OPEN ACCESS
Indonesian Journal of Sport, Health and Physical Education Sciences Vol 4 No 3 (2026) https://doi.org/10.58723/inasport.v4i3.845

Check for updates

Research Article

Bibliometric Analysis of Artificial Intelligence in Sports Talent Identification (2016-2026)

Received: July 30, 2026	Revised: November 1, 2026	Accepted: November 9, 2026	Published: November 13, 2026

Martinez Edison Putra*, Ella Fauziah

Abstract:

Background: Talent identification in sport at an early age is a foundational element of successful long-term athlete development systems, yet research on the application of artificial intelligence (AI) to this process remains fragmented across sports, computational techniques, and types of synthesis study.
Aims: This study maps publication productivity, author and country collaboration, influential sources, and the thematic evolution of research on AI applications in sports talent identification using a bibliometric approach.
Methods: Data were retrieved from Scopus on 17 July 2026 (articles in English, 2016–2026). Document selection was reported using an adapted PRISMA 2020 flow, yielding 83 documents that were analyzed with Bibliometrix/Biblioshiny and VOSviewer.
Result: Annual output was low in 2016–2020 (10 articles in total; 1–3 per year), rose from 2022, and reached 20 articles in 2025; 2026 is a partial year (17 articles up to 17 July 2026) and is not comparable with complete years. Journal of Sports Sciences was the most productive source (7 articles), and the remaining output was dispersed across 63 journals. Hohmann and Pion were the most productive and most cited authors, although several of their documents are shared. By total production, China (21 documents), Germany (12), and the United Kingdom (9) led, and Indonesia was among four countries tied for fourth place with six documents each. Thematic mapping placed sports analytics, player performance metrics, and player recruitment among the motor themes (based on few documents), talent identification and machine learning among the basic themes, and anthropometry and biological maturation among the niche themes. Thematic evolution suggested a shift from data mining and discriminant analysis toward neural networks, artificial intelligence, and deep learning. A cited-reference analysis showed that the intellectual base rests mainly on classic talent identification and development works (each cited by at most seven documents), while machine learning references were cited by four to five documents.
Conclusion: Given the small corpus and the single database, the findings are exploratory. They point to research gaps in terminology standardization, the examination of bias in AI-based assessment, and international collaboration.

Keywords: Artificial Intelligence; Bibliometric Analysis; Machine Learning; Sports Talent Identification; Thematic Evolution

1. INTRODUCTION

Talent identification in sport from an early age is a fundamental component of modern athlete development systems, as elite performance is heavily determined by the ability to identify potential athletes from the earliest stages of development (Till & Baker, 2020). Various countries, including Indonesia through its National Sports Grand Design (DBON), position talent identification as the foundation of Long-Term Athlete Development (LTAD) programs to prepare athletes for international achievement (Bakhtiar et al., 2023). The talent identification process enables the mapping of individual anthropometric, biomotor, technical, and tactical characteristics so that potential can be directed toward the most suitable sport (N. Susanto et al., 2024). Beyond physical aspects, psychological factors, coaching support, and school policy also contribute to successful talent identification, making the process multidimensional and irreducible to a single indicator (Changqing et al., 2022). Talent identification in sport has therefore developed into one of the major fields of study within sport science, as well as a foundation for developing long-term athlete development systems.

Publisher Note:

CV Media Inti Teknologi stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright

©2026 by the author(s).

Licensee CV Media Inti Teknologi, Bengkulu, Indonesia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike (CCBY-SA)license (https://creativecommons.org/licenses/by-sa/4.0/).

Recent developments indicate that sports talent identification has increasingly moved toward a data-driven advanced analytical methods. Traditional approaches that rely mainly on isolated anthropometric or physical indicators have been criticized because athlete development is influenced by complex interactions between biological maturation, training exposure, environmental factors, and performance progression (Barraclough et al., 2022; Fernandez Ortega et al., 2021). Consequently, contemporary talent identification models increasingly emphasize longitudinal monitoring and multidimensional assessment rather than single-time-point evaluation (Kelly et al., 2022). This transition provides an important foundation for the integration of artificial intelligence (AI), as computational approaches allow researchers and practitioners to analyze complex patterns from large and heterogeneous datasets that are difficult to process using conventional statistical approaches.

The advancement of artificial intelligence and machine learning has further accelerated this transformation by enabling the analysis of complex relationships among multiple athlete characteristics. Recent studies in sports science demonstrate that machine learning algorithms are increasingly applied to performance evaluation, athlete monitoring, and predictive modelling because these approaches can identify nonlinear patterns that are difficult to capture using conventional statistical methods (Silvino et al., 2025). However, the effectiveness of these models remains dependent on data quality, methodological robustness, and appropriate validation procedures (Naughton et al., 2024; Tang et al., 2026).

Research on sports talent identification initially focused on unidimensional approaches, particularly anthropometric characteristics and basic physical capacity, as seen in talent-profile studies of primary-school-age children (Bakhtiar et al., 2023) and talent prediction tests in athletics (E. Susanto et al., 2023). The limitations of this unidimensional approach subsequently drove a shift toward more comprehensive models that combine physiological, technical, tactical, and psychological variables simultaneously (Nijenhuis et al., 2024). In parallel, research began exploring computational techniques to support the talent identification process, ranging from multi-criteria decision-making (MCDM) methods such as the Analytic Hierarchy Process (AHP), TOPSIS, and fuzzy logic (Khan et al., 2023), to basic machine learning algorithms such as K-Nearest Neighbour, Support Vector Machine, and Random Forest, which have been applied across various sports (Reyaz et al., 2022). However, a systematic review of talent identification and development (TID) models from 1995-2020 found that the majority of these models still relied on conventional statistical techniques and MCDM, while adoption of more advanced computational techniques remained limited and uneven (Khan et al., 2023).

The increasing availability of athlete tracking systems, wearable technologies, and performance database has accelerated the adoption of machine learning techniques in sports science. Machine learning algorithms such as Random Forest, Support Vector Machine, Artificial Neural Network, and ensemble-based methods have been applied to identify performance patterns, classify athlete characteristics, and support prediction tasks in various sporting contexts (Kelly et al., 2022; Tang et al., 2026). Nevertheless, the implementation of these approaches remains challenging because model accuracy depends strongly on dataset quality, sample representativeness, feature selection, and methodological transparency (Naughton et al., 2024; Tang et al., 2026). Therefore, understanding the development structure of AI-based talent identification research is necessary to determine whether current advances represent a sustainable scientific trend or remain limited to fragmented applications. A shift toward more complex and integrated machine learning applications only became clearly visible in more recent systematic reviews, which mapped the application of various predictive algorithms to integrate multidimensional data in order to improve the accuracy of predicting athletic potential (Tang et al., 2026).

The complexity of sports talent identification lies in its multifactorial nature and the absence of a single consensus on the definition or operational criteria of "talent" itself (Till & Baker, 2020). Factors influencing successful talent identification span biological, technical, psychological, and socio-cultural dimensions that interact dynamically (Tang et al., 2026). A systematic review of computation-based TID models further found no consensus on the parameters used across different sports, along with inconsistencies in age, terminology, and evaluation criteria that complicate the development of a unified cross-sport talent identification model (Khan et al., 2023). At the level of machine learning application,

this complexity is reflected in the still-limited number of models that integrate physical, anthropometric, and psychological parameters simultaneously; a review of cricket, for example, found no machine learning model that predicted young players' talent based on a combination of these relevant parameters (Reyaz et al., 2022). Conditions such as the relative age effect and variation in biological maturity at a young age can further distort identification outcome if not methodologically controlled, requiring a holistic, constraint-led model of talent evaluation (Junior et al., 2021). This multidimensional complexity is what has driven the exploration of AI-based computational approaches, even though the literature indicates that adoption of these techniques remains, in practice, relatively stagnant and not yet commensurate with their potential to address this complexity (Khan et al., 2023).

