

Tracking knowledge evolution, hotspots and future directions of emerging technologies in cancers research: a bibliometrics review

Xiaoyu Wang¹, Jingjing Guo², Dongxiao Gu^{2,3*}, Ying Yang², Xuejie Yang², Keyu Zhu²

¹ The 1st Affiliated Hospital, Anhui University of Traditional Chinese Medicine, Anhui 230031, China.

² The School of Management, Hefei University of Technology, Hefei, Anhui 230009, China.

³ Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui, 230009, China

* Corresponding authors. Email: xywang0551@gmail.com.

Abstract

Due to various environmental pollution issues, cancers have become the “first killer” of human beings in the 21st century and their control have become a global strategy of human health. The increasing development of emerging information technologies has provided opportunities for prevention, early detection, diagnosis, intervention, prognosis, nursing, and rehabilitation of cancers. In recent years, the literature associated with emerging technologies in cancer has grown rapidly, but few studies have used bibliometrics and a visualization approach to conduct deep mining and reveal a panorama of this field. To explore the dynamic knowledge evolution of emerging information technologies in cancer literature, we comprehensively analyzed the development status and research hotspots in this field from bibliometrics perspective. We collected 7,136 articles (2000–2017) from the Web of Science database and visually displayed the dynamic knowledge evolution process via the analysis on time-sequence changes, spatial distribution, knowledge base, and hotspots. Much institutional cooperation occurs in this field, and research groups are relatively concentrated. BMC Bioinformatics, PLOS One, Journal of Urology, Scientific Reports, and Bioinformatics are the top five journals in this field. Research hotspots are mainly concentrated in two dimensions: the disease dimension (e.g., cancer, breast cancer, and prostate cancer), and the technical dimension (e.g., robotics, machine learning, data mining, and etc.). The emerging technologies in cancer research is fast ascending and promising. This study also provides researchers with panoramic knowledge of this field, as well as research hotspots and future directions.

Keywords: cancers, emerging information technology, hotspots, knowledge evolution, bibliometrics.

1. Introduction

Various environmental issues especially air and water pollution combined with widespread use of food additives and pesticides make cancer the top killer in lots of countries. Cancer is a group of diseases that involve abnormal cell growth with the potential to invade or spread to other parts of the human body. Under the influence of various tumorigenic factors, some cells in local tissue usually lose normal control of their growth at the gene level, thereby forming new organisms with clonally abnormal proliferation. They also often manifest as local lumps [1]. The World Health Organization predicted that cancer will become the “first killer” of human beings in the 21st century. Hence, the control of cancer has become a global health strategy [2]. Tumors not only threaten people’s health but also impose a heavy burden on families, society, and countries; thus, they disrupt social and economic development [3]. This prominent issue of public health is an important topic for many scholars [4].

The increasing development of emerging information technologies, such as big data [5], cloud computing [6], the Internet of Things (IoT) [7], and wearable devices [8], has provided opportunities for prevention, early detection, diagnosis, intervention, prognosis, nursing, and rehabilitation of cancers. The increasing number of successful applications of emerging technologies in medicine has provided people with many tools and ways to collect, share, analyze, and utilize health data resources associated with cancers [9]. This phenomenon also provides us with unprecedented opportunities to use big data processing for understanding the genetic characteristics and pathogenesis of cancers and to bring accurate and powerful guidance to individual prevention or treatment of cancers [10].

Currently, research on information technologies for cancer is facing historical opportunity. An increasing number of scholars are devoting themselves to this research field. For example, Parkin et al. [11] investigated the incidence, mortality, and prevalence of 26 cancers worldwide in 2002. Lin et al. [12] developed a machine-learning model to predict breast cancer and assist medical team to make decisions. Gabriela Jurca analyzed genetic breast cancer trends by integrating text mining, data mining, and network analysis [13].

