Int J Med Rev 2026 March;13(1): 1088-1095

10.30491/ijmr.2025.512198.1321

INTERNATIONAL JOURNAL OF MEDICAL REVIEWS

Original Article

A Text Mining Analysis of Iranian Researchers' Scientific Publications in Artificial Intelligence for Medical Sciences

Mostafa Kashani [1], Meisam Dastani [2*]

[1] Sirjan University of medical sciences, Sirjan, Iran
[2] Social Determinants of Health Research Center, Gonabad University of Medical Sciences, Gonabad, Iran

*** Corresponding Author:** Meisam Dastani, Social Determinants of Health Research Center, Gonabad University of Medical Sciences, Gonabad, Iran. E-mail: meisam.dastani@gmail.com

Received March 13, 2025; **Accepted** September 28, 2025; **Online Published** March 30, 2026

Abstract

Introduction: In recent years, the field of artificial intelligence (AI) and medical sciences has faced an increase in scientific production, which indicates the development of this research field. Therefore, this study was conducted with the aim of analyzing the process of Iranian researchers' scientific publications in AI for medical sciences by employing text mining techniques.

Methods: The present work is an applied research and was conducted using text mining techniques, specifically the topic modeling algorithm. Data was extracted until August 24, 2024, using an appropriate search strategy in the WOS database. In addition, the Python programming language was used to analyze the textual data.

Results: According to the outcome of topic modeling on 572 data related to the Iranian researchers' scientific publications in AI for medical sciences, five main topic clusters were obtained, which respectively include "modeling and prediction in medical systems," "AI," "learning, and training," "diagnosis and modeling in digital medicine," "cancer diagnosis using deep learning,"and "modeling and classification of medical images."

Conclusion: The focus of Iranian research in the field of AI in medical sciences is more towards applications that are directly related to the improvement of prediction and modeling of diseases and medical systems. In addition, the trend towards emerging issues, such as medical education and the use of AI for the development of educational skills, indicates attention to the wider applications of this technology in improving the quality of education and learning in medical sciences.

Keywords: Artificial Intelligence, AI, Medical Sciences, Text Mining, Topic Modeling, Iran

Introduction

Artificial intelligence (AI) is one of the leading new technologies in both scientific and industrial fields. It has attracted the attention of many researchers, particularly in medical sciences. AI is now considered a transformative technology in healthcare, with strong potential to change medical practice, research, and management.[1]

The use of AI in medical sciences is rapidly growing. Its unique capabilities include processing big data, analyzing patterns, and predicting clinical outcomes. This innovative technology plays an important role in diagnosing diseases, predicting effective treatments, and improving healthcare processes.[2] Since AI includes machine learning, natural language processing, and intelligent robots, artificial intelligence-based technologies provide numerous opportunities for innovation in the knowledge-based healthcare industry.[3] For instance, AI algorithms have shown exceptional capabilities in analyzing medical images, predicting disease progression, and supporting clinical decision-making.[4]

In recent years, significant advances in computing and data-driven technologies have caused fundamental changes in research approaches in the field of medical sciences. The growth in computing power, together with advances in cloud technologies and the availability of large biomedical datasets, has created new opportunities for research. In this environment, scientists can analyze and interpret medical data using advanced machine learning algorithms.[5,6] These developments have caused AI to become one of the main pillars of research and development in medical sciences so that researchers employ these technologies to improve diagnosis accuracy, predict treatment outcomes, and identify latent patterns in clinical data.[7]

In addition, the increasing development of machine learning and deep learning algorithms, which are capable of processing and analyzing complex data with high accuracy, has led medical researchers to consider using these technologies.[8-10]

