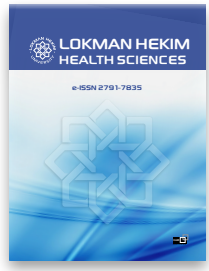
LOKMAN HEKIM HEALTH SCIENCES

DOI: 10.14744/lhhs.2026.66001
Lokman Hekim Health Sci 2026;6(2):264–271



lokmanhekimhs.com

ORIGINAL ARTICLE

Publication Trends and Citation Patterns of Artificial Intelligence and Large Language Model Research in Orthopedics and Traumatology: A Türkiye-based Bibliometric Analysis

Mehmet Can Gezer, Mahircan Demir

Department of Orthopedics and Traumatology, Mamak State Hospital, Ankara, Türkiye

Abstract

Introduction: Artificial intelligence (AI), particularly large language models (LLMs) and generative systems such as ChatGPT, has rapidly transformed medical research and clinical practice. Although global publication trends have been widely examined, country-specific bibliometric analyses in orthopedics and traumatology remain limited. This study aimed to perform a comprehensive bibliometric evaluation of Türkiye-based AI and LLM research in orthopedics and traumatology and compare the findings with global trends.

Methods: A cross-sectional bibliometric analysis was conducted in January 2026 using PubMed, TR Dizin, Scopus, and Web of Science databases. Original research articles from Türkiye-affiliated institutions involving AI, machine learning, deep learning, or large language model applications in orthopedics and musculoskeletal research were included. Journal Impact Factor (2024 JCR), citation counts, and SCImago Journal Rank (SJR) quartile classifications were recorded. Descriptive statistical analyses were performed.

Results: A total of 63 studies were included. Publication volume increased markedly after 2023. The mean Journal Impact Factor was 2.34±1.15 (median: 2.20; range: 0.8–5.4). Of the publications, 38.1% were in Q1 and 46.0% in Q2 journals, with none in Q4. The total citation count was 578, with a mean of 9.17±14.20 (median: 5; range: 0–82), reflecting a right-skewed distribution. Patient education studies were numerically predominant (n=34, 53.9%), while clinical application studies demonstrated the highest mean citation count (10.5±18.1).

Discussion and Conclusion: AI-based research in orthopedics and traumatology in Türkiye has expanded rapidly since 2023, predominantly in Q1–Q2 journals. Citation patterns show considerable heterogeneity, consistent with an early developmental phase. These findings provide an objective bibliometric profile and may inform country-specific scientific strategies in AI-driven orthopedic research.

Keywords: Artificial intelligence; Bibliometrics; ChatGPT; Orthopedics; Surgery

Cite this article as: *Gezer MC, Demir M. Publication trends and citation patterns of artificial intelligence and large language model research in orthopedics and traumatology: A Türkiye-based Bibliometric Analysis. Lokman Hekim Health Sci 2026;6(2):264–271.*

Correspondence: Mahircan Demir, M.D. Mamak Devlet Hastanesi, Ortopedi ve Travmatoloji Kliniği, Ankara, Türkiye
E-mail: mahir19911991@gmail.com **Submitted:** 17.03.2026 **Revised:** 03.05.2026 **Accepted:** 25.05.2026 **Available Online:** 11.06.2026

Artificial intelligence (AI) has evolved over the past decade from an experimental computational approach into a structural component of healthcare systems.[1,2] Machine learning, deep learning, and particularly large language models (LLMs) have generated a significant paradigm shift in modern medicine by enabling the analysis of high-dimensional clinical data, interpretation of imaging findings, comprehension of unstructured medical texts, and context-aware responses to complex clinical scenarios. Image-based diagnostic algorithms, risk prediction models, and natural language processing–driven systems have strengthened clinical decision-support mechanisms and accelerated the digital transformation of healthcare processes.[3,4]

