated in the broader sport and health science literature (Padulo, 2025).

Future research directions

Based on the identified limitations, future research should focus on:

- pre-registered protocols (e.g., PROSPERO),
- full implementation of the PRISMA methodology, including a formal graphical flow diagram,
- adequately powered samples with power analysis,
- use of independent test sets and external validation,
- replication of studies in different populations,
- greater model transparency (including explainable AI approaches),
- integration of physiological, behavioural and, where justified, omics data,
- systematic application of validated risk-of-bias and certainty-of-evidence frameworks, such as ROBINS-I and GRADE, to strengthen the interpretability of findings across this heterogeneous literature (Sterne et al., 2016; Guyatt et al., 2008).

These steps are crucial for moving from conceptual and experimental models to validated tools applicable in practice.

Limitations of this review

This review has several limitations. The included studies are heterogeneous in terms of design, population and outcomes, which prevented meta-analysis. The search was limited to English-language papers, which may introduce selection bias. Additionaly, PRISMA methodology was not fully implemented in this review, most notably in the absence of a formal graphical PRISMA flow diagram (a tabular equivalent is provided in Table 1), which should be considered when interpreting the results. While certainty of evidence was assessed using the GRADE approach (Appendix C, C1), a formal quantitative assessment of publication bias (e.g., funnel plot or Egger's test) was not performed, which should be considered when interpreting the results.

CONCLUSION AND CLINICAL IMPLICATIONS

Conclusion

This systematic review addressed its objective of mapping the existing literature (2021–2026) on the application of digital twin (DT) and artificial intelligence (AI) technologies in sport, defining the role of physiologists in the development of these models, and proposing a conceptual framework integrating biological, physiological and psychological data through AI within a digital patient–athlete model.

With regard to the first part of this objective, the review confirms that DT and AI are increasingly applied in performance analysis, injury prevention and rehabilitation, most consistently in studies incorporating physiological parameters, while evidence for integrated DT+AI, genomic and microbiome applications remains limited.

Regarding the second part, the review shows that the role of the physiologist, although critical to the biological validity of these models, remains systemically underrepresented, with explicit authorship or methodological involvement in only a minority of the studies that rely on physiological data.

With respect to the third part, the findings support the value of the proposed conceptual framework integrating genomic, microbiome, nutritional and psychophysiological data through AI within a digital patient–athlete model, while confirming that no validated system of this kind currently exists.

Overall, this review indicates that while DT and AI hold conceptual promise for sports science and medicine, their full integration with physiological expertise and their translation into validated, clinically applicable tools have not yet been achieved.

Clinical and practical implications

The results of this review suggest that the application of DT and AI in sport currently has the greatest potential in the context of:

• improving performance analysis through the integration of physiological data,

• early recognition of patterns associated with fatigue and injury risk,

• development of personalised training approaches in controlled or high-performance environments.

However, implementation in routine sports and clinical practice remains limited. Wider application requires model validation under real-world conditions, data standardisation and more clearly defined protocols for result interpretation.

The importance of an interdisciplinary approach is particularly emphasised, where the involvement of physiologists and other biomedical experts in the early stages of model development can contribute to their validity and practical applicability.

Final remark

Digital twin and artificial intelligence represent an evolution, not a replacement, of existing approaches in sports science and medicine. Their value depends on data quality, methodological rigour and adequate interpretation of results.

Future progress in this field requires a shift of focus from conceptual and pilot studies to rigorously designed research with larger samples, external validation and transparent models. In this context, the development of reliable and clinically relevant digital twins of athletes remains a long-term but achievable goal.

REFERENCES

1. Bartlett, J. D., O'Connor, F., Pitchford, N. (2020). Monitoring training load to understand fatigue in athletes. *Sports Medicine*, *50*(7), 1243–1258.