These technologies allow faster and more complex analyses. They also serve as powerful tools for discovering new knowledge and improving medical and healthcare processes.[11] The result of this new approach was a significant increase in scientific production in the field of AI and medical sciences, which indicates the growth and expansion of this field in recent years. This growing trend indicates the significance and special place that AI will play in the future shape of medical sciences.[12] Due to the importance and influence of AI in medical sciences, Iranian researchers have also become interested in this field and have made significant efforts to produce science and present scientific achievements in recent years. Tran et al. (2019) demonstrated that Iran is in 12th placeregarding the number of scientific publications in this field.[13] In another investigation, Guo et al. (2020) indicated that Iran ranks 10th in terms of scientific production in the field of AI in healthcare.[14] Although there has been an increase in the scientific publications of Iranian researchers in this scientific field, text mining analysis and topic modeling of Iranian researchers' scientific publications in AI for medical sciences have not been performedso far. Therefore, the scientific evaluation and analysis of the publications of Iranian researchers in this field are considerably important in identifying the trend of the published topics. This analysis can identify the strengths and shortcomings of the conducted research and provide new research fields for future researchers. Text mining is among the methods used to analyze publications and scientific productions.[15] Text mining techniques have become increasingly important for the analysis of scientific publications in medical sciences.[16] As one of the frequently used methods in textual data analysis, text mining is considered an appropriate tool for extracting data and latent patterns in scientific texts. By employing this method, it is possible to perform a deep analysis of the scientific content and identify the main research topics and trends in this field.[17,18] Accordingly, the present research aimed to analyze the Iranian researchers' scientific publications in AI for medical sciences indexed in the WOS database using text mining and topic modeling techniques.

Materials and Methods

The present study is an applied research, which was conducted using text mining techniques. Data was extracted using an appropriate search strategy (Table 1) and advanced search in the WOSCC database. The searches were conducted without time limit and continued until August 24, 2024.

Table 1. Search Strategy in the WOSCC Database

No.	Search Strategy	Number of Retrieved Data
#1	AI (Topic) or Artificial Intelligence (Topic) or Computational Intelligence (Topic) or Machine Intelligence (Topic) or Computer Reasoning (Topic) or Computer Vision System* (Topic) or Knowledge Acquisition (Topic) or Knowledge Representation* (Topic)	493,755
#2	medic* (Topic)	3,111,029
#3	#1 AND #2	41,352
#4	Results were limited to scientific productions in Iran	572

The text mining methods and Latent Dirichlet Allocation (LDA) topic modeling algorithm were applied using Python programming language to analyze textual data. The steps of text mining were as follows:

Data Gathering

The data employed in this study was imported from a CSV file. The Title, Abstract, and Keywords columns were combined for each article to provide a text data set for analysis.

Text Pre-Processing

The pre-processing of the texts included various steps, such as removing punctuation and numbers, deleting special and stop words, and lemmatizing the words. Therefore, noise removal and data preparation before running the algorithms have a direct impact on the quality of topic clustering. They also help reveal the hidden information in scientific texts with greater

clarity.[19] Accordingly, these steps were performed to improve data quality and reduce textual noise.

Development of a Document-Term Matrix

After preprocessing, a document-term matrix was created. This matrix displayed the number of repetitions of each word in each document. Then, the co-word network was visualized before the implementation of the LDA algorithm to examine the relationships between words.

Topic Modeling Using LDA

GridSearchCV was used to find the best LDA model and the optimal number of topic clusters. After the selection of the desired number of topic clusters, the authors prepared the LDA modeling algorithm on the extracted scientific production data, including title, abstract, and keywords. Afterwards, important words and the documents related to each cluster were extracted, and the clusters were labeled based on these words.

LDA is one of the most widely used algorithms in topic modeling, designed to uncover hidden patterns in texts. LDA is based on the assumption that each document is a mixture of several latent topics, and each topic is represented as a set of words with a specific probability distribution. In other words, documents are composed of topics with varying weights, and topics are defined by the likelihood of word occurrences. Accordingly, a document is assigned to the topic for which it has the highest weight.[20,21]

Visualization of the Outcomes

The results of the LDA model were graphically illustrated to facilitate the presentation of each cluster's topics.

Results

According to the search conducted in the WOSCC database, 572 items were extracted. Figure 1 indicates the word cloud of Iranian researchers' scientific publications in AI for medical sciences.

Figure 1. Word cloud of Iranian Researchers' Scientific Publications in AI for Medical Sciences.

