The Research Hotspots and Visualization Analysis Technology of Health Big Data

Weizhang Huang, Hanxiang Gong, Yajun Yang*
The Second Affiliated Hospital of Guangzhou Medical University

*Corresponding author: 2020682075@gzhmu.edu.cn

Abstract. By understanding the international research hotspots in this field, we hope to provide reference information for health big data researchers to grasp the research hotspots and frontiers. By searching the relevant literature in the field of international health big data research in the Web of Science database, using the method of bibliometrics to draw a visual map. Results A total of 3850 articles were retrieved. Since 2013, the number of health big data research papers has shown a rapid growth. Harvard Medical School, University of Michigan and Stanford University are the research institutions with the largest number of papers in this field. The United States is the most active in the field of big data research, with 1506 papers, far higher than other countries, followed by China, the United Kingdom and South Korea. International health big data research hotspots include research on Intelligent wearable devices, design and development of health big data monitoring system, and application of health big data in precision medicine. The research reveals the characteristics and hotspots of international health big data research field, and provides some references for the development direction of this field in China.

Keywords: International, Health big data, hotspot, visualization.

1. Introduction
Health and medical big data refer to the collection of all data related to medical treatment and life and health, as well as the data generated by patients during medical care [1]. With the rapid development of "Internet +", the amount of data generated in the health and medical field has grown rapidly. Big data technology is increasingly applied to research in the fields of biology and medicine, which can provide technical support for the health and medical field and promote the construction of medical information. At the same time, the use of medical and health big data analysis can strengthen the early diagnosis and early treatment of diseases, as well as the health care and nursing care of the population [2]. Health big data research is an important research content in the current academic and business circles, and related theories and application research results are increasing. In order to further promote the research of health big data, it is necessary to analyze the research hotspots and cutting-edge topics of health big data.
2. Information and methods

2.1. Source
Web of Science core collection using Web of Science database. Search conditions: TS= (health OR healthcare OR clinical OR medical OR medicine OR medical care) AND TS= “big data”, search time: May 4, 2021. The total number of documents: 3850, the document type is limited to Article, and the search period is not limited.

2.2. Research methods and tools
The research is based on bibliometric analysis, using software methods such as VOSviewer to systematically analyze the distribution of research institutions, author collaborations, and keyword distribution of health big data-related literature, and draw a visual map. VOSviewer software is a software developed by Professor Ludo Waltman of Leiden University in the Netherlands for visual analysis of documents. This method can more directly evaluate the research hotspots and related characteristics of documents.

3. Results

3.1. Time Distribution
The change in the number of publications can reflect the development status and different stages of development in a field. The number of publications in the international health big data research field is calculated according to the time distribution, and a broken line change chart is drawn (Figure 1). It can be seen from Figure 1 that international papers on health big data were first published in 2009. Since 2013, the number of publications has shown rapid growth and a large increase. Big data has attracted more and more attention in the field of medical and health, and has quickly become a research hotspot. One, the number of articles published in 2020 has reached 1,529. It can be predicted that in the next few years, health big data will continue to become a research hotspot in the medical and health field, and the international health field will have more in-depth and extensive research on big data.

![Figure 1. Changes in the number of papers in the field of health big data research](image)

3.2. Distribution of research institutions
Figure 2 and Table 1 show the distribution of research institutions in the field of health big data research and their publication volume. The research institutions with the most articles in this field are Harvard Med Sch, Univ Michigan and Stanford Univ, with 88, 78 and 72 articles respectively. The most frequently cited are Chinese Acad Sci, Stanford Univ and Univ Bristol, with 3110, 2236 and 2065 citations respectively. The most collaborative research institutions are Harvard Med Sch, Univ Penn and Univ Calif San Francisco. These data show that colleges and universities are the main force in promoting the process of health big data research.
Figure 2 The distribution of research institutions of related literature in the field of health big data research

Table 1 Top 20 research institutions in the field of health big data research

| Serial number | organization                      | documents | citations | total link strength |
|---------------|-----------------------------------|-----------|-----------|---------------------|
| 1             | harvard med sch                   | 88        | 1132      | 252                 |
| 2             | univ michigan                     | 78        | 894       | 191                 |
| 3             | stanford univ                     | 72        | 2236      | 144                 |
| 4             | chinese acad sci                  | 61        | 3110      | 82                  |
| 5             | univ penn                         | 57        | 1190      | 227                 |
| 6             | king saud univ                    | 55        | 1570      | 34                  |
| 7             | univ oxford                       | 53        | 584       | 91                  |
| 8             | univ calif san francisco          | 52        | 1664      | 191                 |
| 9             | duke univ                         | 47        | 662       | 153                 |
| 10            | univ calif los angeles            | 47        | 1019      | 122                 |
| 11            | univ washington                   | 47        | 1023      | 161                 |
| 12            | huazhong univ sci & technol       | 45        | 1519      | 46                  |
| 13            | univ toronto                      | 45        | 633       | 81                  |
| 14            | massachusetts gen hosp            | 42        | 921       | 123                 |
| 15            | columbia univ                     | 41        | 607       | 87                  |
| 16            | univ minnesota                    | 40        | 717       | 78                  |
| 17            | harvard univ                      | 38        | 1740      | 137                 |
| 18            | johns hopkins univ               | 38        | 903       | 123                 |
| 19            | univ melbourne                    | 37        | 943       | 48                  |
| 20            | emory univ                        | 36        | 601       | 113                 |

