Shi *et al. Systematic Reviews (2025) 14:62*
https://doi.org/10.1186/s13643-025-02779-2

Systematic Reviews

RESEARCH

Open Access

Global output of clinical application research on artificial intelligence in the past decade: a scientometric study and science mapping

Ji-yuan Shi[1,5†], Shu-jin Yue[1†], Hong-shuang Chen[2], Fei-yu Fang[4], Xue-lian Wang[3], Jia-jun Xue[4], Yang Zhao[6], Zheng Li[4*] and Chao Sun[1,3*]

Abstract

Background Artificial intelligence (AI) has shown immense potential in the field of medicine, but its actual effectiveness and safety still need to be validated through clinical trials. Currently, the research themes, methodologies, and development trends of AI-related clinical trials remain unclear, and further exploration of these studies will be crucial for uncovering AI's practical application potential and promoting its broader adoption in clinical settings.

Objective To analyze the current status, hotspots, and trends of published clinical research on AI applications.

Methods Publications related to AI clinical applications were retrieved from the Web of Science database. Relevant data were extracted using VOSviewer 1.6.17 to generate visual cooperation network maps for countries, organizations, authors, and keywords. Burst citation detection for keywords and citations was performed using CiteSpace 5.8.R3 to identify sudden surges in citation frequency within a short period, and the theme evolution was analyzed using SciMAT to track the development and trends of research topics over time.

Results A total of 22,583 articles were obtained from the Web of Science database. Seven-hundred and thirty-five AI clinical application research were published by 1764 institutions from 53 countries. The majority of publications were contributed by the United States, China, and the UK. Active collaborations were noted among leading authors, particularly those from developed countries. The publications mainly focused on evaluating the application value of AI technology in the fields of disease diagnosis and classification, disease risk prediction and management, assisted surgery, and rehabilitation. Deep learning and chatbot technologies were identified as emerging research hotspots in recent studies on AI applications.

Conclusions A total of 735 articles on AI in clinical research were analyzed, with publication volume and citation counts steadily increasing each year. Institutions and researchers from the United States contributed the most to the research output in this field. Key areas of focus included AI applications in surgery, rehabilitation, disease diagnosis, risk prediction, and health management, with emerging trends in deep learning and chatbots. This study also provides detailed and intuitive information about important articles, journals, core authors, institutions, and topics in the field through visualization maps, which will help researchers quickly understand the current status, hotspots,

†Ji-yuan Shi and Shu-jin Yue contributed equally to this work.

*Correspondence:
Zheng Li
zhengli@pumc.edu.cn
Chao Sun
sunchaobjyy@163.com
Full list of author information is available at the end of the article

and trends of artificial intelligence clinical application research. Future clinical trials of artificial intelligence should strengthen scientific design, ethical compliance, and interdisciplinary and international cooperation and pay more attention to its practical clinical value and reliable application in diverse scenarios.

Introduction

Artificial intelligence (AI), first proposed in 1956, has experienced substantial advancements and has become a key interdisciplinary and leading-edge scientific domain [1, 2]. Advancements in computing power, data storage, and algorithms have accelerated AI applications in medicine [3–5], resulting in a significant increase in AI publications over the past decade across areas such as disease prediction, medical imaging, health management, diagnosis, robotics, gene sequencing, nursing, and treatment decision-making [6–8], while emerging technologies like medical image diagnostics, clinical decision support systems, robotic assistants, and tools such as ChatGPT are revolutionizing healthcare and being applied in clinical practice to enhance efficiency; however, their effectiveness and clinical value still require further validation [2, 9–13].

Clinical trials are studies conducted on humans to evaluate the safety, efficacy, and side effects of new drugs, treatments, medical devices, or interventions, providing high-quality evidence of their effectiveness. With the growing potential of AI in medicine, an increasing number of researchers are conducting clinical trials to explore new applications of AI-related interventions [8, 14, 15], but the theme and development trend of these studies are not clear. Understanding the research progress and emerging hotspots in AI clinical trials is crucial for guiding both clinical practice and scientific research [15–17]. While qualitative and systematic reviews can provide in-depth analyses of specific aspects or subsets of AI clinical trials, their scope is often limited and may not fully capture the overall landscape or emerging trends in the entire field. In contrast, bibliometric analysis is a statistical and quantitative tool that extracts measurable data through the analysis of published research and the utilization of knowledge within these publications. Bibliometrics involves analyzing data such as the number of publications, distribution, citations, and keywords to reveal patterns, trends, and structures in scientific research activities. It has been widely applied across various research domains to evaluate research performance, assess academic impact, construct collaboration networks, and identify hot topics [18–20]. Previous bibliometric studies have explored AI research across diverse domains [7, 14, 21–27]. However, many of these studies include a significant number of review articles and nonclinical research, which may introduce bias and potentially exaggerate the current state of AI applications in certain areas [15]. To date, no bibliometric analysis has specifically focused on AI clinical trials. This study aims to fill this gap by providing a comprehensive overview of the field's hotspots and trends, evaluating academic impact, identifying research gaps, and offering data support to advance AI innovation and application in healthcare.

Bibliometric analysis was conducted to answer the following research questions of AI clinical trials: (1) Which are the top contributing countries, institutions, and authors? (2) What is the current status of international, institutional, and author collaboration in this field? (3) Which are the most influential clinical trials and top contributing journals? (4) What are the history, hot spots, and future trends of the research topic? It is believed that this bibliometric analysis can not only assist researchers in quickly identifying key articles, journals, potential collaborators, and institutions in this field but also provide valuable insights for selecting future research directions.

Methodology

Search strategy and studies selection

Web of Science is one of the most authoritative academic databases globally, covering multiple disciplines and providing comprehensive citation data and research tools, making it a preferred choice for bibliometric studies [19]. A comprehensive search was conducted in the Web of Science Core Collection database for clinical trials related to AI published between January 2012 and December 2022. The detailed search strategy is provided in Additional file 1: Appendix 1. The search terms were developed through repeated revisions and discussions with experts who have over 10 years of professional background in information retrieval. The main search terms included the following: "Artificial Intelligence" OR AI OR "Computational Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Machine Intelligence" OR "Neural Network*" OR "Natural Language Processing" OR "Computer Vision" OR "Predictive Modeling" OR "Reinforcement Learning" OR "Knowledge Representation*" OR "Knowledge Acquisition" OR "Sentiment Analysis" OR "Expert System*" OR "Fuzzy Logic" etc.) AND ("Clinical Trial*" OR "Randomized Controlled Trial*" OR "Randomised Controlled Trial*" OR RCT OR "Controlled Trial*" etc.).

Studies selection

Studies were included if they met the following criteria: the publication language was restricted to English, and the publication date was limited to the past 10 years; the research involved the application of AI technologies (e.g., natural language processing, machine learning, computer vision); the research focused on the field of medicine; and only clinical trials related to AI were considered. Studies that had not undergone peer review (e.g., protocols, correspondence, theses, and conference papers) were excluded. Titles and abstracts were screened independently by two reviewers to exclude records not meeting the inclusion criteria, with full texts reviewed as needed to determine eligibility. Any disagreements during the selection process were resolved through discussion with a third reviewer.