Table 1. Results Obtained from Topic Modeling of Iranian Researchers' Scientific Publications in AI for Medical Sciences

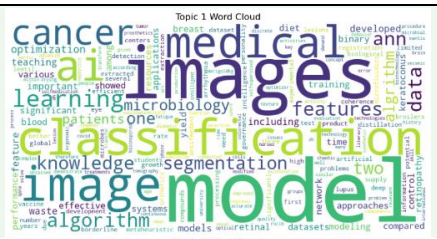
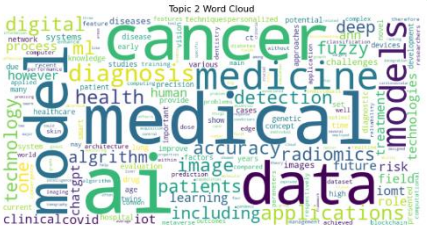
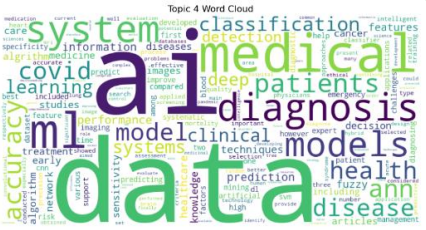
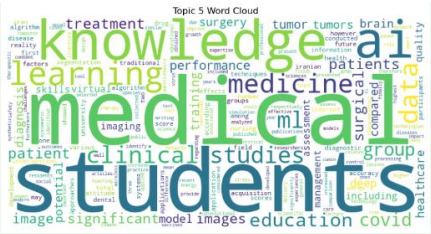
Topic No.	Important Words	WordCloud
0	model, images, classification, image, medical, ai, cancer, learning, algorithm, features	
1	ai, medical, cancer, data, model, medicine, models, diagnosis, applications, digital	
2	ai, cancer, medical, learning, breast, deep, models, imaging, studies, images	
3	data, ai, medical, ml, diagnosis, system, models, patients, accuracy, ann	
4	medical, students, knowledge, ai, learning, medicine, data, clinical, ml, studies	

Figure 1 indicates that the words "ai, machine learning, diagnosis, datamining, and cancer" were the most frequently used words in the scientific productions of Iranian researchers in the field of AI.

According to the results of the implementation ofthe LDA algorithm on the data, five main topic clusters were obtained for the Iranian researchers' scientific publications in AI for medical sciences. Table 1 presents the most important words of each topic cluster along with its word cloud.

According to Table 1, the following labels were selected for each of the topic clusters.

Topic 0: "Modeling and classification of medical images."

Topic 1: "Diagnosis and modeling in digital medicine."

Topic 2: "Cancer diagnosis using deep learning."

Topic 3: "Modeling and prediction in medical systems."

Topic 4: "AI, learning, and training."

Figure 2 indicates the percentages of Iranian researchers' scientific publications in AI for medical sciencesin each of the topic clusters.

Figure 2 indicates that the highest percentage (33%) of Iranian researchers' scientific publications in AI for medical sciences was in Topic Cluster 3, entitled "modeling and prediction in medical systems," and the lowest percentage (8.6%) was in the Topic Cluster 1entitled "modeling and classification of medical images."

Figure 3 indicates the annual trend of topic clusters of Iranian researchers' scientific publications in AI for

