REVIEW ARTICLE

Emerging Trends of the Health Sciences and Machine Learning: A Bibliometric and Visualization Analysis Using CiteSpace II

Yin Lijia[1], Teh Sin Yin[1], Noorfatimah Yahaya[2], Ng Wei Chien[1]

[1] School of Management, Universiti Sains Malaysia,11800 Minden, Penang, Malaysia.
[2] Advanced Medical & Dental Institute, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia.

ABSTRACT

The health sciences field includes several disciplines such as medicine, biology, and psychology. Machine Learning (ML) is essential for the swift analysis of large datasets, facilitating the personalization of medical therapies according to each patient's own clinical characteristics. This project aims to perform a thorough analysis of emerging trends in the application of ML within health sciences and to propose potential future directions. Bibliometric analysis and CiteSpace visualization tools were utilized to analyze 159 peer-reviewed publications published in the Web of Science core collection from 2014 to September 19, 2024. This study examined publication trends, keyword co-occurrence networks, and the progression of research themes. This research clarified notable application trends and primary research focal points in the domain of ML as it pertains to health sciences. The application of visualization techniques provides new insights into the trends of ML applications in health sciences.

Malaysian Journal of Medicine and Health Sciences (2025) 21(SUPP10):161-164. doi:10.47836/mjmhs.21.s10.31

Keywords: Bibliometric, CiteSpace, Human Health, Health Sciences, Machine Learning

Corresponding Author:
Teh Sin Yin, PhD
Email: tehsyin@usm.my
Tel :+604-6532894

INTRODUCTION

Health sciences promote the overall health of individuals and communities, prevent disease, enhance quality of life, and provide a solid scientific foundation for policy development and implementation in the field of health care (1). Machine Learning (ML), as an important branch of artificial intelligence (AI), is continuously advancing and developing. The utilization of machine learning applications in the healthcare domain enables medical professionals to acquire advanced diagnostic and treatment capabilities (2). Current medical service systems have experienced a fundamental shift through the combination of healthcare fields with AI technology. The healthcare sector follows proven disease prevention approaches and policy construction methods (3) until ML changes diagnosis capabilities through precise prediction and person-oriented care approaches. Advanced analytical methods support complex systems and healthcare concepts because they evaluate data networks with numerous interlinking parameters (4).

Literature Review and Research Gaps

ML has shown great promise in healthcare, improving clinical decision support, patient management, and equitable transplantation systems (Pappada et al., 5). Kumar et al. (6) and Arina et al. (7) highlights AI implementation in pharmacology, focusing on drug development, disease diagnosis, personalized treatment, patient risk prediction, surgical algorithm validation, and bias prevention. However, there are several challenges such as inadequate ability of models to generalize across diverse datasets and differences in healthcare data across various racial and geographical groups. In addition, there is a pressing need for adaptation studies to optimize ML implementations in resources-constrained settings. Moreover, the scarcity of data on rare diseases necessitates few-shot learning advancements. While the lack of clear ethical guidelines hinders secure medical data exchange. This study addresses these gaps by examining critical limitations in health science ML research and proposing future directions for improving model robustness, data inclusivity, and ethical frameworks.

Research Objectives and Significance

This study utilizes Citespace II to display a table of journal co-citation, a co-occurrence of keywords, and a temporal distribution map of articles published in Web of Science (WOS) about the application of ML in health sciences.

RQ1: Journal co-citation is employed to illustrate publishing patterns in ML applied to health sciences and to reveal connections among various articles.

RQ2: Keyword co-occurrence is employed to examine and elucidate themes and interconnections in scientific literature pertaining to the application of ML in health sciences.

RQ3: The keyword timeline graph aims to enhance researchers' analyzing of the knowledge structure of ML in health science research and to detect emerging trends that are pushing the boundaries of the field.

ML is transforming health science research through enhancing evaluation, personalized treatment, and predictive methodologies. ML improves disease screening, optimizes resource planning, and supports ethical healthcare frameworks through technological governance. This study highlights ML's impact and identified key challenges to guide future advancements.

Bibliometric Method

Researchers employ bibliometrics to convert document data into knowledge through the application of statistical tools that assess the dissemination of scientific literature across various themes in both spatial and temporal contexts (13). Science mapping methodology visualizes knowledge structures and research evolution, and developing trend patterns, through spatial and relational representations of academic institutions.

The study employs CiteSpace II as the visualization and analysis instrument, that academics frequently utilize for assessing knowledge graphs across several scientific disciplines (14).

Data Sources Search Method

This is the standard for performing research and implementing scientific mapping. The WOS database was selected for scientific mapping purposes. WOS includes distinguished research with selectivity, balance, and comprehensiveness, adhering to stringent requirements for indexing and preserving accurate citation metrics and bibliographic data (15). The phrase refers to the period from 2014 to September 19, 2024. The following keywords are utilized to assess and examine the paper. The use of "ML" and "health sciences" as keywords aim to explore the applications of ML in the field of health sciences, thus ensuring the research topic and scope are distinctly defined. Third, studies published in languages other than English were excluded. Fourth, CiteSpace II offers a deduplication tool to ensure that the dataset is free from redundant data that could affect the analysis outcomes (16). This investigation found no duplicate or extraneous data after data pre-processing phase of CiteSpace II. The comprehensive data resource search protocol is depicted in Figure1.