2. Bjelica, B., Hajder, Đ., Dragosavljević, S., Perović, T., Pržulj, R., Aksović, N., & Bubanj, S. (2025). Use of AI Technology Training on Motor Parameters: Systematic review. *Sportske nauke i Zdravlje 15*(1), 119-127.
3. Glebova, E., Desbordes, M., & Bodet, G. (2023). Digital venue twins: Enhancing spectator experience and event management. *Sport, Business and Management: An International Journal*, *13*(4), 401–418.
4. Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*, *336*(7650), 924-926.
5. Hliš, T., Fister, I., & Fister Jr., I. (2024). Digital twins in sport: Concepts, taxonomies, challenges and practical potentials. *Expert Systems with Applications*, *258*, 125104.
6. Kakavas, G., Malliaropoulo, N., Mitrotasio, M., Tsaklis, P., & Bikos, G. (2022). Is virtual twins technology the future in sports medicine. *Journal of Physical Medicine Rehabilitation & Disabilities*, *8*(1), 75.
7. Lee, C. H., Chang, M. H., & Li, Z. H. (2026). From data to prevention: A systematic review of artificial intelligence applications in sports injury prediction. *Technology and Health Care*, 09287329261455797.
8. Li, X., Loscalzo, J., Mahmud, A. F., Aly, D. M., Rzhetsky, A., Zitnik, M., & Benson, M. (2025). Digital twins as global learning health and disease models for preventive and personalized medicine. *Genome Medicine*, *17*(1), 11.
9. Liu, D., Wang, H., Xiang, X., & Fang, H. (2025). Digital Twin in healthcare and medical field: A review. *Pattern Recognition and Computer Vision in the New AI Era*, 247-271.
10. Liu, L., Zhu, Q., Zong, Y., Chen, X., Zhang, W., & Wang, J. (2026). Machine learning prediction models for preeclampsia: systematic review and meta-analysis. *Journal of Medical Internet Research*, *28*, e78714.
11. Lukač, L., Fister Jr., I., & Fister, I. (2022). Digital twin in sport: From an idea to realization. *Applied Sciences*, *12*(24), 12741.
12. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*.
13. Padulo, J. (2025). Sport and health science: interdisciplinary approaches to modern challenges. *British Medical Bulletin*, *155*(1), ldaf007.
14. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Bmj*, *372*.
15. Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj*, *355*.
16. Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of Experimental Orthopaedics*, *8*(1), 27.
17. Vec, V., Tomažič, S., Kos, A., & Umek, A. (2024). Trends in real-time artificial intelligence methods in sports: a systematic review. *Journal of Big Data*, *11*(1), 148.
18. Zhou, D., Chen, J., Wang, L., et al. (2025). Artificial intelligence in sport: A systematic review of performance analysis and injury prevention. *Journal of Sports Science*, *43*(2), 145–162.

APPENDICES

APPENDIX A – PRISMA 2020 Checklist (completed)

Section/Topic	#	Checklist item	Reported on page
TITLE	1	Title identifies the report as a systematic review	1
ABSTRACT	2	Structured summary	1
INTRODUCTION	3	Rationale	1
	4	Objectives	1
METHODS	5	Eligibility criteria	2
	6	Information sources	2
	7	Search strategy	2
	8	Selection process	2-3
	9	Data collection process	3
	10	Data items	3
	11	Risk of bias assessment (ROBINS-I)	3 + Appendix B (13)
	12	Effect measures – not applicable (qualitative synthesis)	N/A
	13	Synthesis methods	3
	14	Assessment of publication bias – method not pre-specified; disclosed as not performed in Limitations	9
	15	Certainty assessment (GRADE)	3 + Appendix C, C1 (13-15)
RESULTS	16	Study selection (numbers, PRISMA flow – tabular equivalent, no graphical diagram)	2-4 (Table 1)
	17	Study characteristics	4-5 (Table 3)
	18	Risk of bias in studies (summary)	4 (Table 2) + Appendix B (13)
	19	Results of individual studies	6-7
	20	Synthesis of results (quantification)	4, 6 (Table 4)
	21	Reporting of publication bias – not performed	N/A
	22	Certainty of evidence (qualitative)	4
DISCUSSION	23	Interpretation (critical)	7-9
	24	Limitations (study-level and review-level)	9
	25	Conclusions (critical, with recommendations)	9-10
OTHER	26	Funding – not declared	N/A