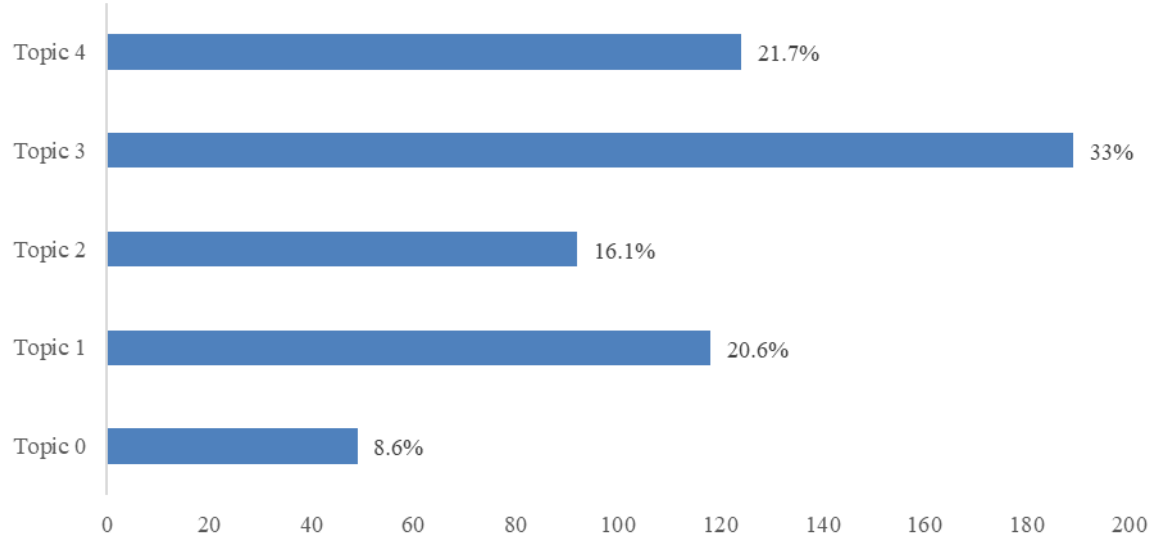


Figure 2. Percentage of Iranian Researchers' Scientific Publications in AI for Medical Sciences in each of the Topic Clusters.

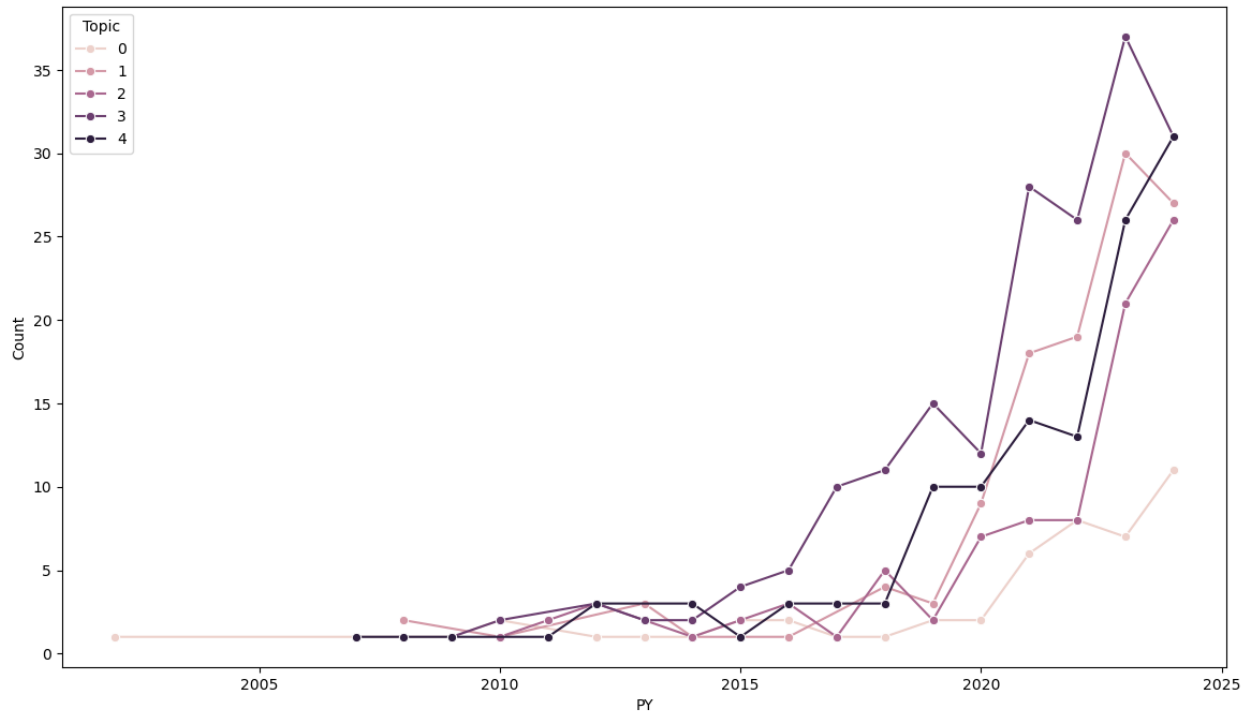


Figure 3. Annual Trend of Topic Clusters of Iranian Researchers' Scientific Publications in AI for Medical Sciences.

medical sciences.

Figure 3 indicates that Topic Cluster 0, entitled "modeling and classification of medical images," has been published since 2002, although the annual publication rate has been low. In addition, Topic Cluster 3, entitled "modeling and prediction in medical systems," has allocated a high publication rate in recent years. After that, Topic Cluster 1, entitled "diagnosis and modeling in digital medicine," and Topic Cluster 4, entitled "AI, learning, and training," have recently involved the majority of publications.