Data analysis

VOSviewer 1.6.17 software was used to analyze the network relationship of authors, countries, keywords, and organizations of the AI clinical trials. The size of the nodes reflects the number of publications or frequency, two nodes were connected by a line if they published the clinical trial together, while the color of the nodes represented different clusters [26–28]. Different expressions for the same author or keyword were standardized into a uniform expression to reduce bias in data analysis. CiteSpace 5.8 R3 software was used to perform burst detection for co-cited references and keywords [19]. The parameters for CiteSpace were set as follows: time slicing (2012–2022), years per slice (1), term source (all selections), node type (analyzed individually), selection criteria (top 50), pruning (none), and visualization (static cluster view with merged networks). Bursts detect, defined as items cited frequently within a specific period, were identified with the top 20 results per slice. Additionally, a dual-map overlay of journals was generated using CiteSpace [18, 29]. Descriptive statistics for the year of publication, number of citations, authors, countries, organizations, and journals were performed using Microsoft Excel 2016. The thematic evolution of AI clinical trials was analyzed using SciMAT (Science Mapping Analysis Software Tool). Keyword expressions were standardized for consistency through a combination of SciMAT's automatic and manual cleaning functions. The thematic evolution map was created by dividing the dataset into three time slices (2012–2016, 2016–2019, and 2019–2022), enabling an analysis of topic progression across these periods. Keywords were selected as the primary unit of analysis, with a co-occurrence matrix and the simple centrality algorithm for clustering. The evolution map was standardized using the Jaccard index, the minimum number of clustering keywords was set to 5, and the maximum was set to 20.

Results

Publication time

A total of 22,583 articles were obtained from the Web of Science database, and 735 articles proved eligible for this study. The average number of related publications per year before 2017 increased slowly and did not exceed 20 times. After 2016, the annual average number of related publications has increased rapidly, with the number of publications exceeding 100 times in 2020 and 200 times in 2021, respectively (Fig. 1). There was a significant correlation between polynomial curve fits and annual literature growth trend, with a high coefficient of determination ($R^2 = 0.9776$). According to the fitting curve, the number of publications may continue to grow rapidly in the coming years. In the past decade, the total number of citations for 735 AI clinical research has reached a staggering 12,698 times, with an average of more than 17 citations per study. The total number of citations per year exceeded 100 times in 2015, over 1000 times in 2019, and reached 4336 times in 2022. The explosive growth in the number of citations also means that related research has received increasing attention (Fig. 1).

Countries

In total, 53 countries were involved in the publication of AI clinical research, and the United States (243 times) published the most articles, followed by China (233 times), the UK (64 times), South Korea (63 times), and Germany (61 times). There were 22 countries that published more than 5 articles in the network map of the country generated by VOSviewer (Fig. 2). The cooperation between countries is divided into four teams, of which the first team is composed of Germany, Japan, Switzerland, Netherlands, Sweden, Denmark, and Norway; the second team is composed of the United States, China, South Korea, Singapore, and Israel; the third team is composed of Australia and Canada; and the fourth team is composed of the UK, France, Spain, and Italy. The cooperation network map is mainly composed of developed countries, with relatively few developing countries.

Institutions

A total of 1764 institutions participated in the relevant research of AI clinical research. The top five institutions involved in AI clinical research were Harvard University of the United States (31 times), California University of the United States (26 times), Seoul National University of South Korea (24 times), Stanford University of the United States (17 times), and University of Pennsylvania of the United States (16 times); the remaining institutions have