APPENDIX B – ROBINS-I domains per study (summary for all 48 original studies)

Study ID	Confounding	Selection	Classification	Deviation	Missing data	Outcome measurement	Selective reporting	Overall rating
1	Low	Low	Moderate	Low	Low	Low	Low	**Low**
2	Moderate	Low	Moderate	Low	Moderate	Low	Low	**Moderate**
3	Moderate	Moderate	Serious	Moderate	Low	Moderate	Low	**Moderate**
4	Low	Low	Low	Low	Low	Low	Low	**Low**
5	Serious	Moderate	Serious	Moderate	Moderate	Moderate	Low	**Serious**
6	Critical	Serious	Serious	Serious	Moderate	Serious	Moderate	**Critical** (excluded)
...	...	...	...	...	...	...	...	...
48	Moderate	Serious	Serious	Moderate	Moderate	Moderate	Low	**Serious**

Note: The full table of all 48 studies (with anonymised identifiers) is available upon request. Studies with critical risk in any domain (two studies) were excluded from the synthesis. Studies with serious risk were included but their findings were interpreted with great caution.

APPENDIX C – GRADE assessment of certainty of evidence

Methodological framework

The certainty of evidence was assessed according to the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach, evaluating the following domains:

- risk of bias
- inconsistency of results
- indirectness of evidence
- imprecision of estimates
- publication bias

Initial level of evidence

Since none of the 46 included studies was a randomised controlled trial (RCT), the initial level of evidence for all domains was classified as **low.** This is in accordance with GRADE methodology, whereby observational studies automatically start with a lower certainty level compared to RCT designs.

Assessment by domain

Domain	Certainty of evidence	Main reasons
Digital twin (DT)	Low	heterogeneity of models, lack of RCT, pilot designs
Artificial intelligence (AI)	Low	high risk of overfitting, lack of external validation
DT + AI integration	Very low	inconsistent results, methodological heterogeneity
Physiological models	Low	laboratory conditions, no longitudinal validation
Genomics / microbiome	Very low	limited number of studies, mostly conceptual
Personalised nutrition	Very low	lack of standardisation and controlled studies

Main reasons for downgrading

All domains were further downgraded from "low" to "very low" in some cases because of:

- absence of randomised controlled trials (0 RCT)
- high or moderate risk of bias in most studies (ROBINS-I analysis)
- substantial heterogeneity in design, population and outcomes
- lack of external validation and replication studies
- presence of publication and selection bias

Overall GRADE conclusion

The overall certainty of evidence in the field of digital twins and artificial intelligence in sport is rated as **low to very low.** This applies even to domains with statistically positive effects, indicating that the magnitude of effect is not accompanied by adequate methodological quality.

Key interpretation

Although some studies show potential for improving performance and predicting injuries, the current evidence base is not robust enough for clinical or practical implementation without additional validation through:

- randomised controlled trials
- longitudinal designs
- external replication of models
- standardisation of AI and DT protocols

PPENDIX C1 – GRADE Summary of Findings (SoF) table

Outcome / domain	Number of studies	Study design	Effect direction	Certainty of evidence (GRADE)	Main limitations	Confidence in conclusion
Digital twin (DT) – performance and training adaptation	30	Observational + pilot	Mostly positive	Low	heterogeneity of models, lack of RCT, pilot designs	Low confidence
Artificial intelligence (AI) – prediction of performance and injuries	35	ML/DL studies (observational)	Positive	Low	overfitting, lack of external validation	Low confidence
DT + AI integration	18	Experimental and simulation	Mixed / inconsistent	Very low	methodological misalignment, small samples	Very low confidence
Physiological models (HRV, VO_2max, lactate)	27	Laboratory studies	Mostly positive	Low	laboratory conditions, no longitudinal validation	Low confidence
Genomics and microbiome	6	Conceptual + pilot	Neutral	Very low	small number of studies, lack of empirical validation	Very low confidence
Personalised nutrition in DT models	12	Observational studies	Inconsistent	Very low	lack of standardisation, heterogeneous protocols	Very low confidence

Received on 21.04.2026.
Accepted on 24.08.2026.