Discussion

The present study examined the publication patterns and trends of Iranian researchers' scientific publications in AI for medical sciences. AI, as one of the advanced technologies in medicine, plays a fundamental role in improving the quality of diagnosis, treatment, and management of diseases.[22] Due to the increasing complexity of medical data and a need for precise and

rapid analyses of them, the use of AI and machine learning tools has been considered an efficient solution.[23]

The results of this study reveal the significant focus of Iranian researchers on various applications of AI, including disease diagnosis, prediction of clinical outcomes, and extraction of knowledge from medical data. In this regard, some topics, including "machine learning" and "data mining," are employed as prominent methods in analyzing complex medical data and discovering new relevant patterns. In addition, the widespread use of "diagnosis" depicts that Iranian researchers have made many efforts to develop algorithms and diagnostic tools based on AI, which can lead to an increase in precision and speed in the diagnosis of disorders. Moreover, a focus on "Cancer" indicates that researchers are actively working on new solutions based on AI to manage and treat related conditions. In this regard, the literature emphasizes that applications related to AI and machine learning in cancer detection and treatment potentially provide therapeutic support for faster planning of a new treatment for each patient.[24] Musa et al. (2022) revealed that the number of studies related to the use of AI and machine learning in cancer-related applications is progressing and expanding.[25]

The results of the LDA analysis on the Iranian researchers' scientific publications in AI for medical sciences indicate the existence of five main topic clusters. These clusters have separately dealt with various topics related to the applications of AI in medicine. The first cluster, "modeling and classification of medical images," deals with the application of machine learning algorithms and artificial neural networks for the analysis and segmentation of medical images, especially in cancer diagnosis. The second cluster, "diagnosis and modeling in digital medicine", focuses on improving the precision of disease diagnosis using AI models in medical data. The third cluster, "cancer diagnosis using deep learning," emphasizes the use of deep learning techniques to analyze medical images and improve the accuracy of cancer diagnosis, especially breast cancer. The fourth cluster, "modeling and prediction in medical systems," is dedicated to the application of data mining and machine learning in predicting diseases and improving the performance of medical systems, especially during the COVID-19 pandemic. Finally, the fifth cluster, "AI, learning, and training," examines the effects of AI in medical education and improving the skills of medical students and the impacts of the COVID-19 pandemic on this process. These outcomes reveal the main applications of AI in medical sciences as well as the interests and expertise of Iranian researchers in this field. In this regard, modeling and classification of medical images is the core of numerous medical image analysis/processing systems. Indeed, this topic defines the required processes, such as coding, compression, contrast enhancement, noise removal, feature extraction, classification, etc., for a medical image.[26] In addition, the topics related to diagnosis and machine learning are of great significance in the field of AI in medical sciences. In this context, machine learning has emerged as a powerful tool in cancer diagnosis, which provides a significant improvement in the accuracy and efficiency of diagnostic processes. Various machine learning techniques, including K-nearest neighbors, convolutional neural networks (CNNs), and support vector machines (SVMs), are employed to analyze medical images, molecular data, and clinical data to identify and classify cancer.[27,28]

Moreover, the focus of scientific production topics on high-risk diseases indicates the attempt to use AI technologies for early diagnosis and management of disease conditions, which has also been discussed in the literature.[29,30] In breast cancer diagnosis, machine learning algorithms have provided great hopes for early detection, high diagnostic accuracy, and planning in the field of personalized medicine. These tools can process large volumes of clinical, genomic, and imaging data to improve diagnostic accuracy and assist in treatment decisions.[31] Moreover, in lung cancer screening, AI-based systems have been developed to analyze CT scans, which can reduce the workload of physicians and reduce the percentage of missed diagnoses.[32]

The distribution of scientific production based on topic clusters indicates a significant focus on "modeling and prediction in medical systems," which accounts for the largest rate of the whole scientific productions of Iranian researchers in the field of AI in medical sciences. This highest rate of production among the clusters reveals the growing significance of AI in predicting and modeling diseases and medical systems.