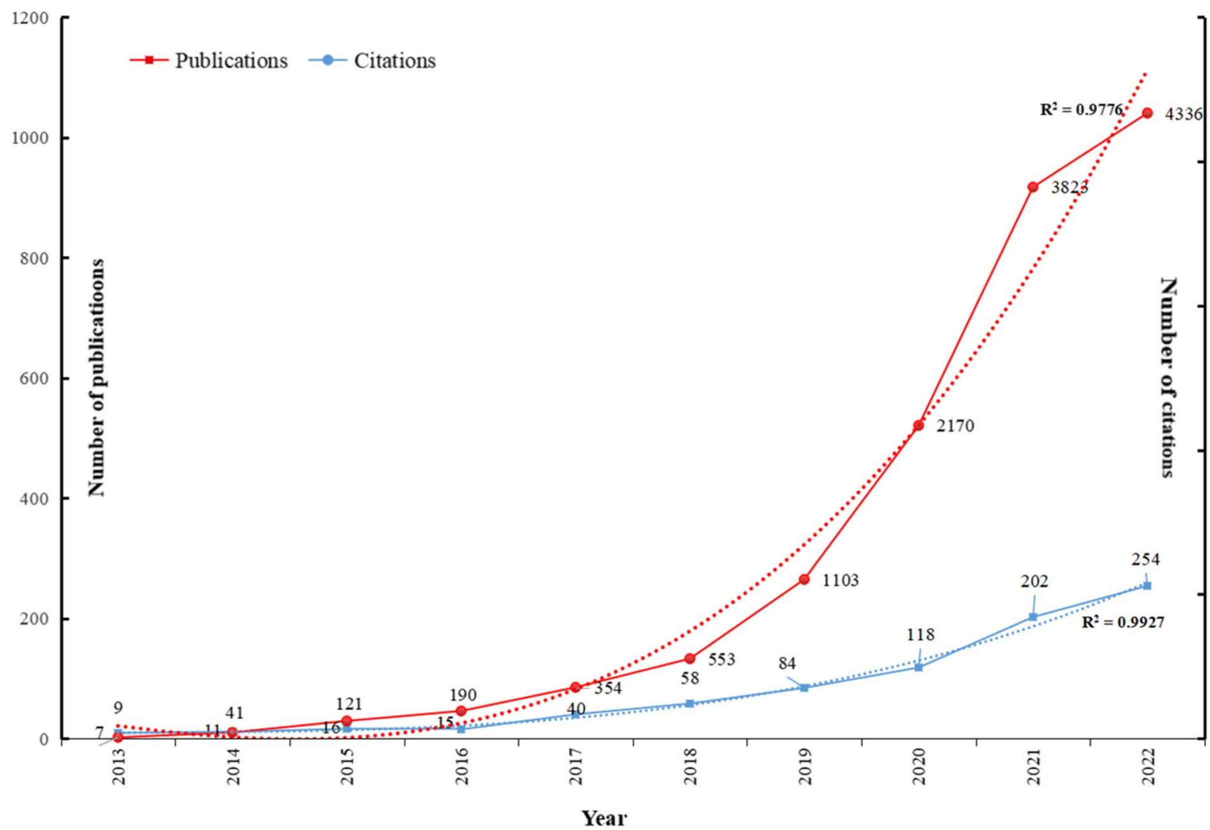

Fig. 1 Number of publications and citations in AI clinical trials

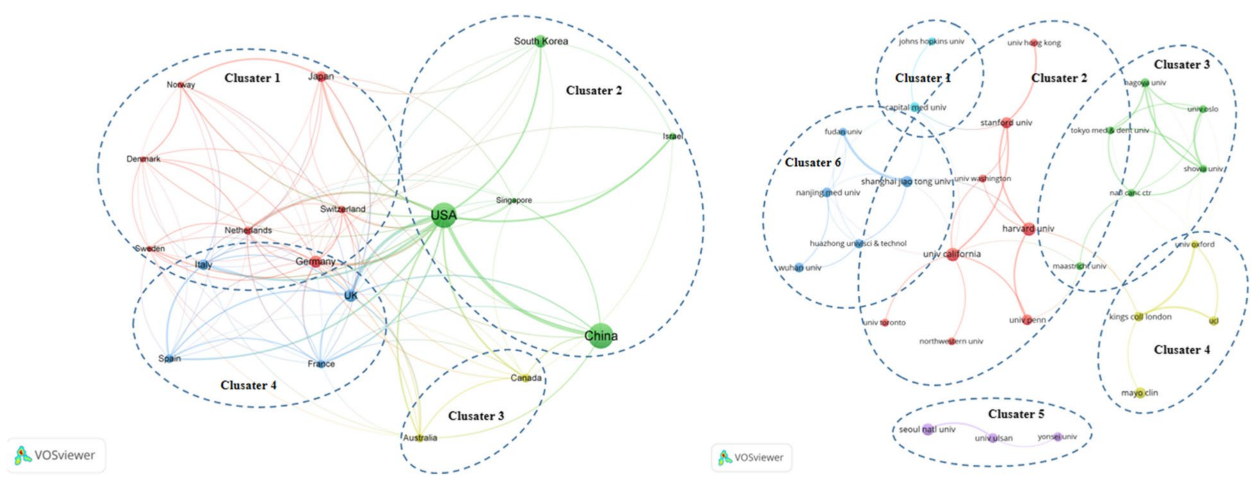

Fig. 2 Network map of cooperation between high-output countries (left) and institutions (right) of AI clinical trials

published less than 15 articles. Based on the analysis of the cooperation of 25 research institutions participating in more than 7 AI clinical research (Fig. 2), these 25 institutions formed 5 clusters, and there were active collaborations among the institutions, especially among the institutions in the same cluster. The Capital Medical University and Johns Hopkins University in Cluster 1 are from China and the United States, respectively. In Cluster 2, except for two universities (the University of Toronto and the University of Hong Kong), the other universities are well-known universities and hospitals from the United States. All universities in Cluster 4, with the exception of the Mayo Clinic, are from the UK. Cluster 3 consists solely of universities from Japan, Cluster 5 comprises only universities from South Korea, and Cluster 6 is made up entirely of universities from China. As can be seen from the institutional cooperation network map, most institutions prefer to cooperate with domestic institutions, and cross-border cooperation needs to be further strengthened.

Authors and co-cited authors

A total of 1706 authors participated in the study of AI clinical trials, and only 6 authors with more than 7 papers, in order: Liu J. ($n=9$ times, Wuhan University), Mori Y. ($n=8$ times, University of Oslo), Mori K ($n=8$ times, Nagoya University), Wu L. ($n=8$ times, Wuhan University), Zhang J. ($n=8$ times, Wuhan University), and Misawa M. ($n=8$ times, Showa University). Based on the analysis of the cooperation of 98 authors who participated in 3 or more articles (Fig. 3), only 40 authors had formed a cooperation network, which indicated that there was a lack of cooperation among other high-yield authors involved in AI clinical research. Although there are few transnational collaborations among high-yielding authors, it is not difficult to see that Sharma P. (University of Kansas) and Mori Yzai (University of Oslo) have played an important role in transnational cooperation (Fig. 3). The collaboration network map of co-cited authors is presented in Fig. 3. Co-cited authors usually indicate that two (or more) authors are cited simultaneously in one or more studies, and collaboration network map of co-cited authors can quickly and intuitively provide information about highly influential research groups and potential collaborators in this field.

Keywords

The main high-frequency keywords include AI (111 times), machine learning (82 times), stroke (47 times), deep learning (46 times), risk (46 times), management (45 times), classification (42 times), rehabilitation (42 times), outcomes (41 times), surgery (37 times), chatbot (36 times), diagnosis (36 times), robotics (36 times), and cancer (35 times). Cluster analysis was carried out on 36 keywords with a frequency greater than 20 and finally clustered into four categories (Fig. 4). The first cluster includes 10 keywords: machine learning, management, chatbot, depression, anxiety, obesity, etc. It mainly focused on evaluating the application effects of AI technologies such as machine learning and chatbots in patient disease management. The second cluster includes nine keywords, such as AI, deep learning, classification, diagnosis, prevention, and disease, which mainly focus on the application effects of AI technology represented by deep learning in disease diagnosis and classification. The third cluster includes nine keywords, such as robotics surgery, system, outcomes, survival, mortality, and RCT. It mainly evaluated the effectiveness and safety of robotic-assisted surgery. The last cluster included eight keywords, such as robotics, rehabilitation, performance, therapy, and reliability, which mainly focused on evaluating the application effect and performance of intelligent robots in patient rehabilitation. In Fig. 5, the keywords with the strongest citation burst were robotics (started in 2014 and ended in 2017), followed by health (started in 2019 and ended in 2020), and randomized control trials (started in 2017 and ended in 2018). The keywords with the strongest burst in recent years were chatbot, gait, mHealth, validation, and intervention.

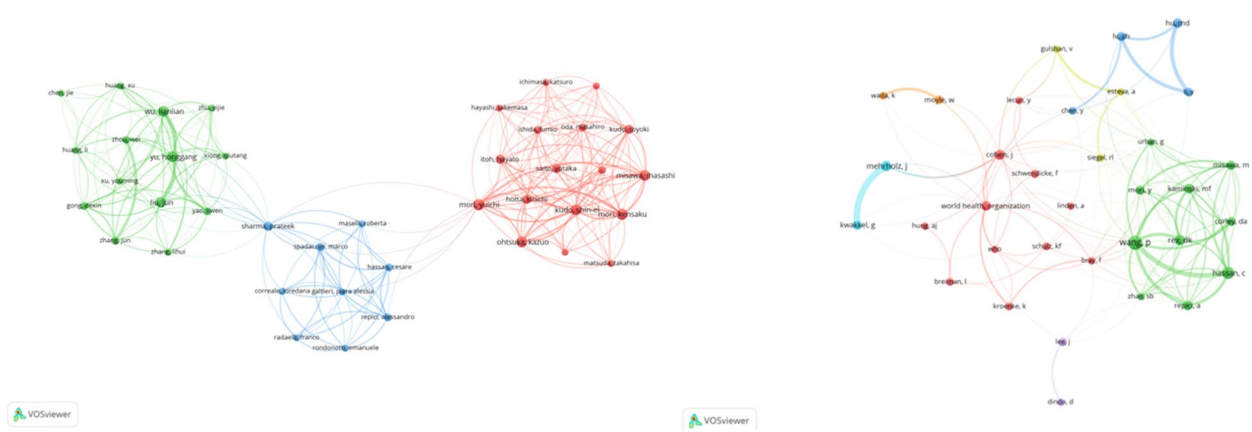

Fig. 3 Network map of cooperation between high-output authors (left) and high co-cited authors (right) of AI clinical trials