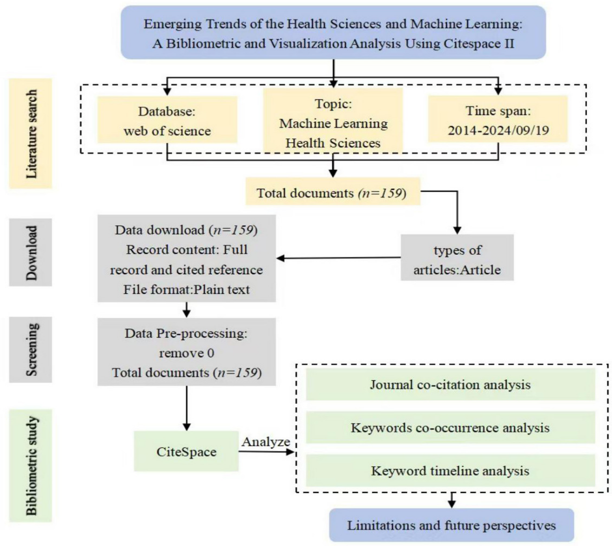

Figure 1: Flowchart of the processes of classifying, screening, and including relevant studies

RQ1: Journal Co-citation Analysis and the Intellectual Structure

Journal co-citation analysis, various fields of health science research including ML were found, hence providing researchers with more routes for exploration. Table 1 presents the framework of ML within health science research knowledge. Despite Journal of the American Medical Association (JAMA) substantial citation count, the peripheral status in the journal co-citation network is shown by a centrality of 0.01 and a total of 28 citations. The New England Journal of Medicine and British Medical Journal (BMJ) Quality & Safety both own 21 citations, with centrality values of 0.03 and 0.02, respectively. Table 1 indicates that the ten leading papers employing ML applications in health sciences are categorized under the Q1 tier of the Science Citation Index.

RQ2: Keywords Co-occurrence Analysis

Researchers commonly employ keyword co-occurrence

Table I: Top 10 journals by total citation

Top 10 by citation	Total Citation	Centrality	Quartile in category
Jama-Journal of the American Medical Association	28	0.01	Q1
New England Journal of Medicine	21	0.03	Q1
BMJ Quality & Safety	21	0.02	Q1
BMJ-British Medical Journal	19	0.03	Q1
Academic Medicine	18	0.12	Q1
Lancet	17	0.02	Q1
Plos One	16	0.06	Q1
Health Affairs	14	0.02	Q1
Annals of Internal Medicine	13	0.16	Q1
Journal of General Internal Medicine	12	0.06	Q1

Figure 2: Keyword co-occurrence network diagram

analysis to investigate the knowledge structure within a discipline, as shown in Figure 2. The ten most common keywords in the fields are quality improvement, care, outcome, impact, prevention, patient safety, mortality, quality, surgery, and health care. “Quality improvement” and “outcome” are intricately linked and signify that the domain prioritizes and emphasizes the effects of enhancing medical quality (18). The repetition of “care” 14 times and “patient safety” 5 times signifies that the assurance of patient safety is a significant research priority in the medical domain. The co-occurrence analysis of these keywords clearly indicates the study focus of ML in the medical domain.

RQ3: Keyword Timeline Analysis

CiteSpaceII software may incorporate temporal elements into the network to generate timeline maps, facilitating the analysis of keyword changes and trends over time (19). Figure 3 illustrates the history of keyword co-occurrence, revealing distinct clusters of keywords. The health cluster exhibited the longest co-occurrence duration, encompassing “communities of practice”, “risk”, and “adjustment”. Secondly, there exists a cluster pertaining to “chronic kidney”, “on-call supervision”, and “health administrative data”. The remaining keyword clusters are “postgraduate education”, “faculty development”, “urinary catheter”, “heart health”, “quality improvement”, “early integration”, and “infection control”.

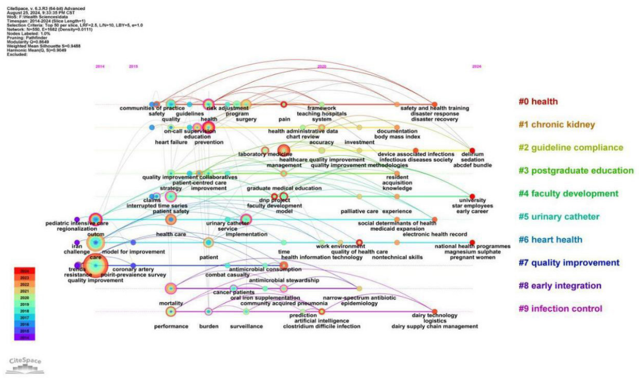

Figure 3: Timeline of keyword co-occurrence

CONCLUSION

The researcher conducts a comprehensive visual analysis of ML literature in health sciences papers from 2014 to September 19, 2024, utilizing the WOS database. The study examines the knowledge repository comprising journal co-citation data, keyword co-occurrence information, and historical keyword details. From 2014 to 2016, the study domain concentrated significantly on “patient safety” and “quality improvement”. Research on “artificial” and “health information technology” issues has gained prominence throughout the period from 2021 to 2024 and indicates significant improvements in this domain.