The annual trend of the topics related to Iranian

researchers' scientific publications in AI for medical sciences shows important developments in various fields of research. In particular, "modeling and prediction in medical systems" has experienced a significant increase in publications in recent years, which may be attributed to the increasing applications and need for predictive models in the face of emerging health challenges, including the COVID-19 pandemic. While the "modeling and classification of medical images" and "diagnosis and modeling in digital medicine" have received less attention in recent years, the need to improve the accuracy and application of AI models in these fields is still evident. Previous findings have also indicated that after the COVID-19 pandemic, the use of digital technologies and AI in healthcare increased significantly. The most important applications have been in predicting disease outbreaks, rapid diagnosis through medical image analysis, and the development of drugs and vaccines.[33,35]

These findings are consistent with the results of this study, which highlight the growing attention to the field of "modeling and prediction in medical systems." This alignment shows that the global COVID-19 crisis acted as a catalyst, further revealing the undeniable role of AI in strengthening health systems' responses to crises and the pressing need for robust predictive models.

The implications of these findings suggest that Iranian researchers are contributing significantly to the global advancement of AI in medical sciences, particularly in predictive modeling and cancer diagnostics. The focus on predictive models could have transformative effects on healthcare delivery in Iran, enabling earlier interventions, optimizing resource allocation, and improving patient outcomes. For instance, AI-driven predictive models could assist in forecasting disease outbreaks or identifying at-risk populations, which is particularly critical in resource-constrained settings. However, the relative decline in focus on medical image analysis and digital medicine suggests a need to balance research efforts to maintain progress across all AI applications in healthcare.

Based on the findings, future research in Iran should focus on three specific areas: (1) expanding AI applications in preventive medicine and population health, (2) strengthening international research networks to increase global visibility and impact, and (3) developing ethical and regulatory frameworks to guide the safe adoption of AI in medical practice.

Conclusion

The results of the present study indicate that the focus of Iranian research in the field of AI in medical sciences is more inclined towards applications that are directly related to the improvement of prediction and modeling of diseases and medical systems. These outcomes show the expertise of Iranian researchers in the related domains. In addition, the trend towards emerging issues, such as medical education and the use of AI for the development of educational skills, presents attention to the broader applications of this technology in improving the quality of education and learning in medical sciences.

Furthermore, the results of the present investigation can help determine research priorities and identify areas that need more focus on the field of AI in medical sciences. In particular, the need to pay more attention to areas with low percentages of scientific production in these fields can lead to the improvement of the research situation and the development of new applications of AI in medicine.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgement

This study was supported by the Sirjan University of Medical Sciences [Grant Number 402SIR30 and Ethical code: IR.SIRUMS.REC.1403.019]. The authors gratefully acknowledge this financial support.

References

1. Chaurasia A. Algorithmic precision medicine: Harnessing artificial intelligence for healthcare optimization. Asian J Biotechnol Bioresour Technol. 2023;9(4):28-43.
2. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6(2):94-98. doi:10.7861/futurehosp.6-2-94
3. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. Int J Environ Res Public Health. 2021;18(1):271. doi:10.3390/ijerph18010271
4. Kaur C, Garg U. Artificial intelligence techniques for cancer detection in medical image processing: A review. Materials Today: Proceedings. 2023;81:806-9.
5. Zhang C, Lu Y. Study on artificial intelligence: The state of the art and future prospects. J Ind Inf Integr. 2021;23:100224. doi:10.1016/j.jii.2021.100224
6. Lötsch J, Ultsch A, Mayer B, Kringel D. Artificial intelligence and machine learning in pain research: a data scientometric analysis. Pain Rep. 2022;7(6):e1044. doi:10.1097/pr9.0000000000001044
7. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-8. doi:10.1038/s41591-