Recent perspectives emphasize that talent identification should be understood as a continuous developmental process rather than a single assessment event. Athlete potential emerges through interactions between biological characteristics, training exposure, environmental conditions, and long-term performance progression. Therefore, contemporary talent development systems increasingly incorporate longitudinal monitoring frameworks that combine physical, technical, psychological, and contextual indicators to improve identification (Henriksen & Stambulova, 2023; Shahidi et al., 2023).

The integration of AI into sports talent identification represents a broader transformation within elite sport, where decision-making increasingly relies on computational analysis rather than solely on expert observation. Artificial intelligence applications have been reported in athlete assessment, tactical analysis, performance prediction, and talent development; however, challenges related to explainability, ethical implementation, and algorithmic transparency remain important considerations (Cossich et al., 2023; Puce et al., 2026).

The current landscape of sports talent identification research is highly diverse, both in terms of sport, methodological approach, and the type of synthesis study produced. Some systematic reviews focus on physical education as a vehicle for talent identification in schools (Prieto-ayuso et al., 2020), while other narrative reviews highlight methodological approaches to talent identification in team sports in general (Barraclough et al., 2022). In specific sports such as basketball, meta-analyses show that anthropometric and physiological variables such as height, body mass, and Yo-Yo test capacity consistently distinguish elite from non-elite athletes (Han et al., 2023), while in racket sports, systematics reviews highlight the added value of multidimensional and longitudinal approaches (Nijenhuis et al., 2024).

This landscape also encompasses reviews in the computational domain, ranging from MCDM techniques in general (Khan et al., 2023) to the application of machine learning in specific sports such as cricket (Reyaz et al., 2022). More specifically, in the context of contemporary AI applications, a recent systematic review mapped the use of machine learning in talent identification of young athletes in team sports, identifying algorithms used ranging from Random Forest and Support Vector Machine to K-Nearest Neighbors, while also noting a high risk of methodological bias across the majority of studies (Tang et al., 2026). At a more macro level, a bibliometric study by Parra-Martinez & Wai, (2023) mapped trends, productivity, and research collaboration in talent identification across scientific disciplines over eight decades, although that study was not specifically focused on the sports domain.

At the national level, a recent systematic review of Scopus-indexed, Indonesian-authored publications, conducted by the present authors, found that software- and machine-learning-based approaches to sports talent identification in Indonesia show adequate validity and accuracy, but that none of the included studies implemented deep learning or neural networks (Putra & Fauziah, 2026). The diversity of focus and method across these studies indicates that research on sports talent identification has developed rapidly, yet its mapping remains fragmented by sport, type of computational technique, and type of approach studied.

The various synthesis studies described above show important contributions, but also leave gaps that remain unaddressed. The systematic review on physical education covers only a single decade and is limited to 11 articles, making its scope narrow (Prieto-ayuso et al., 2020). The narrative review of methodological approaches in team sports does not apply quantitative synthesis or bibliometric mapping of the field's development (Barraclough et al., 2022). Meta-analyses and systematic reviews

on specific sports such as basketball (Han et al., 2023) and racket sports (Nijenhuis et al., 2024) provide deep understanding but are limited to a single sport, and therefore do not depict cross-sport trends in talent identification research. In the computational domain, the systematic review by Khan et al., (2023) maps MCDM techniques in general but does not focus on machine learning or the latest generation of AI, while the review by Reyaz et al., (2022) focuses narrowly on a single sport cricket and therefore does not represent cross-country trends in AI application for sports talent identification.

The systematic review on machine learning in team-sport talent identification (Tang et al., 2026) is the study most relevant to the topic of AI, yet it is a systematic review, focusing on methodological evaluation and bias risk in individual studies rather than on quantitative mapping of publication trends, collaboration, or thematic evolution. On the other hand, the available bibliometric study (Parra-Martinez & Wai, 2023) maps the development of talent identification across disciplines and globally, but its scope is not specific to the sports domain or to the application of AI within it. Consequently, to date no study has specifically mapped the development, productivity, collaboration network, and thematic evolution of research on the application of artificial intelligence in sports talent identification in a comprehensive, cross-sport manner.

A bibliometric approach was chosen because it provides a structural, quantitative picture of a field, covering publication trends, collaboration networks, influential sources, and the evolution of research themes (Parra-Martinez & Wai, 2023). Unlike systematic or narrative reviews, which examine a limited set of studies in depth (Barraclough et al., 2022; Prieto-Ayuso et al., 2020), bibliometric analysis maps macro-level patterns across a research corpus. Because AI-based talent identification involves a growing variety of algorithms, data sources, and sports (Tang et al., 2026), such mapping is useful for revealing growth patterns, collaboration, and shifts in thematic focus. As the corpus in this field is still small, the results are interpreted as exploratory.

Based on the above, this study aims to map the development of research on the application of artificial intelligence in sports talent identification through a bibliometric analysis approach, covering publication productivity trends, author and country collaboration networks, the most influential journal and sources, and the evolution of research themes over time. The results of this study are expected to provide a comprehensive picture of the structure and direction of this field's development, while also identifying research gaps that can serve as a basis for future studies.

2. MATERIAL AND METHOD

2.1 Research Design

This study uses bibliometric analysis approach to map productivity trends, collaboration networks, and the evolution of research themes concerning the application of artificial intelligence (AI) in sports talent identification. The identification, screening, and selection of documents followed an adapted PRISMA 2020 flow.

2.2 Data Sources and Search Strategy

Data were retrieved from the Scopus database on 17 July 2026. The search used the title, abstract, and keywords fields with the following string: TITLE-ABS-KEY ((“artificial intelligence” OR “machine learning” OR “neural network” OR “computer vision” OR “data mining” OR “predictive analytics”) AND (“talent identification” OR “talent detection” OR “athlete identification” OR “athlete selection” OR “sport talent” OR “sports talent”)) AND PUBYEAR > 2015 AND PUBYEAR < 2027. The period 2016–2026 was selected to cover the decade preceding data collection. The query returned 180 records. Document type (article) and language (English) were then applied as filters. Conference papers and reviews were excluded to restrict the corpus to peer-reviewed research articles with consistent metadata, the implication of this restriction is discussed in Section 3.2.8.

2.3 Eligibility Criteria

2.3.1 Inclusion Criteria

(a) articles published in 2016–2026 in English; (b) documents discussing the application of AI, machine learning, or similar computational techniques in sports talent contexts; (c) documents with complete metadata (title, abstract, authors, affiliations, year, keywords).

2.3.2 Exclusion Criteria

(a) non-article document types (conference papers, reviews, book chapters, retracted documents, books, conference reviews, notes; n = 86); (b) non-English documents (n = 4); (c) documents outside the thematic scope, in which “talent identification” referred to human resource management, business,

OPEN ACCESS
Indonesian Journal of Sport, Health and Physical Education Sciences Vol 4 No 3 (2026) https://doi.org/10.58723/inasport.v4i3.845
Check for updates
Research Article

or non-sport education (n = 7), identified by screening titles and abstracts.