Limitations and Future Perspectives

The exclusive use of WOS database could have affected the quantity and scope of available data information. Research following this effort should analyze data sourced from Scopus and China National Knowledge Infrastructure (CNKI). The development of ML models for health sciences requires algorithms that can outperform classic methods. Appropriate analysis methods such as logistic regression or support vector machines should be selected and developed to match every requirement (20).

ACKNOWLEDGEMENT

This work was supported by the Universiti Sains Malaysia, APEX ERA Grant (Grant Number: R502-KR-ARU004-0000000093-K134).

REFERENCES

1. Van Druten VP, Bartels EA, Van de Mheen D, De Vries E, Kerckhoffs APM, Nahar-van Venrooij LMW. Concepts of health in different contexts: a scoping review. BMC Health Serv Res 2022;22(1):389. doi:10.1186/s12913-022-07787-3.
2. Mateen BA, Liley J, Denniston AK, Holmes CC, Vollmer SJ. Improving the quality of machine learning in health applications and clinical research. Nat Mach Intell 2020;2(10):554-6.

doi:10.1038/s42256-020-00267-8.
3. World Health Organization. Global strategy on digital health 2020-2025. Geneva: WHO Press; 2021.
4. Leydesdorff L, Wagner CS, Bornmann L. Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. J Informetr. 2020;14(3):101009. doi:10.1016/j.joi.2020.101009.
5. Scott M Pappada. Machine Learning in Medicine: It Has Arrived, Let's Embrace It. J Card Surg. 2021 Nov;36(11):4121-4124. doi:10.1111/jocs.15918.
6. Mandeep Kumar, T P Nhung Nguyen, Jasleen Kaur, Thakur Gurjeet Singh, Divya Soni, Randhir Singh, Puneet Kumar. Opportunities and Challenges in Application of AI in Pharmacology. Pharmacol Rep. 2023 Jan;75(1):163-178. doi:10.1007/s43440-022-00445-1.
7. Arina P, Kaczorek MR, Hofmaenner DA, Pisciotta W, Refinetti P, Singer M, Mazomenos EB, Whittle J. Prediction of Complications and Prognostication in Perioperative Medicine: A Systematic Review and PROBAST Assessment of Machine Learning Tools. Anesthesiology. 2024 Jan 1;140(1):85-101. doi:10.1097/ALN.0000000000004764.
8. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.
9. Maier A, et al. Few-shot learning in medical imaging. IEEE Trans Med Imaging 2020;39(4):1152-62. doi:10.1109.
10. Price WN, Cohen IG. Privacy in the age of medical big data. Nat Med 2023;29(1):22-30. doi:10.1038/s41591-018-0272-7.
11. Rajpurkar, P., et al. AI in global health: Opportunities and challenges. Science 2022;375(6582):eabn2254. doi:10.1126/science.abn2254.
12. Schulam, P., & Saria, S. Reliable decision support using counterfactual models. Journal of Machine Learning Research 2019;20(1):1-39. doi:10.5555/3322706.3322712.
13. Luo X, Wu Y, Niu L, Huang L. Bibliometric analysis of health technology research: 1990~2020. Int J Environ Res Public Health 2022;19(15):9044. doi:10.3390/ijerph19159044.
14. Dong H, Wu Q, Pang Y, Wu B. A comparative analysis on risk communication between international and Chinese literature from the perspective of knowledge domain visualization. Environ Health Prev Med 2021;26(1):60. doi:10.1186/s12199-021-01115-6.
15. Zhu J, Liu W. A tale of two databases: the use of Web of Science and Scopus in academic papers. Scientometrics 2020;123(1):321-35. doi:10.1007/s11192-020-03512-7.
16. Sun W, Wu W, Dong X, Yu G. Frontier and hot topics in the application of hydrogel in the biomedical field: a bibliometric analysis based on CiteSpace. J Biol Eng 2024;18(1):40. doi:10.1186/s13036-024-00325-5.
17. Kolasa K, Admassu B, Hołownia-Voloskova M, Kędzior KJ, Poirrier JE, Perni S. Systematic reviews of machine learning in healthcare: a literature review. Expert Rev Pharmacoecon Outcomes Res 2024;24(1):63-115. doi:10.1080/14737167.2024.2167397.
18. Hill JE, Stephani AM, Sapple P, Clegg AJ. The effectiveness of continuous quality improvement for developing professional practice and improving health care outcomes: a systematic review. Implement Sci 2020;15:1-14. doi:10.1186/s13012-020-01123-7.
19. Wang W, Lu C. Visualization analysis of big data research based on Citespace. Soft Comput 2020;24(11):8173-86. doi:10.1007/s00500-020-05109-9.
20. Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. Nat Biomed Eng 2022;6(12):1330-45. doi:10.1038/s41551-022-00906-3.