However, no study driven by subjective belief instead of scientific evidence has focused on the current status and future direction of the entire emerging information technologies in cancer. Existing studies are also addressing the application of a specific technology in a certain type of cancer rather than from the scope of all kinds of cancers [14,15]. Emerging information technologies do not indicate a specific technology but various technologies as well [16]. As a result, a holistic understanding of the concept of emerging technologies in cancer has been limited. The lack of comprehensive understanding of emerging technologies in cancer has disturbed an appropriate allocation of research human resources into this area, which causes inefficient works to a great extent.

To fill in this research gap, this study provides a visual analysis for the development status, hotspots, and trend of emerging information technologies’ applications in cancer research by using 7,136 articles from Web of Science database [17]. The main findings of this study are as follows: (a) this number of published articles in this field changed from 49 in 2000 to 1197 in 2017; (b) much institutional cooperation occurs in this field, and research groups are relatively concentrated; (c) the core authors and major contributors in this field

are Menon M, Li CF, Wang J, Madabhushi A, Patel VR, Kim S, and etc. *BMC Bioinformatics*, *PLOS One*, *Journal of Urology*, *Scientific Reports*, and *Bioinformatics* are the top five journals in this field; (d)research hotspots are mainly concentrated in two dimensions: the disease dimension (e.g., cancer, breast cancer, and prostate cancer), and the technical dimension (e.g., robotics, machine learning, data mining, and etc.). The emerging technologies in cancer research is fast ascending and promising. Scholars from different nations and institutes should further enhance exchange and cooperation. The dynamic knowledge structure of emerging technologies' application in cancer research is visualized and unveiled, which will be helpful to understand current state-of-art developments [18] [19]. Future research directions of this research field for scholars in e-health and medical informatics are also discussed.

The rest of the paper is organized as follows. The next section introduces the methodology we used in this study. Section 3 introduces the knowledge map of the time-and-space analysis. Section 4 presents the research hotspot analysis. The last section elaborates the conclusions of the study and presents the trends for future research.

2. Methodology

2.1 Data source

The data we used are collected from Web of Science, which is an authoritative retrieval platform for scientific documents worldwide [20]. We collected data on articles published in core international journals between January 1, 2000 and January 25, 2018 from the *Web of Science* (SCI-EPANDED, CPCI-S, CCR-EXPANDED, and Index Chemicus included for our literature analysis).

We performed a literature search on emerging technologies in cancer research. The search strategy we used is (cloud computing OR big data OR Artificial Intelligence OR internet of things OR wearable device* OR #) AND (tumor* OR cancer* OR neoplasm OR ##) AND the literature type: (Article OR proceedings paper). “#” represents more than 14 keywords related to emerging information technologies, and “##” represents the 11 keywords related to tumors. A panel consisting of four reviewers screened all titles and abstracts for potentially eligible studies. Full texts of all potentially eligible studies were then assessed by the same reviewers. We eventually obtained 7,136 journal articles.

2.2 Toolkits

Bibliometrics is a cross-discipline that quantitatively analyzes all knowledge carriers with the academic literature as research object by using mathematics, statistics, and philology comprehensively [21,22]. Using bibliometrics, we can mine the tacit knowledge in literature and obtain hotspots and future trends of a certain discipline [23,24]. In this study, we used *Excel 2010*, *Histcite*, and *CiteSpace*.

CiteSpace is citation-based visual analysis software based on the JAVA platform that focuses on analyzing the hidden knowledge in scientific research literature and becoming popular in the context of scientometrics, data and information visualization. This tool can calculate the number of articles in a specific field and explore critical paths of knowledge