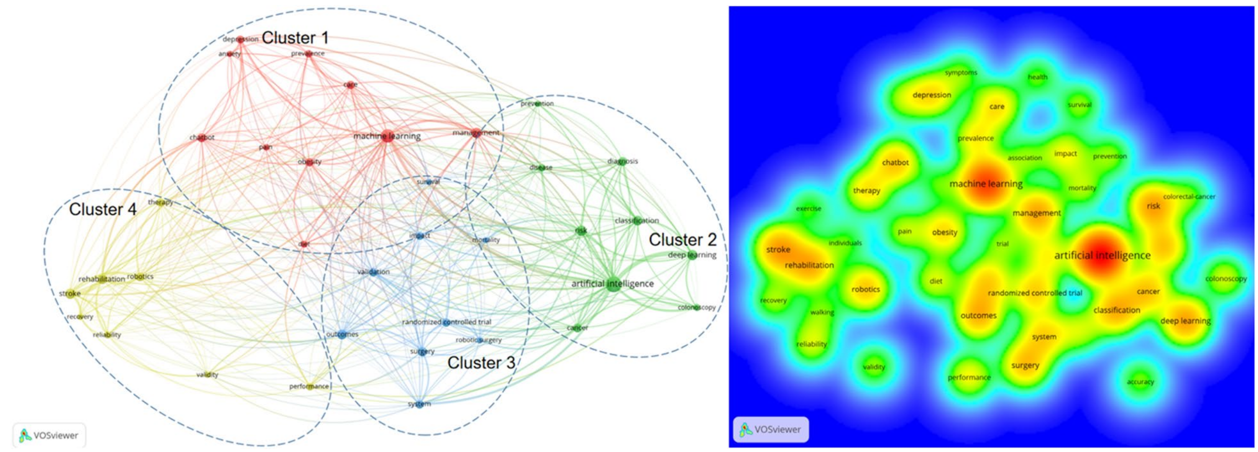

Fig. 4 Network map (left) and density map (right) of keywords in AI clinical trials

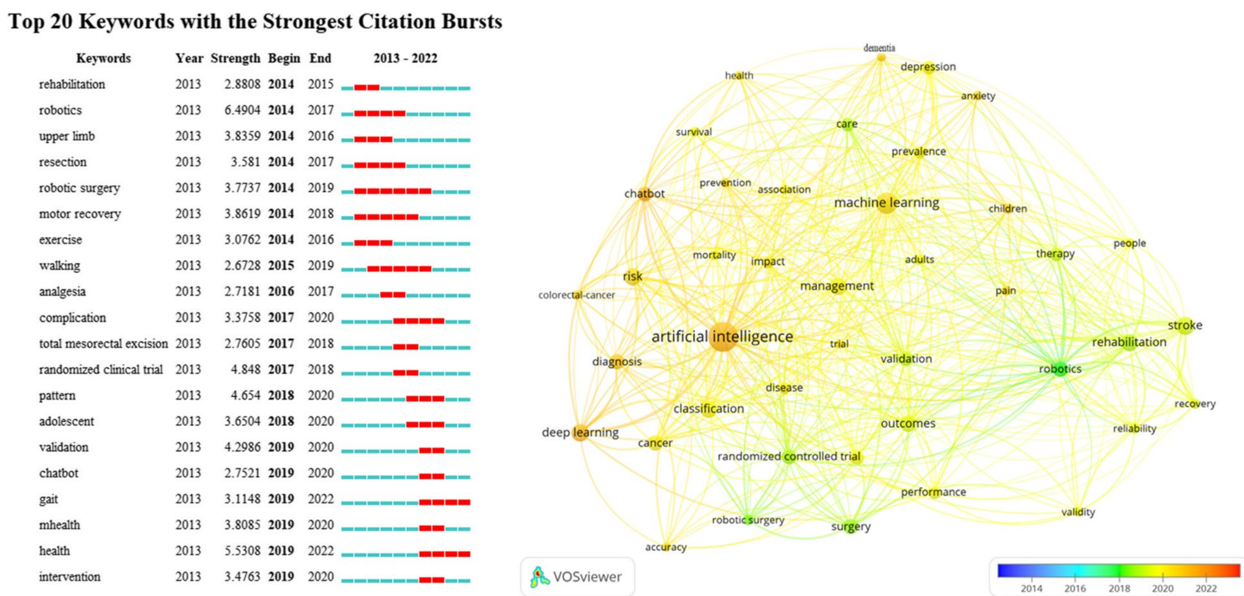

Fig. 5 Keywords burst map (left) and time overlay network map (right) of keywords in AI clinical trials

Keywords burst and theme evolution

The dynamic evolution map can intuitively analyze the evolution process of keywords in the field of AI clinical research in different periods (divided into three periods: 2012–2016, 2016–2019, 2019–2022). Each node represents a cluster of similar keywords, and the size of the node is consistent with the frequency of keyword occurrences; Tthe thickness and color depth of the connecting lines between nodes are proportional to the strength of the correlation between themes. Dark and thick lines indicate that these two adjacent themes have a high degree of similarity and strong evolutionary ability. The solid line represents the mainstream evolution direction of the keyword, while the dotted line represents the evolution direction of the tributary. Isolated points indicate that the topics that appear separately in this research stage are new themes that have not been followed up in the future. Most of the topic words, such as robot-assisted surgery, disease risk prediction, and management, have certain continuity, while keywords such as chatbot and deep learning have been emerging research topics in recent years (Fig. 6).

Journal

Seven-hundred thirty-five AI clinical trials have been published in 342 journals. The top 5 journals with the