2.4 Data Selection Process

2.4.1 Identification

The search returned 180 records; after the article filter (n = 86 excluded) and the language filter (n = 4 excluded), 90 records remained.

2.4.2 Screening

Records were exported from a single database and checked for duplicates by DOI and title similarity (90% threshold); no duplicates were found. Titles and abstracts of the 90 records were screened by the second author (EF) and checked by the first author (MEP); disagreements were resolved by discussion. Seven records were excluded as off-topic. Records with weaker thematic fit that matched the search terms were retained (see Section 3.2.8).

2.4.3 Included

83 documents were retained for analysis.

Table 1. PRISMA 2020 Document Selection Summary

Stage	Action	n retained	N excluded (reason)
Identification	Scopus database search (2016-2026)	180	-
	Document-type filter (article only)	94	86 (not an article)
	Language filter (English only)	90	4 (not English)
Screening	Duplicate check (DOI + title similarity, 90% threshold)	90	0 (no duplicates)
	Title/abstract screening (topical relevance)	83	7 (off-topic use of “talent identification”)
Included	Final corpus for bibliometric analysis	83	-

2.5 Analysis and Bibliometric Parameters

The 83 documents were exported from Scopus on 17 July 2026 as CSV files containing full records and cited references. Citation counts reflect the values recorded by Scopus on the export date. Analyses were performed using the Bibliometrix R package (version 5.5.0) through its Biblioshiny interface in R (version 4.6.1) and VOSviewer (version 1.6.20). Author keywords were analyzed as supplied by the authors; no thesaurus was applied, so spelling variants (e.g., neural network/neural networks, soccer/football) were treated as separate items. Seventy-seven of the 83 documents provided author keywords (315 unique keywords). Because 2026 is a partial year, it was shown in the annual production analysis but excluded from the growth-rate estimation (2016-2025).

Performance analysis covered annual scientific production, productive sources (number of documents), productive authors (number of documents and total citations), and countries. For countries, corresponding-author articles were classified as single-country publications (SCP) or multiple-country publications (MCP), and the MCP ratio was calculated for each country. Total country production was also counted from all author affiliations, counting each country once per document.

Science mapping comprised four analyses. First, a thematic map was built from author keywords in Biblioshiny with 250 words, a minimum cluster frequency of five per thousand documents (equivalent to a minimum of two documents in this corpus), and Walktrap clustering. Second, thematic evolution was analyzed across four periods (2016-2022, 2023-2024, 2025, and 2026) defined by cutting years 2022, 2024, and 2025. Third, a keyword co-occurrence network was generated in VOSviewer from author keywords using full counting and a minimum of two occurrences (41 of 315 keywords met this threshold), with association strength normalization, a resolution of 1.00, and a minimum cluster size of one which yielded nine clusters (Table 5). Fourth, a country collaboration network (co-authorship, full counting) was generated with a minimum of four documents per country, and the cited references of the 83 documents were counted to identify the most frequently cited works (Table 6). Because Biblioshiny and VOSviewer use different clustering procedures, their clusters are not

expected to coincide. Given the small corpus (n = 83), the science mapping results are treated as exploratory.

3. RESULTS AND DISCUSSION

3.1 Results

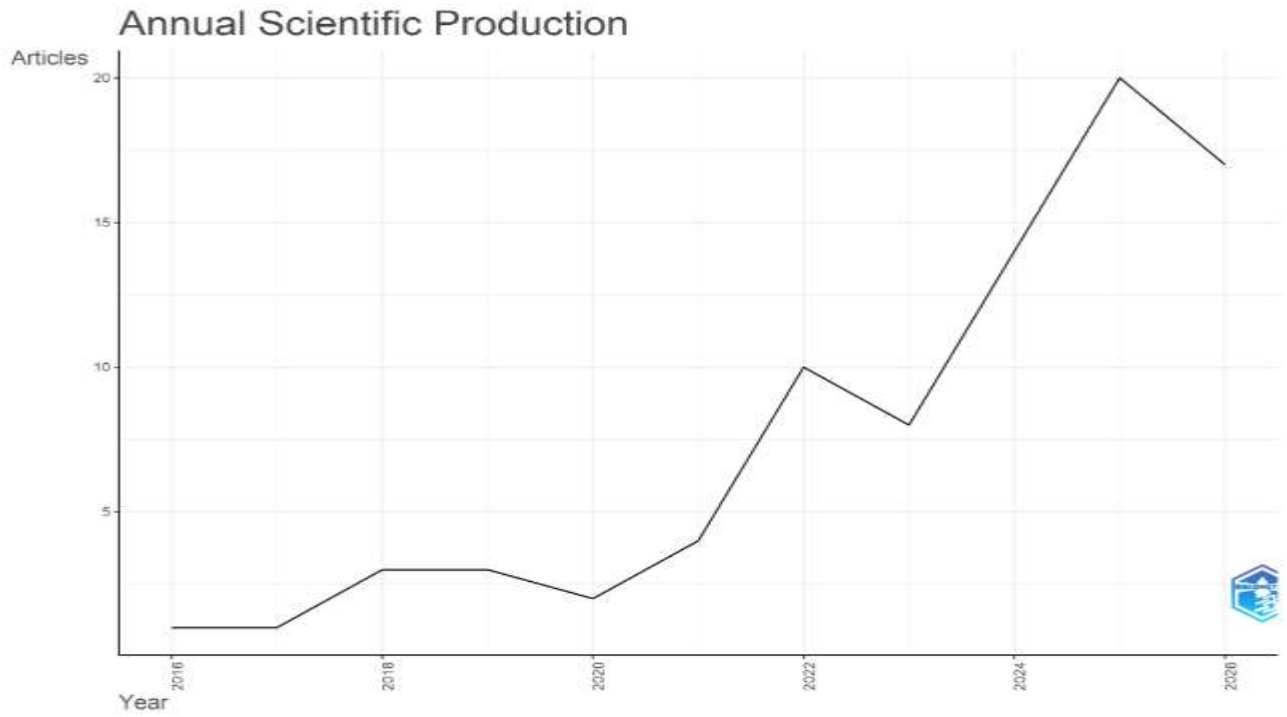


Figure 1. Annual Publication Productivity Trend

Annual output was low in 2016-2020 (10 documents in total; 1-3 per year), reached 10 documents in 2022 after four in 2021, dipped slightly to eight in 2023, and then rose to 14 in 2024 and 20 in 2025 (compound annual growth rate 2016-2025; 39.5%, calculated from a base of one document and therefore indicative only). The 2026 count (17 documents) covers only the period up to 17 July 2026 and is already close to the 2025 total; it is therefore shown separately in Figure 1 and excluded from growth estimates, and Scopus indexing lag may further affect recent years. This pattern is consistent with the narrative in the introduction that machine learning in sports talent identification has only recently begun to grow rapidly (Tang et al., 2026).