evolution on the basis of co-citation analysis and PFNET [25], then through the drawing of a series of visual maps, the analysis of the potential dynamic mechanism of the subject evolution and the exploration of the frontier of the discipline development are formed. In this study, the data obtained in Web of Science is processed and imported into citespace software. After selecting the parameters, the subject distribution and cooperation analysis, journal co-citation analysis and keyword co-occurrence analysis were successively completed, and then the distribution structure, evolution path, key turning points and development rules of subject domain knowledge were displayed by visual graphs. The nodes represent the objects of the analysis, and the more frequencies (or citations) appear, the larger the nodes. The color and thickness in the inner circle of the node indicate the frequency of occurrence in different time periods. The line between the nodes represents a co-occurrence (or co-citation) relationship, and the thickness of the nodes indicates co-occurrence (or co-citation) strength. Centrality is an indicator of the importance of measuring nodes in the network (in addition to degree centrality, proximity, etc.). This metric is used in *CiteSpace* to discover and measure the importance of the literature, and to highlight the literature (or authors, journals, institutions, etc.) with a purple circle. In this study, we used *CiteSpace III* to analyze the knowledge structure and basis of emerging technologies in cancer research. We identified major issues in this field by using co-occurrence network maps and co-occurrence time zone maps of keywords obtained by *CiteSpace III*.

3. Knowledge map of time-and-space analysis

3.1 Time distribution map

To evaluate the outcomes of emerging technologies in cancer research, we tracked annual publishing trends and changes (Fig. 1). In 2000, 49 articles were published in this research field. Fig. 1 shows the growth trend of related literature between 2000 and 2017. Annual volume of literature has generally grown exponentially, and the number of articles associated with emerging technologies in cancer is increasing vigorously.

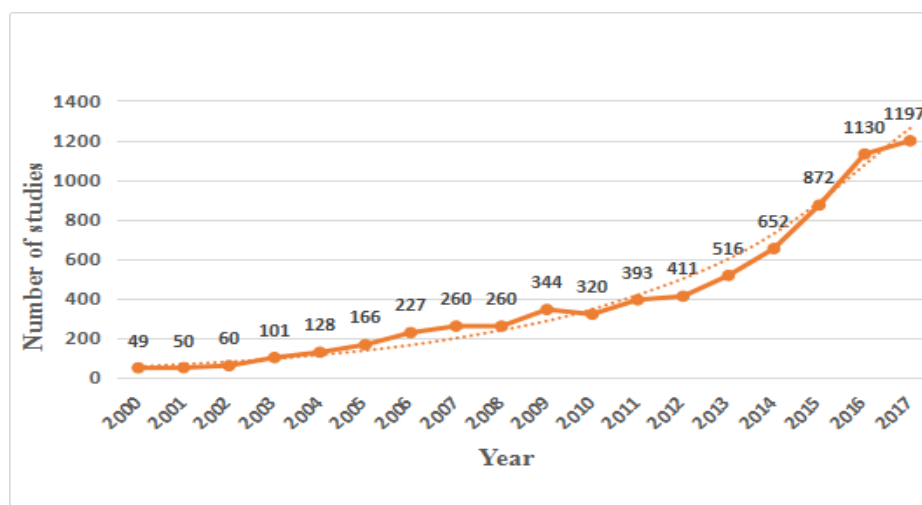


Fig. 1. Annual number of published articles.

Fig. 2 shows the annual total number of authors. The trend of number of authors is consistent with that of number of articles. A total of 208 authors published works in 2000. By 2017, the number grew to 6276, an increase of 30 times that in 17 years ago.