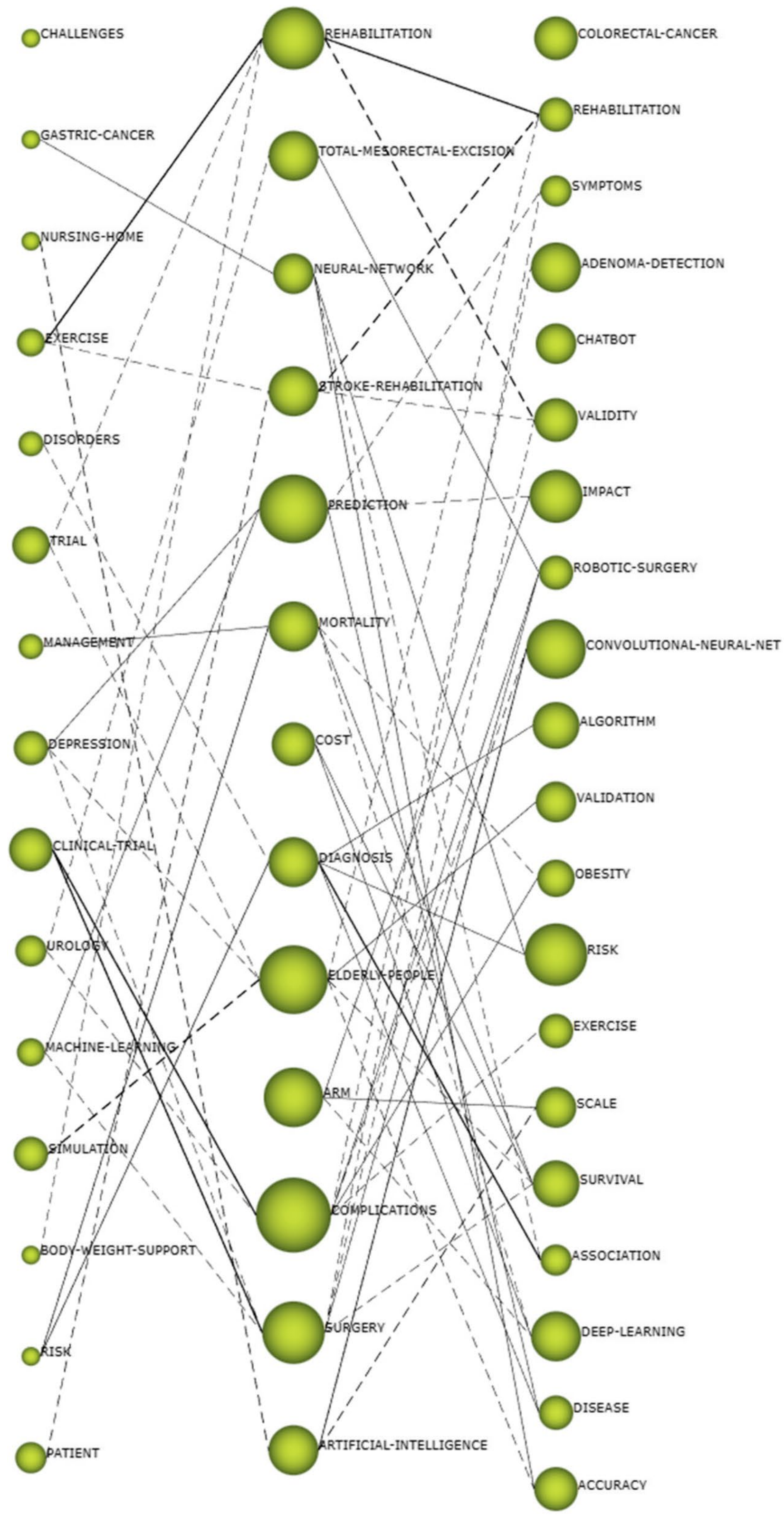

Fig. 6 The thematic evolution map for AI clinical trials — key theme evolution across three periods

largest number of AI clinical trials were *Computational and Mathematical Methods in Medicine* (26 times), *Contrast Media Molecular Imaging* (23 times), *Journal of Healthcare Engineering* (21 times), *Journal of Medical Internet Research* (17 times), and *JAMA Network Open* (13 times). *The New England Journal of Medicine* was the most co-cited journal, with 294 co-citations, followed by *Gastrointestinal Endoscopy* (251 co-citations), *Gastroenterology* (246 co-citations), *JAMA* (221 co-citations), and *Lancet* (219 co-citations). Figure 7 shows the dual-map overlay of journals. The left side represented the map of citing journals, and the right side represented the map of cited journals. The label represented the subject covered by the journal. Curves represent paths of references, originating from the citing map on the left and pointing to the cited map on the right. There were three main citation paths shown on the map. The green paths mean papers published in medicine/medical/clinical mostly cited journals in molecular/biology/genetics, health/nursing/medicine, and psychology/education/social.

References citations

The number of citations for clinical research of AI that we have included has increased rapidly. Among the TOP 10 most cited articles, the most cited research (645 citations), performed by Jayne (2017) and published in JAMA, evaluated the safety and effectiveness of robotic-assisted laparoscopic surgery in patients undergoing rectal cancer resection [29]. The second most-cited research, conducted by Haenssle (2018) and published in the Annals of Oncology, compared deep learning convolutional neural networks' diagnostic performance with a diagnostic team of 58 dermatologists [30]. The third most-cited research, performed by Wang (2019) and published in Gut, compared the difference in detection rates between a deep learning-based automatic polyp detection system and a standard colonoscopy [31]. Four of the remaining high-cited articles assessed the effectiveness of AI in the diagnosis of colorectal cancer [32–35]. The other three high-cited papers evaluated the effect of the application of companion robots on the mental health of the elderly, the safety and effectiveness of robotic-assisted minimally invasive thoracoscopic esophagectomy, and the application effect of the machine learning-derived early warning system for intraoperative hypotension [36–38]. Citation bursts can help researchers quickly identify hot topics and important references in different periods. In Fig. 8, the time period in which a reference was found to have a burst is displayed by a red line, indicating the first year and the last year of the duration of the burst [19]. The strongest burst started in 2019 due to a 2021 paper published in Gut [31].

Discussion

AI technology has flourished, transforming traditional medical practices and shifting from technical development to clinical applications, showing great potential for future healthcare use. Before 2018, the annual publication count was under 50, but in the following years, there was a rapid surge in the number of articles (Fig. 1). It is anticipated that the volume of literature in this field will continue to grow. Clinical trials on AI have primarily been published in recent years, yet the total number of citations has risen sharply, reaching 4336 in 2022. The average citation count per study was 17, reflecting the widespread attention and significant academic impact of the research. Despite challenges such as data privacy, security concerns, low acceptability, and ethical issues, the immense potential and benefits of AI applications in medicine continue to drive researchers' enthusiasm,

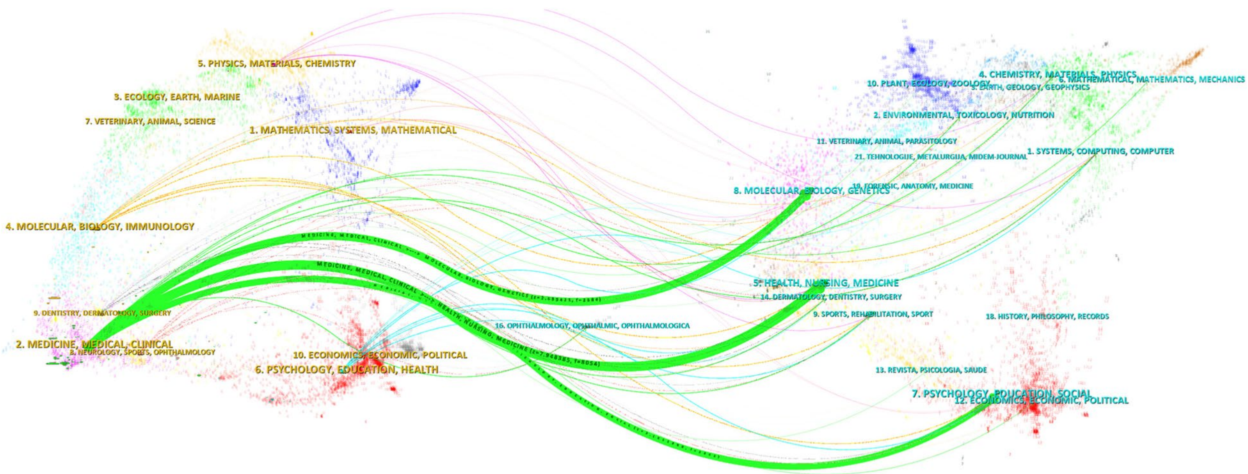

Fig. 7 The dual-map overlay of journals related to AI clinical trials — the citation relationship between journals