Table 2. Most Productive Source (Journals)

No	Source	Number of Articles
1	Journal of Sports Sciences	7
2	Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte	3
3	Sensors	3
4	Applied Sciences (Switzerland)	2
5	Computational Intelligence and Neuroscience	2
6	Frontiers in Physiology	2
7	Frontiers in Psychology	2
8	International Journal of Sports Science and Coaching	2
9	Journal of Applied Sport Management	2
10	Journal of Human Sport and Exercise	2
11	PLOS ONE	2
12	Sport TK	2
13	Sports	2

The 83 documents were published in 63 journals (after merging two spelling variants of Journal of Sports Human and Exercise). Journal of Sport Sciences contributed the most (7 articles), followed by Revista Internacional de Medicina y Ciencias de la Actividad Fisica y del Deporte and Sensors with

OPEN ACCESS
Indonesian Journal of Sport, Health and Physical Education Sciences Vol 4 No 3 (2026) https://doi.org/10.58723/inasport.v4i3.845
Check for updates
Research Article

three articles each. Fifty journals contributed one article each and 13 journals contributed two or more (33 articles in total), so output is widely dispersed across sports science, computing, and multidisciplinary outlets.

Table 3. Most Productive Authors

No	Authors	Year	Freq.	Total Citation	Total Citation Per Year Publication
1	Hohmann, Andreas	2017	1	35	3.50
		2019	1	50	6.25
		2021	1	22	3.67
		2023	1	11	2.75
2	Pion, Johan	2017	1	35	3.50
		2019	1	50	6.25
		2025	1	8	4.00
3	Zhao, Kewei	2019	1	50	6.25
		2023	1	11	2.75
4	Lenoir, Matthieu	2017	1	35	3.50
		2025	1	8	4.00
5	Siener, Maximilian	2021	1	22	3.67
		2023	1	11	2.75
6	Faber, Irene	2021	1	22	3.67
		2025	1	8	4.00
7	Bonilla, Diego A.	2020	1	10	1.43
		2022	1	9	1.80
8	Cannataro, Roberto	2020	1	10	1.43
		2022	1	9	1.80
9	Petro, Jorge L.	2020	1	10	1.43
		2022	1	9	1.80
10	Urrutia-Mosquera, Wilson	2020	1	10	1.43
		2022	1	9	1.80
11	Cameron, Shaun	2025	2	3	1.50
12	Mara, Jocelyn	2025	2	3	1.50
13	Radwan, Ibrahim	2025	2	3	1.50

Hohmann was the most productive and most cited author (4 documents; 118 citations; 2017–2023). The remaining eleven authors each contributed two documents. Authors with identical citation counts appear to co-author the same documents: Hohmann, Pion, and Lenoir (2017); Hohmann, Pion, and Zhao (2019); Hohmann, Siener, and Faber (2021); Bonilla, Cannataro, Petro, and Urrutia-Mosquera (2020 and 2022); and Cameron, Mara, and Radwan (2025). The ranking therefore reflects a few research groups rather than independent contributors. Citation counts of recent documents are lower because they have had less time to accumulate citations.

Table 4. Corresponding Author's Countries

No	Country	Articles	SCP	MCP	MCP ratio	No	Country	Articles	SCP	MCP	MCP ratio
1	China	20	16	4	0.200	11	Belgium	2	1	1	0.500
2	Australia	5	2	3	0.600	12	Brazil	2	0	2	1.000
3	United Kingdom	5	4	1	0.200	13	Colombia	2	0	2	1.000
4	Germany	4	4	0	0.000	14	India	2	0	2	1.000
5	Indonesia	4	2	2	0.500	15	Netherlands	2	0	2	1.000
6	Malaysia	4	3	1	0.250	16	Romania	2	1	1	0.500
7	Canada	3	2	1	0.333	17	Algeria	1	1	0	0.000
8	Italy	3	1	2	0.667	18	Denmark	1	0	1	1.000
9	Turkey	3	3	0	0.000	19	Greece	1	1	0	0.000
10	USA	3	3	0	0.000	20	Iran	1	1	0	0.000

By corresponding-author affiliation, China led (20 articles; 16 single-country and 4 multiple-country; MCP ratio 0.20), followed by Australia (5; 0.60) and the United Kingdom (5; 0.20), with Germany, Indonesia, and Malaysia at 4 articles each. China's output was thus largely domestic, whereas Australia, Italy, and several countries with two articles (Brazil, Colombia, India, and the Netherlands) had MCP ratios of 0.60-1.00. By total production counted from all author affiliations (each country once per document), China ranked first with 21 documents (25% of the corpus), followed by Germany (12) and the United Kingdom (9); Australia, Indonesia, Italy, and Spain each contributed six documents.

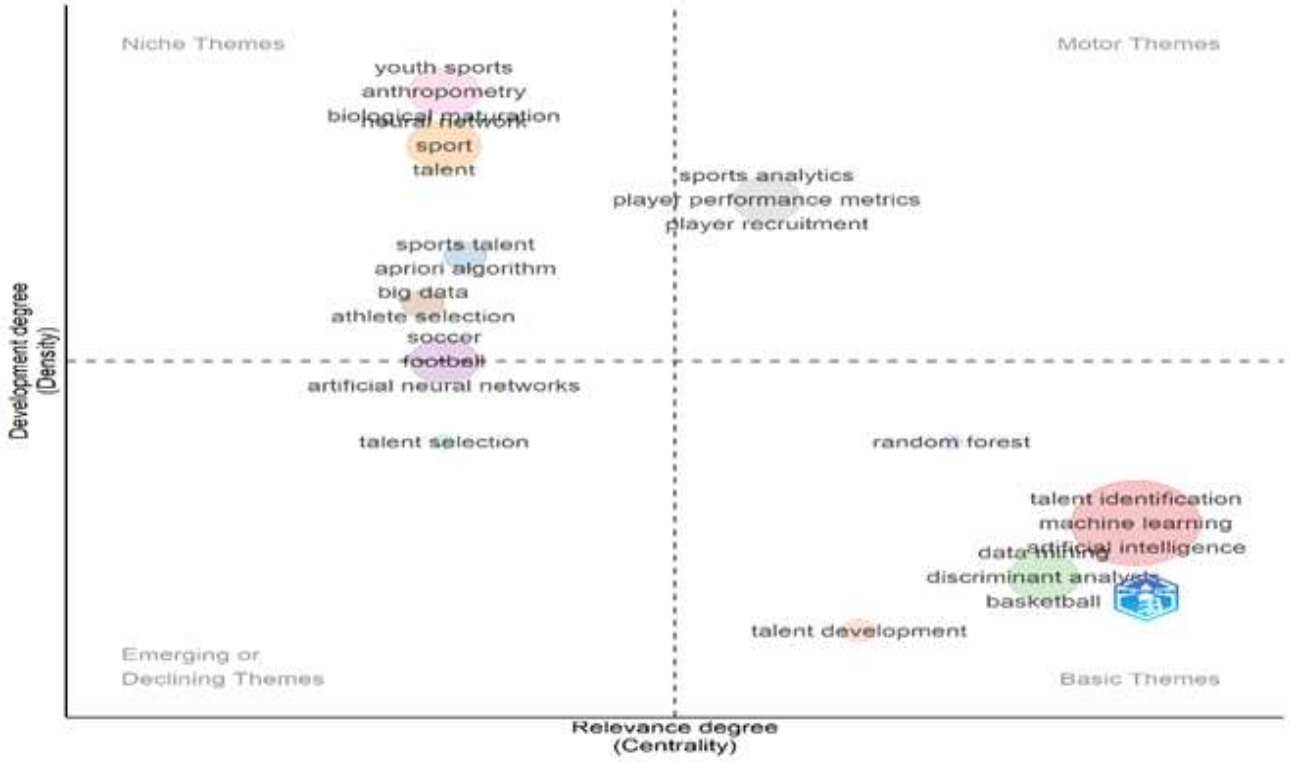


Figure 2. Research Theme Map

Thematic mapping based on author keywords places the field's thematic structure into four density centrality quadrants. In the motor themes quadrant (high density and centrality), the themes of sports analytics, player performance metrics, and player recruitment emerge as clusters that are both well developed and central to the field, indicating that the application of data analytics to support recruitment decisions and player performance evaluation has become mainstream research. In the basic themes quadrant (high centrality, relatively lower density), a large cluster containing talent identification, machine learning, artificial intelligence, data mining, discriminant analysis, and basketball confirms that AI-based talent identification is the main thematic foundation connecting nearly all other clusters, supported by co-occurrence of talent development and random forest in the same quadrant.