3.3. Country/Region Distribution
Figure 3 and Table 2 show the distribution of countries/regions in the field of health big data research and the volume of publications. It can be seen from Figure 3 and Table 2 that in foreign countries, the United States is the most active in the field of big data research on Taihealth, and its output is in a dominant position (the number is 1506), accounting for about 39.12% of the total, ranking first. It is much higher than other countries, further confirming the level and influence of its health big data research. China (850 articles, accounting for 22.08%), the United Kingdom (401 articles, accounting for 10.42%), and South Korea (263 articles, accounting for 6.83%) are next. Analyzing the total citation frequency and single citation frequency of hot papers in various countries, it is found that the United States still ranks first in terms of total citation frequency, with a total citation frequency of 29,835 times. The total citations of China, the United Kingdom, and Australia are followed by 15,711, 8,596, and 4,079, respectively.
Figure. 3 Distribution of countries/regions in the field of health big data research

Table. 2 Top 20 countries/regions in the field of health big data research

| Serial number | country          | documents | citations | total link strength |
|---------------|------------------|-----------|-----------|---------------------|
| 1             | usa              | 1506      | 29835     | 1329                |
| 2             | peoples r china  | 850       | 15711     | 695                 |
| 3             | england          | 401       | 8596      | 978                 |
| 4             | south korea      | 263       | 3861      | 282                 |
| 5             | australia        | 234       | 4079      | 543                 |
| 6             | canada           | 216       | 3846      | 407                 |
| 7             | italy            | 204       | 2742      | 589                 |
| 8             | india            | 198       | 2625      | 313                 |
| 9             | germany          | 178       | 3169      | 617                 |
| 10            | spain            | 155       | 2058      | 444                 |
| 11            | netherlands      | 153       | 2288      | 567                 |
| 12            | france           | 135       | 1816      | 468                 |
| 13            | saudi arabia     | 100       | 2262      | 155                 |
| 14            | taiwan           | 98        | 1771      | 112                 |
| 15            | switzerland      | 97        | 1519      | 324                 |
| 16            | japan            | 96        | 2036      | 295                 |
| 17            | sweden           | 72        | 965       | 383                 |
| 18            | scotland         | 70        | 1202      | 246                 |
| 19            | belgium          | 66        | 1410      | 286                 |
| 20            | brazil           | 61        | 859       | 240                 |

3.4. Research topic distribution

Figure 4 and Table 3 show the distribution and frequency of the subject terms of the literature in the field of health big data research. A total of 14,762 keywords were extracted from 3850 documents, with a total word frequency of 37,141 times, and 29 subject words with a word frequency greater than 100 times. It can be considered that these keywords can reflect the current hottest health big data research topics.
Figure. 4 The distribution of the subject words of the literature in the field of health big data research

Table. 3 Top 40 keywords appearing frequently in the subject headings of the literature in the field of health big data research