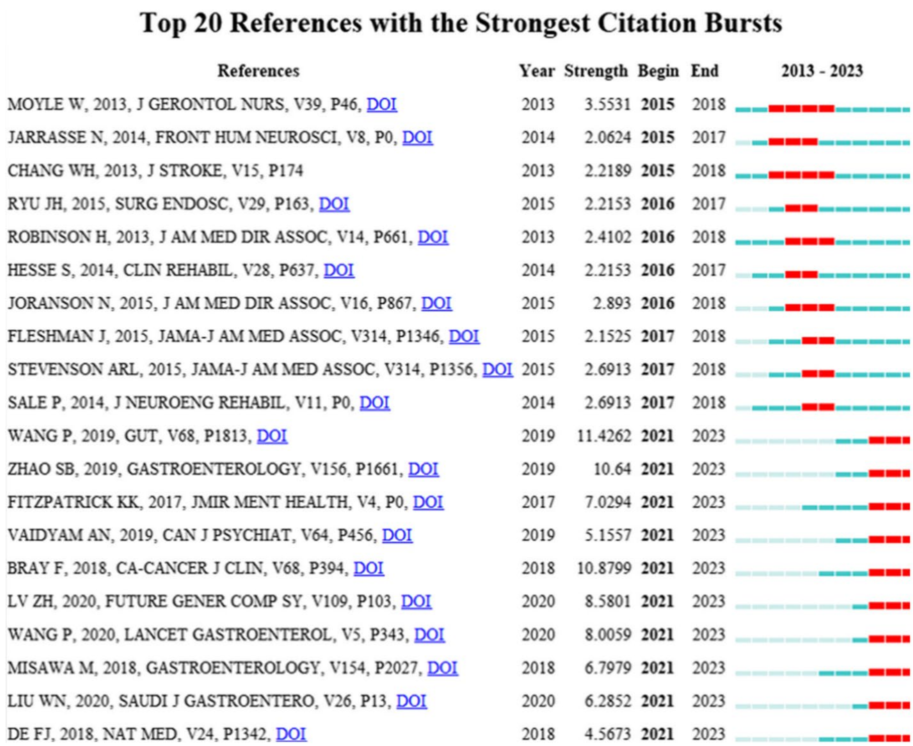

Fig. 8 Citation burst of references in AI clinical trials — references suddenly cited frequently in different periods

with many striving to integrate AI into clinical practice. While previous studies have analyzed AI research across various fields, such as ethics, prostate cancer, COVID-19, and cerebrovascular and heart disease, they often include reviews, commentaries, and technical studies, making it difficult for researchers and clinicians to pinpoint trends in AI clinical trials [27]. This can lead to an exaggerated view of the current state of AI's clinical applications. The number of articles included in this study is significantly lower than in previous bibliometric analyses, primarily due to our strict exclusion of technologies unrelated to AI and nonclinical trial studies.

Among the top 10 countries contributing the most to AI clinical research, 4 are from Europe, 2 from the Americas, and 3 from Asia. Only 23 countries have published more than 5 articles, while 43.4% of countries have published fewer than 6. The United States leads with 33.1% of the total AI clinical research publications, reflecting its dominant position in the field. China stands as the only developing country among the top 10 contributors. The US influence on AI clinical research is significant, likely due to its open policies, strong technological capabilities, and substantial financial support for researchers. International cooperation has formed four main collaborative groups, primarily among developed countries, highlighting the need for stronger partnerships, particularly between developed and developing nations, to prevent further widening the gap in medical science and technology (Fig. 2). The study involved 1706 authors, with only 6 publishing more than 7 papers; Liu J. ($n=9$ times, Wuhan University), Mori Y. ($n=8$ times, University of Oslo), and Mori K. ($n=8$ times, Nagoya University) contributed most articles, and an analysis of collaboration patterns showed limited cooperation among high-yield authors (Fig. 3), with only 40 out of 98 forming networks. Since many AI technologies require vast amounts of data for model training, their clinical applications necessitate extensive collaboration among international, interdisciplinary teams. However, current institutional cooperation remains insufficient. Future international organizations and government bodies should focus on promoting talent exchange and facilitating project collaboration to enhance global efforts in AI healthcare development. Machine learning, deep learning, chatbot, and robotics were the most frequently used AI technology keywords, while stroke and cancer were the most frequently used disease keywords. Other high-frequency keywords used were risk, management, classification, rehabilitation, outcome, surgery, and diagnosis.

Research hotspots are defined as scientific problems or topics extensively discussed in a series of publications within a specific timeframe, with these works demonstrating inherent interconnections [39]. Based on the density map and cluster map of co-occurring keywords

(Figs. 4, 5, and 6), previous AI clinical trials primarily focused on several key hotspots:

(1) Cluster 1 focuses on the application of AI in disease risk prediction and management. For example, Popp et al. compared a machine learning-based personalized diet targeting postprandial glycemic response with standard low-fat diets in adults with abnormal glucose metabolism and obesity, finding no significant differences [40]. Similarly, another study found that a machine learning-based monitoring system improved compliance and cost-effectiveness in obstructive sleep apnea management [41]. Additionally, AI-powered chatbots, utilizing technologies such as deep learning and natural language processing, have emerged as a prominent area of research, with preliminary studies exploring their potential in managing conditions like depression, anxiety, obesity, and pain [42–46].

(2) Cluster 2 focuses on evaluating the application of AI in disease diagnosis and classification. AI algorithms trained for medical image analysis have demonstrated value in diagnosing conditions such as colorectal cancer, pulmonary nodules, chronic atrophic gastritis, and early gastric cancer, as confirmed by clinical trials [33, 47–50]. Additionally, other studies have assessed the accuracy and effectiveness of AI tools in disease diagnosis and therapeutic support, further underscoring their broader potential in this field [51].

(3) Cluster 3 mainly focused on evaluating the application value of robot-assisted surgery. Multiple randomized controlled trial RCTs have also verified the effectiveness and cost-effectiveness of robot-assisted minimally invasive surgery for patients [52, 53]. Although surgical robots have integrated certain AI technologies, they primarily function as facilitators in the surgical field and cannot fully replace surgeons. However, with the rapid advancement of AI, it is anticipated that this technology will fundamentally transform traditional surgical practices.

(4) Cluster 4 focuses on the application of AI technology in rehabilitation. Several studies have explored the effectiveness of integrating AI algorithms, such as machine learning, with other rehabilitation interventions [54].

For instance, Liu et al. demonstrated that machine learning-powered rehabilitation robotic beds significantly improved motor function and lower limb activity in patients with stroke and hemiplegia [55]. Similarly, Bai et al. found that integrating AI-driven limb rehabilitation systems with virtual reality effectively enhanced physical movement ability and activities of daily living in stroke patients [56]. While several studies have explored the use of AI-integrated robots in rehabilitation, most have been limited by small sample sizes, underscoring the need for large-scale randomized controlled trials (RCTs) to further validate their safety and efficacy [57]

In summary, the intervention effects of artificial intelligence in disease risk prediction and management, disease diagnosis and classification, robot-assisted surgery, and rehabilitation training have been preliminarily validated through clinical trials, and these directions may become key areas of clinical application. Moreover, generative AI has also demonstrated significant potential for future applications. The validation of the effectiveness and cost-effectiveness of artificial intelligence in clinical settings through high-quality clinical trials could drive increased investment and support for AI healthcare technologies from governments and healthcare institutions. Therefore, to facilitate the translation of more AI technologies into clinical practice and enhance efficiency, future researchers should broaden the scope of AI clinical trial topics to further explore its potential. Additionally, governments must strengthen regulation to address ethical issues such as data privacy, algorithmic bias, informed consent, and transparency while ensuring fairness and inclusivity to prevent discrimination in patient care.