By contrast, the niche themes quadrant (high density, low centrality) is dominated by themes that are specific and tend to stand alone, such as youth sports, anthropometry, biological maturation, sport, and talent, as well as a second cluster containing sport talent, apriori algorithm, big data, athlete selection, soccer, football, and artificial neural networks. The position of these themes in the niche quadrant indicates that research on biological maturity and anthropometry in a particular sport (soccer) has developed considerably on its own but has not yet engaged much with the large AI/ML cluster in the basic themes' quadrant, potentially forming a separate research silo. Meanwhile, talent selection located in the emerging or declining themes quadrant (low density and centrality) indicates a theme that remains small in scale and has not yet found a clear position within the field's intellectual structure, which can be interpreted either as a newly emerging topic or, conversely, as one that is beginning to be abandoned.

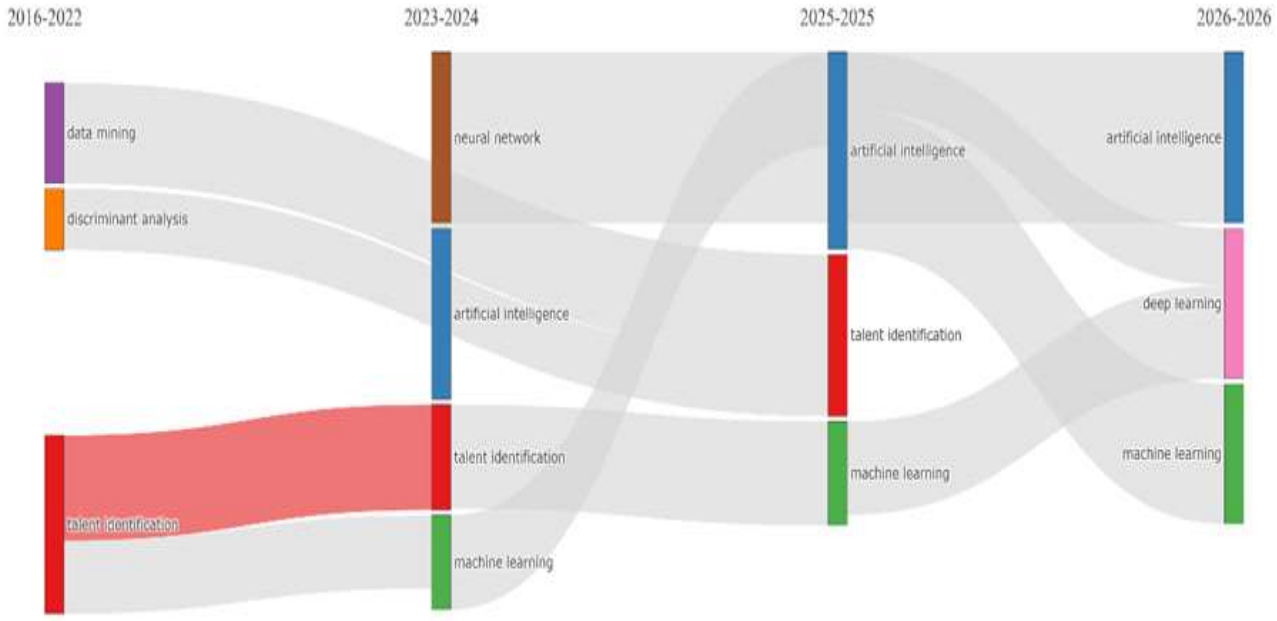


Figure 3. Thematic Evolution

The thematic evolution map across periods shows a shift in research focus from themes oriented more toward conventional physical/biological measurement in the earlier periods toward themes oriented more toward computational techniques (machine learning, data mining) in later periods. This shift pattern generally reinforces the argument developed in the introduction regarding the transition from unidimensional to AI-based approaches (Khan et al., 2023; Tang et al., 2026), while also providing quantitative evidence that had previously only been supported by small scale narrative/systematic reviews.

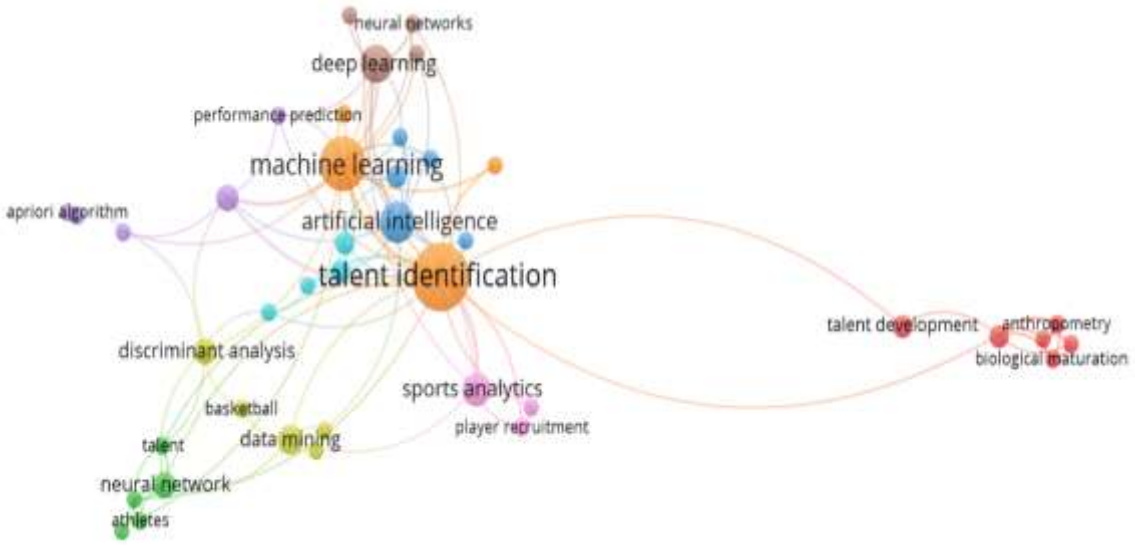


Figure 4. Keyword Co-Occurrence Network

The network visualization of author keywords in VOSviewer reinforces the thematic map findings: keywords such as talent identification, machine learning, artificial intelligence, and neural networks form large, tightly interconnected nodes, indication that these terms serve as the main connectors among sub-themes. Supporting keywords such as deep learning, performance prediction, apriori algorithm, discriminant analysis, and data mining are distributed around these main nodes, illustrating the variety of computational techniques used, while sport domain keywords (basketball, athletes, talent development) occupy positions connecting the techhnical side with the application side.