| Serial number | keyword | occurrences | total link strength | Serial number | keyword | occurrences | total link strength |
|---------------|---------|------------|---------------------|---------------|---------|------------|---------------------|
| 1             | big data| 1769       | 7476                | 21            | privacy | 112        | 576                 |
| 2             | machine learning | 339 | 1820              | 22            | electronic health records | 109 | 619            |
| 3             | health  | 259        | 1140                | 23            | precision medicine | 109 | 574            |
| 4             | risk    | 223        | 1077                | 24            | medicine | 107 | 601            |
| 5             | classification | 190 | 959               | 25            | diagnosis | 106 | 565            |
| 6             | care    | 184        | 990                 | 26            | cancer    | 105 | 479            |
| 7             | system  | 162        | 831                 | 27            | information | 102 | 552            |
| 8             | deep learning | 157 | 767               | 28            | big data analytics | 100 | 436            |
| 9             | internet | 155 | 997                | 29            | impact    | 100 | 515            |
| 10            | challenges | 151 | 923                | 30            | epidemiology | 98  | 434            |
| 11            | model   | 149        | 783                 | 31            | internet of things | 98  | 600            |
| 12            | health-care | 147 | 726                | 32            | outcomes  | 98  | 486            |
| 13            | prediction | 147 | 789                | 33            | healthcare | 97  | 502            |
| 14            | artificial intelligence | 146 | 764 | 34 | analytics | 93 | 509 |
| 15            | management | 145 | 802 | 35 | association | 92 | 492 |
| 16            | disease | 137        | 719                 | 36            | cloud    | 84  | 423            |
| 17            | cloud computing | 136 | 656 | 37 | things | 84 | 598 |
| 18            | mortality | 134 | 709                | 38            | systems  | 83  | 440            |
| 19            | data mining | 130 | 613                | 39            | algorithm | 79 | 384            |
| 20            | framework | 118 | 661                | 40            | design    | 78  | 390            |
3.5. Distribution of research institutes
The color system in the keyword superimposed time view in Figure 5 indicates that the displayed color represents the time when the node appears, the research hotspots, evolutionary context, and the prediction of the research trend of the topic in the same color area. According to Figure 4 and related literature on health big data research, it can be seen that the research and practice of health big data presents three stages of practice evolution: 2014 and before, 2015-2018, and 2019 to the present. Among them, 2019 has been an important year for health big data research, with the expansion and scale of research hotspots.

![Figure 5](image)

**Figure 5** A visual map of the key word overlay time of the literature in the health big data research field

4. An Analysis of Hotspots in International Health Big Data Research

4.1. Research on Smart Wearable Devices
With the continuous improvement of people's health awareness, their attention and demand for various intelligent health service management tools continue to increase. Medical and health wearable devices and big data technology have created huge development opportunities for the medical and health industry [3-4]. Portable mobile medical and health wearable devices can collect users' health data and behavior information in real time through the application of various emerging information technologies. They are an important data collection source for smart medical service platforms based on medical and health big data analysis. Portable mobile medical and health wearable devices and their applications have become the focus of medical and health service-related companies and research institutions. Research on smart medical services based on smart wearable devices is becoming one of the frontier research topics in the field of health big data.

4.2. Research on Design, Development and Application of Health Big Data Monitoring System
With the advancement of wireless communication technologies such as wearable and implantable biosensors, it is possible to promote the design, development and implementation of the human body local area network. The human body local area network has paved the way for the development of
innovative applications for medical health monitoring [5]. At present, the representative application fields of big data include enterprise management, Internet of Things, online social networks, media applications, group wisdom and smart grid [6]. The medical and health field is rich in big data resources, and as people’s health awareness has increased, various health monitoring systems have emerged. Medical health management is increasingly becoming a key area for the application of big data technology, and the design and development of health big data application systems has gradually become the focus. Research topics of interest.

4.3. Application of health big data in precision medicine

In 2015, US President Barack Obama proposed the "Precision Medicine" plan, and precision medicine began to attract the world's attention. With the rapid development of high-throughput biotechnology, various omics, bioinformatics, and systematic methods, precision medicine is gradually becoming a new paradigm for future medical development [7]. Precision medicine affects the development of the medical field in the form of network, promotes the cross-cooperation of biomedicine and information science with new ideas, integrates data-driven scientific research paradigm, and strives to overcome many shortcomings of the traditional symptom-oriented disease diagnosis and treatment model, thereby Effectively use medical resources and improve medical standards.

References
[1] Liu Ning, Chen Min. Research on the topic of medical and health big data application and related data sources [J]. China Digital Medicine, 2016, 11(8): 6-9.
[2] Li Chunying, Zhang Weiwei. Cluster analysis of research hotspots in global big data and health management [J]. Chinese Hospital Management, 2016, 36(10): 63-65.
[3] ROSKI J, BO-LINN G W, ANDREWS T A. Creating Value in Health Care through Big Data: Opportunities and Policy Implications [J]. Health Affairs, 2014, 33 (7) :1115-1122.
[4] WU J, LI H, CHENG S R, et al. The promising future of healthcare services: When big data analytics meets wearable technology [J]. Information&Management, 2016, 53 (8):1020-1033.
[5] CHEN M, GONZALEZ S, VASILAKOS A, et al.Body Area Networks:A Survey [J].Mobile Networks and Applications, 2011, 16 (2) :171-193.
[6] CHEN M, MAO S, LIU Y. Big Data: A Survey [J]. Mobile Networks and Applications, 2014, 19 (2) :171-209.
[7] CHEN R, SNYDER M. Promise of personalized omics to precision medicine [J]. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2013, 5(1): 73-82.

Liu Zhuang, Zhang Yue, Tian Lei, Sun Baozhi, Zhao Yuhong. Analysis of influential factors on self-evaluation of post competency of clinicians in northwest China [J]. China health statistics, 2017,34(03): 397-399 + 403.