Limitations

This study has several limitations. First, due to the limitations of the research methodology and objectives, the search was restricted to the Web of Science database, which may introduce bias by excluding relevant studies from other databases or sources. Second, while efforts were made to manually standardize varying expressions and ambiguous definitions of the same concept prior to data analysis, this process likely reduced but did not entirely eliminate errors. Third, the exclusion of conference papers, commentaries, and protocols may have led to the omission of some important publications. Fourth, the determination of whether a technology was AI related relied on the authors' descriptions and expert judgment, which may have introduced subjective bias. Additionally, the quality of the included studies was not assessed, meaning the reliability of some studies cannot be guaranteed. With the development of the SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence) extension and the CONSORT-AI (Consolidated Standards of Reporting Trials-Artificial Intelligence) [58, 59], future research should aim to enhance transparency, quality, and completeness in reporting AI-related studies.

Conclusions

This study offers a comprehensive overview of key articles, journals, leading authors, institutions, and themes within the field. While the number of publications and citations related to AI in clinical trials has grown rapidly, significant opportunities for further advancement remain. Continuous research, comprehensive regulation, and interdisciplinary collaboration are crucial for fully utilizing the potential of AI in advancing healthcare. Close collaborative relationships between institutions and authors are primarily concentrated in developed countries, and international and interdisciplinary cooperation needs to be further strengthened. Additionally, there is a pressing need to cultivate interdisciplinary, composite talents to drive collaboration and innovation across various fields. Previous AI clinical trials have been predominantly focused on the application of AI in surgery, and rehabilitation, disease diagnosis, disease risk prediction, and health management, chatbots, and deep learning are emerging hot research topics in recent years. In the future, more high-quality RCTs are still required to verify the clinical application value of AI technology across different fields.

Abbreviations

AI	Artificial intelligence
RCT	Randomized controlled trial
WOS	Web of Science

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13643-025-02779-2.

Additional file 1. Appendix 1: The detailed search strategy for Web of Science Core Collection database.

Acknowledgements

Thanks to Jihan Shi (School of Computer Science, Central South University), Jiachen Xie (Yale University), Hongshuang Chen (PUMC), and researchers from the Chinese Academy of Medical Sciences and Lanzhou University for their help and guidance.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting, or dissemination plans of this research.

Authors' contributions

Contributors, all authors made a substantial contribution to the research; concept and design, JYS, ZL, and YZ; acquisition, analysis, or interpretation of data, JYS and FYF; drafting of the manuscript, HSC, SJY, XLW, and JYS; and revised manuscript, JYS, SJY, HSC, ZL, CS, and JJX.

Funding

National High Level Hospital Clinical Research Funding (BJ-2024-167; BJ-2024-250).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Nursing, Beijing University of Chinese Medicine, Beijing, China. [2]Nursing Department, Chinese Academy of Medical Sciences and Peking Union Medical Hospital, Beijing 100144, China. [3]Nursing Department, Institute of Geriatric Medicine, National Center of Gerontology, Beijing Hospital, Chinese Academy of Medical Sciences, Beijing, China. [4]School of Nursing, Chinese Academy of Medical Sciences and Peking Union Medical School, Beijing, China. [5]Collaborating Centre of Joanna Briggs Institute, Beijing University of Chinese Medicine, Beijing, China. [6]School of Nursing, Southern Medical University, Guangzhou, China.

Received: 28 September 2024 Accepted: 27 January 2025
Published online: 15 March 2025

References

1. Xiang Y, Zhao L, Liu Z, Wu X, Chen J, Long E, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. Artif Intell Med. 2020Jan;102: 101780.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019Jan;25(1):44–56.
3. Winkler DA. Role of artificial intelligence and machine learning in nanosafety. Small. 2020Sep;16(36): e2001883.
4. Beam AL, Drazen JM, Kohane IS, Leong TY, Manrai AK, Rubin EJ. Artificial intelligence in medicine. N Engl J Med. 2023;388(13):1220–1.
5. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230–43.
6. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC state-of-the-art review. J Am Coll Cardiol. 2021Jan 26;77(3):300–13.
7. Saheb T, Carpenter DO. Mapping research strands of ethics of artificial intelligence in healthcare: a bibliometric and content analysis. Comput Biol Med. 2021;135: 104660.
8. Fosso Wamba S, Queiroz MM. Responsible artificial intelligence as a secret ingredient for digital health: bibliometric analysis, insights, and research directions. Inf Syst Front. 2021May;15:1–16.
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015May 28;521(7553):436–44.
10. Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. J Oral Pathol Med. 2020Oct;49(9):849–56.
11. Hung AJ, Liu Y, Anandkumar A. Deep learning to automate technical skills assessment in robotic surgery. JAMA Surg. 2021Nov 1;156(11):1059–60.
12. Cha JS, Athanasiadis D, Anton NE, Stefanidis D, Yu D. Measurement of nontechnical skills during robotic-assisted surgery using sensor-based communication and proximity metrics. JAMA Netw Open. 2021;4: e2132209.
13. Decuyper M, Maebe J, Van Holen R, Vandenberghe S. Artificial intelligence with deep learning in nuclear medicine and radiology. EJNMMI Phys. 2021;8(1).
14. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res. 2020;22(7): e18228.
15. Plana D, Shung D, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. JAMA Netw Open. 2022Sep 1;5(9): e2233946.
16. Liu N, Shapira P, Yue X. Tracking developments in artificial intelligence research: constructing and applying a new search strategy. Scientometrics. 2021Feb 25;126(4):3153–92.
17. West E, Mutasa S, Zhu Z, Ha R. Global trend in artificial intelligence-based publications in radiology from 2000 to 2018. AJR Am J Roentgenol. 2019Dec;213(6):1204–6.
18. Yeung AWK, Kulnik ST, Parvanov ED, Fassl A, Eibensteiner F, Völkl-Kernstock S, et al. Research on digital technology use in cardiology: bibliometric analysis. J Med Internet Res. 2022May 11;24(5): e36086.