With the widespread adoption of generative AI applications, transformation has extended beyond diagnostic and analytical domains to include scientific writing, academic productivity, patient education, and clinical decision-support systems.[5,6] This development indicates that AI is no longer merely an auxiliary tool but has become an active stakeholder in healthcare delivery and the research ecosystem.[7]

Orthopedics and traumatology are among the surgical disciplines in which AI applications have been rapidly integrated. Deep learning–based models have demonstrated promising levels of accuracy and efficiency in areas such as automated fracture classification, assessment of spinal deformities, arthroplasty planning, implant survival prediction, and interpretation of sports injuries. More recently, the evaluation of LLMs for generating responses to patient inquiries, assessing examination performance, providing clinical decision suggestions, and editing scientific manuscripts suggests that digital transformation in orthopedics has entered a new phase.[8–12]

However, the need for systematic evaluation of these technologies in terms of accuracy, methodological rigor, clinical applicability, ethical considerations, data security, and scientific impact has been increasingly emphasized in the literature. In particular, concerns regarding the potential for generative AI systems to produce "hallucinated" outputs, as well as issues related to source reliability and clinical validity, underscore the necessity for discipline-specific analyses.

In the international literature, AI-based publications in the field of musculoskeletal research have gained notable momentum, particularly in late 2022. Following the public release of generative AI platforms, studies evaluating diagnostic accuracy, comparative clinical performance, quality of patient information, and educational reliability have increased rapidly.[12] However, existing bibliometric analyses indicate that AI-related research has been predominantly produced in North America, Western Europe, and East Asia, while contributions from developing countries have been analyzed only to a limited extent. Evaluating country-specific scientific positioning is important not only in terms of research output but also for assessing publication quality and international visibility.[13]

A noticeable increase in AI-related publications has also been observed in Türkiye-based orthopedics and traumatology research in recent years. Nevertheless, there is no comprehensive evaluation of the impact level of the journals in which these studies were published, their quartile distribution, citation performance, or their relative positioning within the global literature. In the current literature, Türkiye's contribution in this field is largely interpreted through quantitative growth, whereas comparative analyses based on objective metrics of publication quality and scientific impact remain limited. This gap represents an important shortcoming in the development of country-specific strategic science policies.

Bibliometric analysis is a powerful methodological approach that enables the quantitative evaluation of scientific production within a specific thematic field through objective indicators.[12] Publication counts, citation metrics, journal impact indicators, quartile distributions, and temporal production trends provide an objective framework for assessing the scientific visibility and academic positioning of a research community. Particularly in rapidly evolving and interdisciplinary domains, bibliometric evaluations play a crucial role in identifying scientific orientation, quality trends, and potential areas for development.[13]

Although ChatGPT is one of the most prominent examples of generative AI, the present study was designed to include a broader spectrum of AI and LLM-based research to provide a more comprehensive bibliometric evaluation.

This study aimed to conduct a comprehensive bibliometric analysis of AI-based publications in orthopedics and traumatology affiliated with institutions in Türkiye. Accordingly, publication trends over time, journal quartile distributions (Q1–Q4), impact factor characteristics, and citation performance were evaluated. The findings were compared with general trends reported in the global literature in order to objectively determine Türkiye's scientific positioning in AI-driven orthopedic research.

Materials and Methods

This study was designed as a cross-sectional and descriptive bibliometric analysis to evaluate the bibliometric characteristics of AI–based publications in orthopedics and traumatology affiliated with institutions in Türkiye. This study was based on publicly available data and did not require ethical approval. The study selection process was reported using a PRISMA 2020 flow diagram (Fig. 1).

The literature search was performed in January 2026 using the PubMed, TR Dizin, Scopus, and Web of Science databases. For bibliometric verification, Journal Impact Factors (JIFs) and citation data were obtained from the Web of Science Core Collection, while quartile classifications were retrieved from the SCImago Journal Rank (SJR) database.

The search strategy was developed to capture studies related to AI, machine learning, deep learning, LLMs, generative AI, and ChatGPT within the fields of orthopedics, traumatology, and the musculoskeletal system.