Table 4. Thematic Keywords Clusters (VOSviewer, 9 clusters)

Cluster	Number of Items	Members
1	6	Anthropometry, Biological maturation, Somatotype, Sports medicine, Talent development, Youth sports
2	5	Athletes, Neural network, Performance, Sport, Talent
3	5	Artificial Intelligence, Bias, Classification, Sports performance, Support vector machine
4	5	Athlete selection, Basketball, Big data, Data mining, Discriminant analysis
5	5	Apriori algorithm, Performance analysis, Performance prediction, Random Forest, Sports talent
6	4	Artificial neural network, Football, Multilayer perception, Soccer
7	4	Change of direction speed, Machine learning, Talent identification, Talent selection
8	4	Deep learning, Neural network, Rowing, Sports
9	3	Player performance metric, Player recruitment, Sports analysis

Referring to Table 4, Cluster 7 (machine learning, talent identification, talent selection, change of direction speed) functions as the thematic core linking methodological terms to this study's core domain terms. Three clusters show groupings based on a specific sport combined with a particular computational technique: Cluster 4 (basketball with data mining, discriminant analysis), Cluster 6 (soccer with artificial neural networks), and Cluster 8 (rowing with deep learning); this pattern thus reinforces the finding of research fragmentation by sport discussed in paragraphs 4-5 of the Introduction. Cluster 3 (artificial intelligence, bias, classification, support vector machine) indicates that a small portion of the literature has begun to raise the issue of algorithmic fairness (bias) in AI based sports performance assessment a relevant direction for future research. Cluster 1 (anthropometry, biological maturation, somatotype, sports medicine, talent development, youth sports) contains no AI/ML terms at all, confirming that conventional TID variables remain a data foundation even as research increasingly shifts toward AI based approaches.



Figure 5. Country Collaboration Network (Co-Authorship)

The country level co-authorship map shows that research contributions are dominated by a small number of countries, with China appearing most prominently, followed by Malaysia, Germany, Canada, Australia, Italy, and the UK. This distribution indicates that research collaboration on AI in sports talent identification remains concentrated in Eas Asia and parts of Europe/North America, with very limited representation from other regions (including Southeast Asia/Indonesia). Confirming the geographic gap already identified by Parra-Martinez & Wai, (2023) at the broader talent research level.



Figure 6. Reference Co-Citation Analysis

The reference co-citation analysis identified a number of core reference works (including works discussing machine learning, big data and analytics, and talent identification) that are consistently cited together across the document corpus. The emergence of these references as major co-citation nodes shows that the literature in this field still relies on a small number of "anchors" references a pattern commonly found in relatively young fields of study that do not yet have a broad, well established literature base.

3.2 Discussions

3.2.1 Productivity trends

The low output in 2016–2020 followed by growth from 2022 suggests that AI applications in sports talent identification have only recently expanded, which is in line with reviews reporting limited and uneven adoption of advanced computational techniques until 2020 (Khan et al., 2023) and rapid growth of machine learning applications more recently (Tang et al., 2026). Individual studies applying classification algorithms to sporting contexts illustrate this trend (Utomo & Wiradinata, 2023), and wider AI uptake in sports science, supported by wearable technology and tracking databases, may also contribute (Dwyer et al., 2022; Wilson et al., 2026). Because the corpus is small (83 documents), year-to-year changes are sensitive to a few publications, and growth should not be read as evidence of a paradigm change by it self.

3.2.2 Source and authors

Output is dispersed across many journals, including information engineering and sports management outlets (Utomo & Wiradinata, 2023), which is consistent with the cross-disciplinary nature of the topic. The author pattern is dominated by a small network around Hohmann, Pion, and Lenoir, which was active from 2017, while most other authors appear only from 2019 onward. Indonesian studies cited in this paper (Buhari et al., 2024, 2025) do not appear in the analyzed corpus.

3.2.3 Geographic distribution and Indonesia

The prominence of China, Germany, and the United Kingdom in total production, and of China, Australia, and the United Kingdom among corresponding authors, likely reflects established sport science research capacity, but this study did not test the reasons. China's low MCP ratio (0.20) indicates that its output is mostly domestic, whereas several countries with smaller output show higher international collaboration. Indonesia was among four countries tied for fourth by total production (six documents) and had four corresponding-author articles. Among the six Indonesia-affiliated documents, only one applied deep learning (an e-sports study), and none of the sport talent studies did, which is consistent with the earlier review by the present authors (Putra & Fauziah, 2026). This suggests that Indonesia output is concentrated in conventional machine learning, which leaves room for research using more advanced techniques within the Long-Term Athlete Development framework (Bakhtiar et al., 2023).

3.2.4 Thematic structure

The motor themes show a concentration of studies on performance metrics and recruitment analytics. The basic themes show that talent identification, machine learning, and AI form the connecting core of the field. Anthropometry and biological maturation appear as niche themes, suggesting that they are developed in separate lines of work with limited link to the machine learning cluster. This indicates that AI is used alongside, rather than instead of, established indicators, which is a pattern in our data and not a test of their relative value. Variation in biological maturity can distort identification outcomes if uncontrolled (Junior et al., 2021). The position of talent selection among emerging or declining themes may reflect conceptual overlap between identification, selection, detection, and development (Zhao et al., 2024).

3.2.5 Thematic evolution

This shift from data mining and discriminant analysis toward neural networks, artificial intelligence, and deep learning is consistent with the transition toward more advanced computational techniques described in prior reviews (Khan et al., 2023; Tang et al., 2026), and the present mapping adds a bibliometric view of that change within the corpus. Classification algorithms applied to football player position prediction (Utomo & Wiradinata, 2023) illustrate how computational approaches may complement coaches' judgment. In Indonesia, computerized systems based on biomotor and anthropometric data have been described (Buhari et al., 2024), though the studies reviewed did not use deep learning (Putra & Fauziah, 2026).

3.2.6 Keyword co-occurrence network

Talent identification, machine learning, and artificial intelligence act as the main connecting nodes. Three clusters link a sport to a computational technique, which indicates some sport-specific concentration, in line with the single-sport focus of earlier reviews (Han et al., 2023; Nijenhuis et al., 2024; Reyaz et al., 2022). The keyword "bias" appears in two documents with different meanings (decision-making bias and bias in AI-supported selection), so it cannot by itself show that algorithmic fairness is an emerging topic. Methodological problems such as sample limits, imbalance, and outcome definition can limit the interpretability of predictive models (Naughton et al., 2024; Tang et al., 2026), supporting the need for explainable and validated approaches.

3.2.7 Cited References and Intellectual Base

The most cited references are classic works on talent identification and development in sport rather than machine learning methodology. Vaeyens et al. (2008), Pion et al. (2015), Meylan et al. (2010), Williams and Reilly (2000), and Pion et al. (2017) were each cited by seven of the 83 documents, whereas Bunker and Thabtah (2019) and Breiman (2001) were cited by five and four documents. This suggests that AI-based talent identification research rests mainly on the sport science tradition and adopts machine learning methods from broader sports analytics rather than building talent-specific methods. Several of the most cited works are by Pion and colleagues, who also appear among the most productive authors, so part of this ranking may reflect citations within one research network. Because the counts are small, the ranking is indicative only.