19. Shi J, Gao Y, Ming L, Yang K, Sun Y, Chen J, et al. A bibliometric analysis of global research output on network meta-analysis. BMC Med Inform Decis Mak. 2021;21:144.
20. Cao J, Lim Y, Sengoku S, Guo X, Kodama K. Exploring the shift in international trends in mobile health research from 2000 to 2020: bibliometric analysis. JMIR Mhealth Uhealth. 2021Sep 8;9(9): e31097.
21. Shen Z, Wu H, Chen Z, Hu J, Pan J, Kong J, et al. The global research of artificial intelligence on prostate cancer: a 22-year bibliometric analysis. Front Oncol. 2022Mar;1(12): 843735.
22. Islam MM, Poly TN, Alsinglawi B, Lin LF, Chien SC, Liu JC, et al. Application of artificial intelligence in COVID-19 pandemic: bibliometric analysis. Healthcare (Basel). 2021Apr 9;9(4):441.
23. Tran BX, McIntyre RS, Latkin CA, Phan HT, Vu GT, Nguyen HLT, et al. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int J Environ Res Public Health. 2019;16(12):2150.
24. Tran BX, McIntyre RS, Latkin CA, et al. The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis. Int J Environ Res Public Health. 2019;16(12):2150.
25. Zhang J, Zhang Y, Hu L, Huang X, Liu Y, Li J, et al. Global trends and performances of magnetic resonance imaging studies on acupuncture: a bibliometric analysis. Front Neurosci. 2021Jan;20(14): 620555.
26. Poly TN, Islam MM, Walther BA, Lin MC, Li YCJ, et al. Artificial intelligence in diabetic retinopathy: bibliometric analysis. Comput Methods Programs Biomed. 2023;231: 107358.
27. Shi J, Wei S, Gao Y, Mei F, Tian J, Zhao Y, Li Z, et al. Global output on artificial intelligence in the field of nursing: a bibliometric analysis and science mapping. J Nurs Scholarsh. 2023;55(4):853–63.
28. Zhao W, Yang H, He S, Zhao Q, Wei L. A review of biochar in anaerobic digestion to improve biogas production: performances, mechanisms and economic assessments. Bioresour Technol. 2021;341: 125797.
29. Jayne D, Pigazzi A, Marshall H, Croft J, Corrigan N, Copeland J, et al. Effect of robotic-assisted vs conventional laparoscopic surgery on risk of conversion to open laparotomy among patients undergoing resection for rectal cancer: the ROLARR randomized clinical trial. JAMA. 2017;318(16):1569–80.
30. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018;29(8):1836–42.
31. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomized controlled study. Gut. 2019;68(10):1813–9.
32. van der Sluis PC, van der Horst S, May AM, Schippers C, Brosens LAA, Joore HCA, et al. Robot-assisted minimally invasive thoracolaparoscopic esophagectomy versus open transthoracic esophagectomy for resectable esophageal cancer: a randomized controlled trial. Ann Surg. 2019;269(4):621–30.
33. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med. 2018;169(6):357–66.
34. Repici A, Badalamenti M, Maselli R, Correale L, Radaelli F, Rondonotti E, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology. 2020;159(2):512–20.e7.
35. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomized study. Lancet Gastroenterol Hepatol. 2020;5(4):343–51.
36. Gong D, Wu L, Zhang J, Mu G, Shen L, Liu J, et al. Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomized controlled study. Lancet Gastroenterol Hepatol. 2020;5(4):352–61.
37. Wijnberge M, Geerts BF, Hol L, Lemmers N, Mulder MP, Berge P, et al. Effect of a machine learning-derived early warning system for intraoperative hypotension vs standard care on depth and duration of intraoperative hypotension during elective noncardiac surgery: the HYPE randomized clinical trial. JAMA. 2020;323(11):1052–60.
38. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc. 2013;14(9):661–7.
39. Wu L, Zhang J, Zhou W, An P, Shen L, Liu J, et al. Randomized controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gut. 2019;68(12):2161–9.
40. Popp CJ, Hu L, Kharmats AY, Curran M, Berube L, Wang C, et al. Effect of a personalized diet to reduce postprandial glycemic response vs a low-fat diet on weight loss in adults with abnormal glucose metabolism and obesity: a randomized clinical trial. JAMA Netw Open. 2022;5(9): e2233760.
41. Turino C, Benítez ID, Rafael-Palou X, Mayoral A, Lopera A, Pascual L, et al. Management and treatment of patients with obstructive sleep apnea using an intelligent monitoring system based on machine learning aiming to improve continuous positive airway pressure treatment compliance: a randomized controlled trial. J Med Internet Res. 2021;23(10): e24072.
42. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: a randomized controlled trial. JMIR Ment Health. 2018;5(4): e64.
43. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth. 2020;8(4): e15806.
44. Liu H, Peng H, Song X, Xu C, Zhang M. Using AI chatbots to provide self-help depression interventions for university students: a randomized trial of effectiveness. Internet Interv. 2022;27: 100495.
45. Fitzsimmons-Craft EE, Chan WW, Smith AC, et al. Effectiveness of a chatbot for eating disorders prevention: a randomized clinical trial. Int J Eat Disord. 2022;55(3):343–53.
46. To QG, Green C, Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. JMIR Mhealth Uhealth. 2021;9(11): e28577.
47. Wu L, He X, Liu M, Xie H, An P, Zhang J, et al. Evaluation of the effects of an artificial intelligence system on endoscopy quality and preliminary testing of its performance in detecting early gastric cancer: a randomized controlled trial. Endoscopy. 2021;53(12):1199–207.
48. Wu L, Shang R, Sharma P, Zhou W, Liu J, Yao L, et al. Effect of a deep learning-based system on the miss rate of gastric neoplasms during upper gastrointestinal endoscopy: a single-centre, tandem, randomized controlled trial. Lancet Gastroenterol Hepatol. 2021;6(9):700–8.
49. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. Lancet Oncol. 2019;20(12):1645–54.
50. Nakashima H, Kitazawa N, Fukuyama C, Kawachi H, Kawahira H, Momma K, et al. Clinical evaluation of computer-aided colorectal neoplasia detection using a novel endoscopic artificial intelligence: a single-center randomized controlled trial. Digestion. 2023;1–9.
51. Yao X, McCoy RG, Friedman PA, Shah ND, Barry BA, Behnken EM, et al. ECG AI-guided screening for low ejection fraction (EAGLE): rationale and design of a pragmatic cluster randomized trial. Am Heart J. 2020;219:31–6.
52. Chandra A, Snider JT, Wu Y, Jena A, Goldman DP. Robot-assisted surgery for kidney cancer increased access to a procedure that can reduce mortality and renal failure. Health Aff (Millwood). 2015;34(2):220–8.
53. Zhang X, Wang B. Robot-assisted surgery for renal cell carcinoma with caval thrombosis. Eur Urol Focus. 2018;4(5):639–40.
54. Chae SH, Kim Y, Lee KS, Park HS. Development and clinical evaluation of a web-based upper limb home rehabilitation system using a smartwatch and machine learning model for chronic stroke survivors: prospective comparative study. JMIR Mhealth Uhealth. 2020;8(7): e17216.
55. Liu G, Cai H, Leelayuwat N. Intervention effect of rehabilitation robotic bed under machine learning combined with intensive motor training on stroke patients with hemiplegia. Front Neurorobot. 2022;16: 865403.
56. Bai Y, Liu F, Zhang H. Artificial intelligence limb rehabilitation system on account of virtual reality technology on long-term health management of stroke patients in the context of the Internet. Comput Math Methods Med. 2022;2022:2688003.
57. Papadopoulos C, Castro N, Nigath A, Davidson R, Faulkes N, Menicatti R, et al. The CARESSES randomized controlled trial: exploring the

health-related impact of culturally competent artificial intelligence embedded into socially assistive robots and tested in older adult care homes. Int J Soc Robot. 2022;14(1):245–56.
58. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26:1351–63.
59. Liu X, Faes L, Calvert MJ, Denniston AK. Extension of the CONSORT and SPIRIT statements for reporting of trials involving artificial intelligence: the CONSORT-AI and SPIRIT-AI guidelines. Lancet. 2019;394:1225.

Publisher's Note