No country- or affiliation-based restriction was applied during the initial database search. Instead, eligibility regarding Türkiye affiliation was determined after record retrieval through manual screening of institutional addresses. Publications authored by Turkish researchers but not affiliated with institutions in Türkiye were excluded.

The search strategy was designed to identify studies related to AI, machine learning, deep learning, LLMs, generative AI, and ChatGPT within the fields of orthopedics, traumatology, and the musculoskeletal system.

Database-specific search strategies were adapted according to the indexing structure of each database. In TR Dizin, both Turkish and English keywords were used to enhance search sensitivity.

The complete search strategies for all databases are provided in Supplementary File 1.

Search results were filtered to include all available years without language restriction. No publication year restriction was applied during the search process. Due to substantial overlap across databases, duplicate records were identified and removed through both automated and manual screening. A total of 356 records were identified across all databases. After removal of duplicate records and exclusion of non-Türkiye affiliations, 133 records remained. Following screening, 63 studies were included. Following duplicate removal, records were screened based on title, abstract, and full text. In addition, institutional affiliations were evaluated to determine eligibility according to predefined inclusion criteria. The number of records

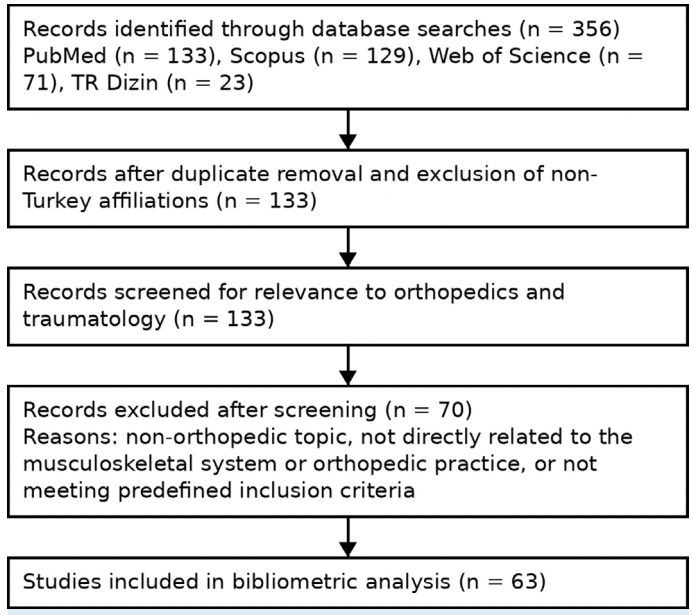


Figure 1. PRISMA 2020 flow diagram of the study selection process.

retrieved from each database and the complete search strategies for all databases are provided in Supplementary File 1 to ensure full reproducibility.

All identified records were independently screened at the title and abstract level by two researchers. Studies considered potentially eligible were subsequently assessed through full-text review. In cases of disagreement, consensus was achieved through discussion and, when necessary, consultation with a third reviewer.

Studies were included if they were conducted at institutions affiliated with Türkiye, were directly related to orthopedics, traumatology, or the musculoskeletal system, involved AI or related computational approaches, and were published as original research articles in peer-reviewed journals.

Exclusion criteria comprised letters to the editor, review articles, case reports, conference abstracts, studies unrelated to orthopedics, and publications not affiliated with institutions in Türkiye.

For each included study, the following variables were recorded: year of publication, journal name, country of publication, national or international classification of the journal, JIF (2024 JCR data), total citation count in the Web of Science, SJR quartile classification (Q1–Q4), and primary thematic category.