3.2.8 Limitations and future research

First, only Scopus was used, and only English-language articles were included. Second, conference papers and reviews were excluded (86 of 180 records, 48%); because proceedings are important in AI research, relevant work may be missing. Third, the search terms did not include terms such as "talent development", "talent prediction", or "scouting", which may limit coverage. Fourth, the corpus is small (n = 83), so the thematic map and cited-reference results should be read as exploratory. Fifth, 2026 is a partial year. Sixth, studies from developing countries are sometimes published in regional journals that are not widely indexed (Buhari et al., 2024, 2025). Seventh, author keywords were not harmonized with a thesaurus, so spelling variants (e.g., neural network/neural networks, soccer/football) appear as separate items and may fragment clusters; six documents had no author keywords. Eighth, some of the 83 retained documents are only loosely related to sports talent identification, for example studies of decision bias in a professional league expansion draft, e-sports performance, sport technology systems, health screening of athletes, and a bibliometric analysis of somatotype. They were retained because they matched the search terms and passed the topical screening, which removed only records that used "talent identification" outside sport (seven records). Keyword clusters, country counts, and author rankings therefore include some peripheral topics, and results that rest on e few such documents (for example, they keyword "bias") should be read with caution. Future research should standardize terminology for identification, selection, and development (Zhao et al., 2024), examine how bias is defined and handled in AI-based assessment, strengthen international collaboration, and develop integrated frameworks combining biological, physical, technical, and contextual variables (Bjørndal & Ronglan, 2021; Cahill & MacNamara, 2024).

4. CONCLUSION

This study maps the structure and direction of research development on the application of AI in sports talent identification through 83 Scopus documents (2016-2026), and finds that this field has only recently entered a growth phase following a gradual period from 2016-2020, with a productivity structure still dominated by a long tail pattern across journal sources, authors, and citation networks. Geographically, this research remains concentrated in China, the United Kingdom, Germany, and Australia, although Indonesia shows

a fairly significant contribution consistent with the positioning of talent identification as the foundation of the Long-Term Athlete Development program within the National Sports Grand Design, even though, as a companion national level review shows, this contribution has not yet extended to deep learning bassed AI specifically. Thematically, this study confirms a genuine transition from conventional anthropometric or biological approaches toward computational techniques such as machine learning and data mining, although conventional biological variables remain a data foundation that has not been fully replaced, and although the concepts of identification, selection, detection, and talent development still lack consistent terminological boundaries across studies.

5. ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not for profit sectors.

6. AUTHOR CONTRIBUTION STATEMENT

M.E.P. was responsible for data processing and anlysis, and drafted the article in Indonesian. EF collected the data and drafted the article in English. Both authors reviewed and approved the final version of the manuscript.

AUTHOR INFORMATION

Corresponding Authors

Martinez Edison Putra, Universitas Katolik Santo Agustinus Hippo, Indonesia.
https://orcid.org/0009-0009-3445-6974
Email: m.putra@sanagustin.ac.id

Authors

Ella Fauziah, Universitas Katolik Santo Agustinus Hippo, Indonesia
https://orcid.org/0000-0001-6886-6802
Email: e.fauziah@sanagustin.ac.id

REFERENCES

Bakhtiar, S., Syahputra, R., Putri, L. P., Mardiansyah, A., Atradinal, Hendrayana, A. A., Afrian, H., Mardela, R., & Pion, J. (2023). Sports talent profile of 7-12 years old: Preliminary study of talent identification in Indonesia. *Journal of Physical Education and Sport*, *23*(12), 3167–3177. https://doi.org/10.7752/jpes.2023.12361

Barraclough, S., Till, K., Kerr, A., & Emmonds, S. (2022). Methodological Approaches to Talent Identification in Team Sports: A Narrative Review. In *Sports* (Vol. 10, Issue 6, p. 81). https://doi.org/10.3390/sports10060081

Bartlett, Jonathan D, & Murray, Andrew. (2025). An assessment of anthropometric and physical profiles of highly trained youth male and female basketball athletes: Implications for talent selection and development. *International Journal of Sports Science & Coaching*, *20*(3), 1037–1046. https://doi.org/10.1177/17479541251333885

Bergkamp, T. L. G., Niessen, A. S. M., Hartigh, R. J. R. Den, Frencken, W. G. P., & Meijer, R. R. (2019). Methodological Issues in Soccer Talent Identification Research. *Sports Medicine*, *49*(9), 1317–1335. https://doi.org/10.1007/s40279-019-01113-w

Bjørndal, C. T., & Ronglan, L. T. (2021). Engaging with uncertainty in athlete development – orchestrating talent development through incremental leadership. *Sport, Education and Society*, *26*(1), 104–116. https://doi.org/10.1080/13573322.2019.1695198

Buhari, M. R., Hamdiana, Ismawan, H., & Khumaini, M. H. F. (2025). Digital Technology and Local Policy : An Evidence-Based Collaborative Model for Sports Talent Identification. *Journal of Coaching and Sports Science*, *4*(2), 161–175. https://doi.org/10.58524/jcss.v4i2.903

Buhari, M. R., Sapulete, J. J., & Hasanuddin, M. I. (2024). Smart Sports Recruitment: Leveraging Software for Talent Precision. *Journal of Coaching and Sports Science*, *3*(2), 135–152. https://doi.org/10.58524/jcss.v3i2.529

Cahill, G., & MacNamara, Á. (2024). Contextual factors in sport talent identification research. *Sport in Society*, *27*(7), 1180–1216. https://doi.org/10.1080/17430437.2023.2288149

Changqing, X., Kamalden, T., Fadilah, T., Liu, H., & Ismail, N. (2022). Exploring the Multidisciplinary Factors Affecting Sports Talent Identification. *Frontiers in Psychology*, *13*(July), 1–14. https://doi.org/10.3389/fpsyg.2022.948121

Cossich, V. R. A., Carlgren, D., Holash, R. J., & Katz, L. (2023). applied sciences Technological Breakthroughs in Sport : Current Practice and Future Potential of Artificial Intelligence , Virtual Reality , Augmented Reality , and Modern Data Visualization in Performance Analysis. *Applied Sciences*, *13*(23), 12965. https://doi.org/10.3390/app132312965

Dwyer, D. B., Kempe, M., & Knobbe, A. (2022). Editorial: Using Artificial Intelligence to Enhance Sport Performance. *Frontiers in Sports and Active Living*, *Volume 4-2022*. https://www.frontiersin.org/journals/sports-and-active-living/articles/10.3389/fspor.2022.886730

Eetvelde, V., Exp, J., Eetvelde, H. Van, 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.

https://doi.org/10.1186/s40634-021-00346-x

Fernandez Ortega, J. A., Rodriguez Buitrago, J. A., & Sanchez Rodriguez, D. A. (2021). Aspectos centrales de la identificación y desarrollo de talentos deportivos: revisión sistemática (Central aspects of the identification and development of sports talents: a systematic review). *Retos, 39 SE-Re*, 915–928. https://doi.org/10.47197/retos.v0i39.79323

Han, M., Calvo, A. L., & Calvo, J. L. (2023). Basketball talent identification: a systematic review and meta-analysis of the anthropometric, physiological and physical performance factors. *Frontiers in Sport and Active Living*, *5*(November). https://doi.org/10.3389/fspor.2023.1264872

Hauser, L.-L., Harwood, C. G., Höner, O., O'Connor, D., & Wachsmuth, S. (2024). Talent development environments within sports: a scoping review examining functional and dysfunctional environmental features. *International Review of Sport and Exercise Psychology*, *17*(2), 1105–1131. https://doi.org/10.1080/1750984X.2022.2129423