021-01614-0
8. Jang HJ, Cho KO. Applications of deep learning for the analysis of medical data. Arch Pharm Res. 2019;42(6):492-504. doi:10.1007/s12272-019-01162-9
9. Taye MM. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. Computers. 2023;12(5):91.
10. Pramod A, Naicker HS, Tyagi AK. Machine learning and deep learning: Open issues and future research directions for the next 10 years. Computational analysis and deep learning for medical care: Principles, methods, and applications. 2021:463-90.
11. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, et al. AI applications to medical images: From machine learning to deep learning. Physica Medica. 2021;83:9-24. doi:10.1016/j.ejmp.2021.02.006
12. Tapeh ATG, Naser M. Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. Arch Comput Methods Eng. 2023;30(1):115-59.
13. Tran BX, Vu GT, Ha GH, Vuong QH, Ho MT, Vuong TT, et al. Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. J Clin Med. 2019;8(3):360. doi:10.3390/jcm8030360
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. doi:10.2196/18228
15. Dastani M, Mardaneh J, Pouresmaeil O. Detecting latent topics and trends in global publications on brucellosis disease using text mining. Interdiscip Perspect Infect Dis. 2022;2022(1):7274734. doi:10.1155/2022/7274734
16. Cheerkoot-Jalim S, Khedo KK. A systematic review of text mining approaches applied to various application areas in the biomedical domain. J Knowl Manag. 2021;25(3):642-68.
17. Martinelli DD. Evolution of Alzheimer's disease research from a health-tech perspective: Insights from text mining. International Journal of Information Management Data Insights. 2022;2(2):100089. doi:10.1016/j.jjimei.2022.100089
18. Zengul FD, Zengul AG, Mugavero MJ, Oner N, Ozaydin B, Delen D, et al. A critical analysis of COVID-19 research literature: Text mining approach. Intell Based Med. 2021;5:100036. doi:10.1016/j.ibmed.2021.100036
19. Hickman L, Thapa S, Tay L, Cao M, Srinivasan P. Text preprocessing for text mining in organizational research: Review and recommendations. Organ Res Methods. 2022;25(1):114-46.
20. Blei DM, Lafferty JD. Topic models. Text mining: Chapman and Hall/CRC; 2009. pp. 101-24.
21. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res. 2003;3:993-1022.
22. Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nat Med. 2020;26(8):1224-8. doi:10.1038/s41591-020-0931-3
23. Rahmani AM, Yousefpoor E, Yousefpoor MS, Mehmood Z, Haider A, Hosseinzadeh M, et al. Machine learning (ML) in medicine: Review, applications, and challenges. Mathematics. 2021;9(22):2970.
24. Shastry KA, Sanjay H. Cancer diagnosis using artificial intelligence: a review. Artif Intell Rev. 2022;55(4):2641-73.
25. Musa IH, Afolabi LO, Zamit I, Musa TH, Musa HH, Tassang A, et al. Artificial Intelligence and Machine Learning in Cancer Research: A Systematic and Thematic Analysis of the Top 100 Cited Articles Indexed in Scopus Database. Cancer Control. 2022;29:10732748221095946. doi:10.1177/10732748221095946
26. Amini Z, Rabbani H. Classification of medical image modeling methods: A review. Curr Med Imaging. 2016;12(2):130-48.
27. Mitra D, Sharma N, Rashid M, Singh R, editors. Classification rules based breast cancer detection using machine learning approach. 2022 5th International Conference on Contemporary Computing and Informatics (IC3I); 2022: IEEE.
28. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell. 2023;186(8):1772-91. doi:10.1016/j.cell.2023.01.035
29. Houssami N, Kirkpatrick-Jones G, Noguchi N, Lee CI. Artificial Intelligence (AI) for the early detection of breast cancer: a scoping review to assess AI's potential in breast screening practice. Expert Rev Med Devices. 2019;16(5):351-62. doi:10.1080/17434440.2019.1610387
30. Leatherdale ST, Lee J. Artificial intelligence (AI) and cancer prevention: the potential application of AI in cancer control programming needs to be explored in population laboratories such as COMPASS. Cancer Causes Control. 2019;30:671-5. doi:10.1007/s10552-019-01182-2
31. Gandhi H, Kumar K. Artificial Intelligence for the Management of Breast Cancer: An Overview. Curr Drug Discov Technol. 2024;21(4):e031123223115. doi:10.2174/0115701638262066231030052520
32. Wu L, Li X, editors. Research on early screening of lung cancer based on artificial intelligence. Proceedings of the 2021 international conference on bioinformatics and intelligent computing; 2021.
33. Demişcan D, Lozan O. The use of artificial intelligence in coordinating COVID-19 prevention measures at the territorial level. Revista de Ştiinţe ale Sănătăţii din Moldova. 2024 (4):44-8.
34. Wang L, Zhang Y, Wang D, Tong X, Liu T, Zhang S, et al. Artificial intelligence for COVID-19: a systematic review. Front Med. 2021;8:704256. doi:10.3389/fmed.2021.704256