JIF values were obtained from the Journal Citation Reports (JCR), while quartile classifications were derived from the SJR database. These metrics were used complementarily to provide both impact- and ranking-based perspectives. Quartile classification was determined according to the

Table 1. Most cited studies included in the analysis

Study title	Citation count	IF	Quartile
ChatGPT performance in the medical specialty exam: An observational study	82	1.4	Q2
Use of Deep Learning Methods for Hand Fracture Detection from Plain Hand Radiographs	43	1.0	Q3
Large Language Models in Healthcare: A Bibliometric Analysis	45	2.4	Q2
Evaluating DeepResearch and DeepThink in Anterior Cruciate Ligament Surgery Patient Education: ChatGPT-4o Excels in Comprehensiveness, DeepSeek R1 Leads in Clarity and Readability of Orthopaedic Information	29	5.0	Q1
ViVGG19: Novel Exemplar Deep Feature Extraction-Based Shoulder Rotator Cuff Tear and Biceps Tendinosis Detection Using Magnetic Resonance Images	22	2.3	Q3
ChatGPT Can Offer At Least Satisfactory Responses to Common Patient Questions Regarding Hip Arthroscopy	22	5.4	Q1
Artificial Intelligence, Machine Learning, and Deep Learning in Orthopedic Surgery	21	1.0	Q2
Achieving High Accuracy in Meniscus Tear Detection Using Advanced Deep Learning Models with a Relatively Small Data Set	19	5.0	Q1
The Diagnosis of Developmental Dysplasia of the Hip From Hip Ultrasonography Images With Deep Learning Methods	16	1.5	Q2

IF: Impact factor; Q: SCImago journal rank.

journal's highest ranking within the orthopedics category or the most relevant subject category.

Data extraction was performed independently by two reviewers. Studies were categorized into three predefined thematic groups: patient-related applications, examination/education-related studies, and clinical applications. The classification process was conducted independently by two reviewers. In studies encompassing multiple thematic domains, categorization was determined based on the predominant content and the primary outcome of the study. Discrepancies between reviewers were resolved through structured discussion, and when necessary, consensus was achieved with the involvement of a third reviewer.

Studies employing various AI approaches (e.g., machine learning, deep learning, and LLMs) were included, and no subtype-specific categorization was performed, as the aim was to evaluate overall bibliometric characteristics.

None of the included studies was authored or co-authored by the present investigators.

Statistical Analysis

The obtained data were analyzed using descriptive statistical methods. The distribution of continuous variables was assessed using both visual methods (histograms and Q–Q plots) and the Shapiro–Wilk test. Normally distributed variables were presented as mean±standard deviation (SD), whereas non-normally distributed variables were expressed as median (minimum–maximum) values.

Given the right-skewed nature of citation data, both mean and median values were reported to provide a comprehensive representation of central tendency. Categorical variables were presented as frequency (n) and percentage (%). Journal quartile distributions (Q1–Q4) were evaluated using frequency analysis.

Citation analysis included the total number of citations, mean and median citation counts per study, citation distribution according to quartile categories, and the proportion of uncited publications. The annual distribution of publications was illustrated graphically, and temporal trends were descriptively evaluated.

Due to the descriptive design of the study, no inferential statistical analyses were performed. All analyses were conducted using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA).

Results

A total of 356 records were identified through database searches, including PubMed (n=133), Scopus (n=129), Web of Science (n=71), and TR Dizin (n=23). Following the removal of duplicate records and studies not affiliated with institutions in Türkiye, 133 unique records remained for further evaluation. These records were screened for relevance to orthopedics and traumatology based on study titles, abstracts, and author affiliations. Studies not directly related to the musculoskeletal system or orthopedic practice were excluded. After eligibility assessment, 63 studies met the predefined inclusion criteria and were

Table 2. Bibliometric and Citation Analysis of Türkiye-Based Orthopedic AI and LLM-Based Studies (n=63)