Henriksen, K., & Stambulova, N. (2023). The social environment of talent development in youth sport. *Frontiers in Sports and Active Living*, *5*. https://doi.org/10.3389/fspor.2023.1127151

Junior, D. B. R., Werneck, F. Z., Oliveira, H. Z., & Panza, P. S. (2021). From Talent Identification to Novo Basquete Brasil (NBB): Multifactorial Analysis of the Career Progression in Youth Brazilian Elite Basketball. *Frontiers in Psychology*, *12*(March), 1–12. https://doi.org/10.3389/fpsyg.2021.617563

Kelly, A. L., Williams, C. A., Cook, R., Sáiz, S. L. J., & Wilson, M. R. (2022). A Multidisciplinary Investigation into the Talent Development Processes at an English Football Academy: A Machine Learning Approach. *Sports (Basel, Switzerland)*, *10*(10). https://doi.org/10.3390/sports10100159

Khan, N. J., Ahamad, G., & Naseem, M. (2023). Computational efficiency in sports talent identification – a systematic review. *International Journal Applied Decision Sciences*, *16*(3), 358–384. https://doi.org/10.1504/IJADS.2023.130600

Larkin, P. (2025). Talent identification of 12-year old male Australian rules footballers : Physical advantages and prognosis for junior and senior national-level selection. *PLoS ONE*, 1–12. https://doi.org/10.1371/journal.pone.0317336

Naughton, M., Salmon, P. M., Compton, H. R., & McLean, S. (2024). Challenges and opportunities of artificial intelligence implementation within sports science and sports medicine teams. *Frontiers in Sports and Active Living*, *Volume 6-2024*. https://doi.org/10.3389/fspor.2024.1332427

Nijenhuis, S. B., Koopmann, T., Mulder, J., Elferink-Gemser, M. T., & Faber, I. R. (2024). Multidimensional and Longitudinal Approaches in Talent Identification and Development in Racket Sports : A Systematic Review. *Sports Medicine - Open*, *10*(4). https://doi.org/10.1186/s40798-023-00669-2

Parra-Martinez, F. A., & Wai, J. (2023). Talent identification research: a bibliometric study from multidisciplinary and global perspectives. *Frontiers in Psychology*, *14*(May), 1–22. https://doi.org/10.3389/fpsyg.2023.1141159

Prieto-ayuso, A., Pastor-vicedo, J. C., & Gonz, S. (2020). Are Physical Education Lessons Suitable for Sport Talent Identification ? A Systematic Review of the Literature. *International Journal of Environmental Research and Public Health*, *17*(6). https://doi.org/10.3390/ijerph17061965

Puce, L., Żmijewski, P., Cotellessa, F., Schenone, C., Ceylan, H. I., Bragazzi, N. L., & Trompetto, C. (2026). The role of artificial intelligence in sports training: opportunities, challenges and future applications for competitive swimming. *Biology of Sport*, *43*, 355–367. https://doi.org/10.5114/biolsport.2026.152352

Putra, M. E., & Fauziah, E. (2026). Pemetaan Riset Identifikasi Bakat Olahraga Berbasis Teknologi di Indonesia : Tinjauan Sistematis Berbasis Scopus. *Jurnal Ilmiah Multidisiplin Ilmu*, *3*(4), 222–229. https://doi.org/10.69714/yegdgj07

Reyaz, N., Ahamad, G., Khan, N. J., & Naseem, M. (2022). Machine Learning in Sports Talent Identification : A Systematic Review. *Conference: 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET)*, 15–16. https://doi.org/10.1109/ICEFEET51821.2022.9848247

Shahidi, S. H., Carlberg, B., & J, D. K. (2023). Talent Identification and Development in Youth Sports: A Systematic Review. *International Journal of Kinanthropometry*, *3*(1 SE-Articles), 73–84. https://doi.org/10.34256/ijk2318

Silvino, V. O., da Silva Sousa, L. G., Ferreira, C. P., dos Santos, L. H. O., Apaza, H. M., Almeida, S. S., & dos Santos, M. A. P. (2025). The Use of Machine Learning in Sports Performance: A Systematic Review. *Translational Journal of the American College of Sports Medicine*, *10*(2). https://doi.org/10.1249/TJX.0000000000000304

Sulistyowati, E. M., Suherman, W. S., Sukamti, E. R., Orhan, B. E., & Geantă, V. A. (2025). Developing an Android-Based Test for Identifying Rhythmic Gymnastics Potential in Young Children: Physical and Psychological Influences. *International Journal of Human Movement and Sports Sciences*, *13*(1), 202–214. https://doi.org/10.13189/saj.2025.130122

Susanto, E., Bayok, M., Satriawan, R., Festiawan, R.,

Kurniawan, D. D., & Putra, F. (2023). Talent Identification Predicting in Athletics: A Case Study in Indonesia. *Annals of Applied Sport Science*, *11*(1), 1–11. https://doi.org/10.52547/aassjournal.1102

Susanto, N., García-Jiménez, J. V., Nowak, A. M., Setyawan, H., Pavlovic, R., Rusdiawan, A., & Syaukani, A. A. (2024). Development Assessment Model for Talent Identification of Young Indonesian Basketball Players: Anthropometrics, Biomotor, Technical, and Tactical Skills. *International Journal of Human Movement and Sports Sciences*, *12*(4), 625–635. https://doi.org/10.13189/saj.2024.120403

Tang, Q., Wei, X., & Tan, B. (2026). The Role of Machine Learning in Talent Identification for Team Sports : A Systematic Review. *Journal of Sports Science and Medicine*, *25*(November 2025), 58–83. https://doi.org/doi.org/10.52082/jssm.2026.58

Till, K., & Baker, J. (2020). Challenges and [Possible] Solutions to Optimizing Talent Identification and Development in Sport. *Frontiers in Psychology*, *11*(April), 1–14. https://doi.org/10.3389/fpsyg.2020.00664

Utomo, K. S., & Wiradinata, T. (2023). Optimal Playing Position Prediction in Football Matches: A Machine Learning Approach. *International Journal of Information Engineering and Electronic Business*, *15*(6), 30–47. https://doi.org/10.5815/ijieeb.2023.06.03

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. https://doi.org/10.1186/s40634-021-00346-x

Wilson, Vinu, Kabeer, Dilshith Azeezul, Narippatta Kappoor, Ashif Ali, Annar Kandi, Safad, Kalambankottarath, Shareef, Mattuchira Kumaran, Aravind, G, Manibharathi, Murugan, Pramitha, KA, Husaif, & Govindasamy, Karuppasamy. (2026). Artificial intelligence in sports: A systematic review of performance and injury prevention. *International Journal of Sports Science & Coaching*, 17479541261450376. https://doi.org/10.1177/17479541261450376

Yuwono, C., Billiandri, B., Putri, D. T., & Adi, S. (2025). A Machine Learning - Based Framework for Optimizing Sports Talent Identification and Development. *Cuestiones de Fisioterapia*, *54*(2), 3934–3952. https://doi.org/10.48047/8sxdfm81

Zhao, J., Xiang, C., Fadilah, T., & Kamalden, T. (2024). Differences and relationships between talent detection, identification, development and selection in sport: A systematic review. *Heliyon*, *10*(6), e27543. https://doi.org/10.1016/j.heliyon.2024.e27543