	Total (n=63)	Patient education (n=34)	Exam/Board (n=9)	Clinical application (n=20)
Mean IF (±SD)	2.34±1.15	2.41±1.10	2.18±1.20	2.33±1.21
Median IF	2.20	2.30	2.00	2.20
IF (min–max)	0.8–5.4	0.9–5.4	0.8–4.7	0.8–5.4
Q1, n (%)	24 (38.1)	14 (41.2)	3 (33.3)	7 (35.0)
Q2, n (%)	29 (46.0)	15 (44.1)	4 (44.4)	10 (50.0)
Q3, n (%)	10 (15.9)	5 (14.7)	2 (22.2)	3 (15.0)
Q4, n (%)	0 (0)	0	0	0
Total citations	578	290	78	210
Mean citations (±SD)	9.17±14.20	8.53±12.5	8.66±7.2	10.5±18.1
Median citations	5	5	6	4
Citations (min–max)	0–82	0–60	0–21	0–82

IF: Impact factor; SD: Standard deviation; Q: SCImago Journal Rank; LLM: Large language models. Continuous variables are presented as mean±SD; categorical variables as n (%).

included in the final analysis. The study selection process is summarized in Figure 1.

The distribution of publications by year is presented in Table 1. The number of publications by year was as follows: 2019 (n=1), 2020 (n=1), 2021 (n=2), 2022 (n=4), 2023 (n=10), 2024 (n=25), and 2025 (n=20). The distribution of publications by year is presented in Figure 2.

All included studies were published in international peer-reviewed journals. According to the 2024 JCRs data, the mean JIF was 2.34±1.15, with a median of 2.20 (minimum: 0.8–maximum: 5.4).

The quartile distribution was as follows: Q1: 24 publications (38.1%), Q2: 29 publications (46.0%), and Q3: 10 publications (15.9%), while no studies were identified in Q4 journals. Publications in Q1–Q2 journals accounted for 84.1% of all studies (n=53).

The total number of citations was 578. The mean citation count per study was 9.17±14.20, with a median of 5 (minimum: 0–maximum: 82). Eleven studies (17.5%) had not received any citations. The most cited studies included in the analysis are summarized in Table 1. Citation counts among the top-cited studies ranged from 16 to 82, encompassing publications related to specialty examinations, patient education, deep learning applications, and bibliometric analyses.

The distribution of studies across thematic subgroups was as follows: patient-related queries and patient education (n=34), specialty examination/board assessment (n=9), and clinical applications and diagnostic performance (n=20).

Bibliometric and citation characteristics according to subgroup classification are presented in Table 2.

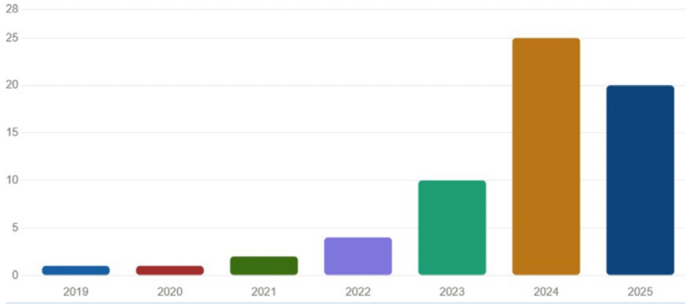


Figure 2. Annual distribution of included publications (n=63).

Discussion

In this study, the bibliometric profile of AI-based publications in orthopedics and traumatology affiliated with institutions in Türkiye was systematically analyzed. The findings indicate a noticeable increase in research output after 2023, with publications predominantly positioned in Q1–Q2 journals. The fact that 84.1 % of the studies were published in Q1–Q2 categories may suggest that Türkiye-based orthopedic research has achieved a certain level of visibility and journal selectivity, rather than demonstrating a definitive competitive advantage. Large-scale bibliometric analyses of the global ChatGPT literature have reported Q1 publication rates generally ranging between 30% and 50%.[12,13] For example, a Web of Science-based analysis examining 2,465 articles reported a Q1 rate of approximately 40%.[12] Similarly, a more recent analysis of 3,231 articles demonstrated that publications were predominantly concentrated in Q2–Q3 categories, with journals indexed in ESCI representing a substantial proportion.[13] Early Scopus-based analyses have also reported relatively higher proportions of

Q3–Q4 publications.[14] The international comparisons presented in this study are based on studies using different databases, time frames, and inclusion criteria. Therefore, these comparisons do not represent direct statistical comparisons and should be interpreted with caution in light of methodological heterogeneity across studies.

In this context, the observed Q1 rate of 38.1% in our study appears to be within the range reported in the literature, rather than positioning Türkiye in a clearly defined global segment. Notably, no publications were identified in Q4 journals. International reports have indicated that a proportion of studies published in early 2023 appeared in lower-impact or rapidly reviewed journals.[14] The absence of Q4 publications among Türkiye-based orthopedic studies may reflect differences in publication patterns; however, this observation should be interpreted with caution, given the limited sample size and the early developmental stage of the field. The mean Impact Factor of 2.34±1.15 indicates that publications are clustered in journals with moderate impact. In the global ChatGPT literature, mean Impact Factor values have been reported to concentrate within the 2–3 range,[12,15] with slightly higher averages observed in analyses focused on healthcare disciplines.[15] Therefore, the Impact Factor profile of Türkiye-based orthopedic publications appears broadly comparable with the broader health sciences literature. In addition, recent studies evaluating the global bibliometric landscape of AI research in orthopedics have also reported a marked increase in publication output in recent years, with most studies being published in journals with moderate-to-high impact levels. The findings of the present study are generally consistent with global research trends in AI within the field of orthopedics.[16]

Studies focused on patient education were observed to have relatively higher mean Impact Factor values, paralleling global thematic trends. International bibliometric analyses have identified "medical education," "ethics," and "patient communication" as emerging motor themes in the generative AI literature.[13,17] Moreover, publications centered on education and patient communication have been reported to achieve early visibility and citation traction.[18–20] This pattern may indicate earlier adoption of generative AI in domains involving natural language generation and patient communication. The overall mean citation count of 9.17±14.20, with a median of 5, reflects a right-skewed distribution. Given the right-skewed distribution of citation data, mean values may be influenced by a small number of highly cited studies. Therefore, median values were also reported to provide a more robust representation of central tendency, and both measures were considered in the interpretation of the findings. Similarly, global analyses have reported high variance in citation distribution, with a limited number of studies accounting for a substantial proportion of citations.[12,15,21] Early-phase bibliometric evaluations have generally reported single-digit mean citation values.[14,22]

The presence of studies receiving up to 82 citations suggests that some early-published studies have reached relatively high citation levels within the field. Conversely, the finding that 17.5% of publications had not yet received citations is consistent with the phenomenon of "citation window bias." Indeed, it has been reported that studies published after 2024 have, in the global literature as well, accumulated only limited citations to date.[13,23]

The higher mean citation rate observed for publications in Q1 journals is consistent with the commonly reported association between journal visibility and citation performance. Global bibliometric analyses have similarly demonstrated that ChatGPT-related publications in Q1 journals tend to accumulate citations more rapidly.[12,15,24] However, the observation of meaningful citation performance in Q2–Q3 journals suggests that multiple factors, including topical relevance, may contribute to citation patterns. In particular, healthcare-focused ChatGPT studies have been reported to achieve rapid citation uptake irrespective of JIF.[15,25] The thematic subgroup analysis indicates that research production in Türkiye initially evolved around patient education and information accuracy. This pattern aligns with global thematic network analyses, in which keyword co-occurrence networks have identified "education," "ethics," "accuracy," "hallucination," and "medical students" as dominant early-phase themes.[12,26] The relatively smaller number of studies focusing on clinical performance and diagnostic accuracy may reflect the current distribution of research topics within the dataset. Nevertheless, the higher mean citation value observed in this subgroup should be interpreted cautiously, as it may be influenced by variability and the presence of highly cited individual studies. International analyses have reported that clinically oriented or diagnostic accuracy studies demonstrate greater citation potential.[12,26]

The marked increase in publications after 2023 parallels the global production curve. However, the relative decline in the number of publications in 2025 should be interpreted with caution, as the literature search was conducted in January 2026 and some publications may not yet have been fully indexed across databases. In the international literature, 2023 has been characterized as the "year of exponential

growth" in ChatGPT-related publishing.[12,15] While the United States, China, and the United Kingdom have been reported as leading contributors in overall output, Türkiye's simultaneous adaptation within a discipline-specific context appears to follow this global trend.[13,15]

This study has several important limitations. First, the relatively small number of included studies, particularly within certain subgroups, may affect the reliability of the descriptive statistics (mean, SD, and percentage distributions). Therefore, the presented findings should be interpreted with caution, and the generalizability of the results is limited. Subgroup-level analyses should be considered exploratory, and further studies with larger sample sizes are needed to validate these findings. Another important limitation of this study is the time-dependent nature of citation analysis. Since the included publications span different years, more recent studies have had a shorter period to accumulate citations. This may result in comparatively lower citation counts for studies published in 2025–2026. Citation counts were not normalized by publication year, which may further contribute to this bias. Therefore, direct comparisons of citation counts across publication years should be interpreted with caution. The JIF and quartile classifications used in this study are journal-level metrics and may not directly reflect the scientific quality of individual articles. This limitation has also been emphasized in the San Francisco Declaration on Research Assessment. Therefore, these metrics should be interpreted with caution when evaluating the findings. In this study, the bibliometric analysis was limited to descriptive statistics, and network-based analyses (such as co-authorship, co-citation, and keyword co-occurrence mapping) were not included. The primary aim of the study was to present the fundamental bibliometric characteristics of AI-based orthopedic research affiliated with Türkiye. Network-based bibliometric analyses (such as co-authorship, co-citation, and keyword co-occurrence mapping) may provide additional insights and could be considered in future studies.

Conclusion

This bibliometric analysis demonstrates that AI-based research in orthopedics and traumatology in Türkiye has increased rapidly since 2023, with publications predominantly appearing in Q1–Q2 journals. Patient-related studies constitute the largest proportion of the literature, while citation counts show a heterogeneous distribution across studies. These findings reflect the current state of the field based on descriptive bibliometric indicators.

Ethics Committee Approval: This study was based on publicly available data and did not require ethical committee approval.

Conflict of Interest: None declared.

Financial Disclosure: The author declared that this study has received no financial support.

Use of AI for Writing Assistance: During the preparation of this manuscript, AI-based writing tools were used for language editing and text improvement. The scientific content, data analysis, and conclusions are entirely the responsibility of the authors.

Authorship Contributions: Concept: MCG, MD; Design: MCG, MD; Supervision: MCG., MD; Data collection and/or processing: MCG; Analysis and/or interpretation: MD; Literature review: MCG; Writing: MD; Critical review: MCG.

Peer-review: Double blind peer-reviewed.

References

1. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-18. [CrossRef]
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. [CrossRef]
3. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng 2017;19:221-48. [CrossRef]
4. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A 2018;115(45):11591-6. [CrossRef]
5. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. PLoS Med 2018;15(11):e1002699. [CrossRef]
6. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthop 2018;89(4):468-73. [CrossRef]
7. Seah JCY, Tang CHM, Buchlak QD, Holt XG, Wardman JB, Aimoldin A, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. Lancet Digit Health 2021;3(8):e496-e506. [CrossRef]
8. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198. [CrossRef]
9. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312. [CrossRef]
10. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating GPT as an adjunct for radiologic decision making: GPT-4

versus GPT-3.5 in a breast imaging pilot. J Am Coll Radiol 2023;20(10):990-7. [CrossRef]
11. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023;47(1):33. [CrossRef]
12. Nan D, Zhao X, Chen C, Sun S, Lee KR, Kim JH. Bibliometric analysis on ChatGPT research with CiteSpace. Information 2025;16(1):38. [CrossRef]
13. Koo M. ChatGPT research: a bibliometric analysis based on the Web of Science from 2023 to June 2024. Knowledge 2025;5(1):4. [CrossRef]
14. Farhat F, Silva ES, Hassani H, Madsen DØ, Sohail SS, Himeur Y, et al. The scholarly footprint of ChatGPT: a bibliometric analysis of the early outbreak phase. Front Artif Intell 2024;6:1270749. [CrossRef]
15. Liu L, Qu S, Zhao H, Kong L, Xie Z, Jiang Z, Zou P. Global trends and hotspots of ChatGPT in medical research: a bibliometric and visualized study. Front Med (Lausanne) 2024;11:1406842. [CrossRef]
16. Regmi A, Jain V, Baral S, Niraula BB. Global research landscape of artificial intelligence in orthopaedics: a decade of trends from a Scopus-based bibliometric analysis (2015-2025). J Clin Orthop Trauma 2026;73:103342. [CrossRef]
17. Khan N, Khan Z, Koubaa A, Khan MK, Salleh RB. Global insights and the impact of generative AI-ChatGPT on multidisciplinary fields: a systematic review and bibliometric analysis. Connect Sci 2024;36:1-50. [CrossRef]
18. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res 2021;133:285-96. [CrossRef]
19. Aria M, Cuccurullo C. Bibliometrix: an R-tool for comprehensive science mapping analysis. J Informetr 2017;11(4):959-75. [CrossRef]
20. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84(2):523-38. [CrossRef]
21. Garfield E. The history and meaning of the journal impact factor. JAMA 2006;295(1):90-3. [CrossRef]
22. Bornmann L, Daniel HD. What do we know about the h index? J Am Soc Inf Sci Technol 2007;58(9):1381-5. [CrossRef]
23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [CrossRef]
24. Huang X, Han F, Chen YF, et al. Bibliometric analysis of the application of artificial intelligence in orthopedic imaging. Quant Imaging Med Surg 2025;15(5):3993-4013. [CrossRef]
25. Bernstein E, Ramsamooj A, Millar KL, Lum ZC. Identification and categorization of the top 100 articles and the future of large language models: thematic analysis using bibliometric analysis. JMIR AI 2025;4:e68603. [CrossRef]
26. Zhang Y, Xie X, Xu Q. ChatGPT in medical education: bibliometric and visual analysis. JMIR Med Educ 2025;11:e72356. [CrossRef]

Supplementary File

Full Search Strategy and Database Records

1. Databases

The literature search was conducted in January 2026 using PubMed, Web of Science Core Collection, Scopus, and TR Dizin.

2. Search Strategies

PubMed:

("artificial intelligence"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "large language model"[Title/Abstract] OR "generative AI"[Title/Abstract] OR "ChatGPT"[Title/Abstract]) AND ("orthopedic"[Title/Abstract] OR "orthopaedic"[Title/Abstract] OR "orthopedics"[Title/Abstract] OR "orthopaedics"[Title/Abstract] OR "traumatology"[Title/Abstract] OR "musculoskeletal"[Title/Abstract])

Web of Science:

TS=("artificial intelligence" OR "machine learning" OR "deep learning" OR "large language model" OR "generative AI" OR "ChatGPT") AND TS=("orthopedic" OR "orthopaedic" OR "orthopedics" OR "orthopaedics" OR "traumatology" OR "musculoskeletal")

Scopus:

TITLE-ABS("artificial intelligence" OR "machine learning" OR "deep learning" OR "large language model" OR "generative AI" OR "ChatGPT") AND TITLE-ABS("orthopedic" OR "orthopaedic" OR "orthopedics" OR "orthopaedics" OR "traumatology" OR "musculoskeletal")

TR Dizin: Turkish and English keywords were used.

3. Record Counts

Database	Records Identified
PubMed	133
Web of Science	71
Scopus	129
TR Dizin	23

Total identified records across all databases: 133

4. Data Processing

Duplicate records were removed using both manual and automated screening.

5. Final Dataset

Final included studies: 63