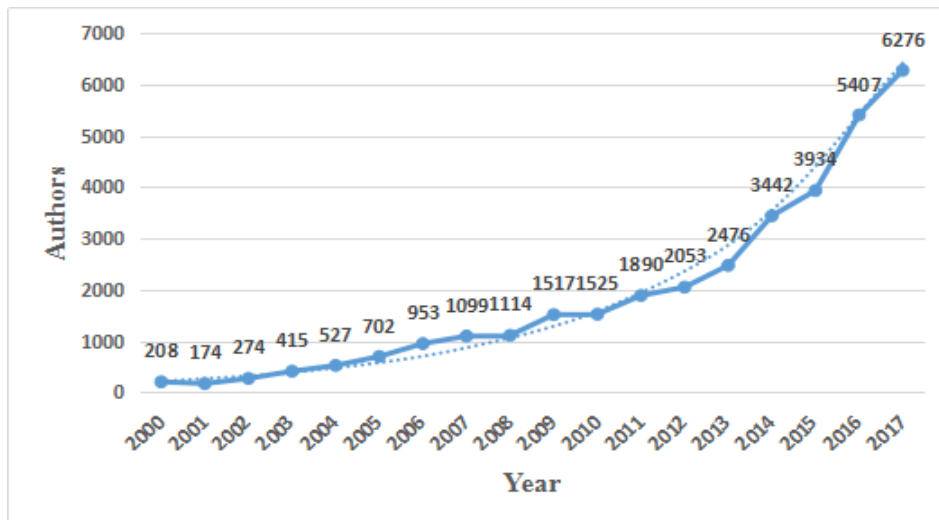


Fig. 2. Annual number of authors.

To evaluate the input–output ratio of researchers in the field, we summarized the average number of co-authors per article (Fig. 3). In 2000, the average number of co-authors per article was 4.24. This number increased to 5.24 in 2017. Fig. 3 reveals a trend of collaboration among authors in the field. The complexity and comprehensiveness of scientific research has led academic researchers to gradually abandon the traditional individual combat model. The trend of scientific research cooperation is unavoidable and gradually increasing. Only by continuously promoting positive cooperation among academic personnel from all walks of life can we resolve more academic issues and promote the development of scientific research. Scientific cooperation is unstoppable, and the scale and scope of scientific research cooperation activities will continue to grow with the rapid development of academic activities [26,27,28].

Collaboration, to some extent, ensures the quality of the published articles. The trend in Fig. 3 reflects the emphasis on the application research of emerging information technologies in cancer.

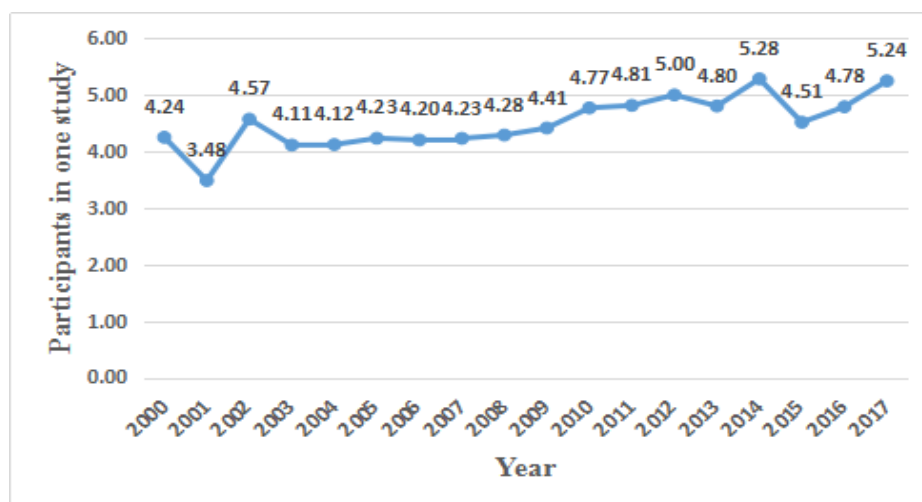


Fig. 3. Average number of co-authors per article.

3.2 Space distribution map

3.2.1 Institutional distribution

Table 1 lists the top nine academic groups and institutions in terms of number of published journal articles on emerging technologies in cancer. Harvard University with 112 papers ranks first. Stanford University with 30 articles ranks second. National Cancer Institute and Johns Hopkins University with 89 and 86 articles, respectively, occupy the third and fourth. Five other institutes with 71–79 articles rank fifth to ninth. Except University of Toronto from Canada and Yonsei University from South Korea, all seven other institutes are from the United States. Therefore, the institutions in the United States show strong research potential and academic exchange awareness in the research field of emerging technologies in cancer.

Citation frequency in the *Histcite* system can be divided into two. One is *Local Citation Score (LCS)*, and the other is *Global Citation Score (GCS)* [25]. LCS refers to the citation frequency of an article in the local database, whereas GCS refers to the citation frequency of an article in the Web of Science database. Except *University Toronto* with an LCS of 37, the LCS of all eight other institutions is above 90. Therefore, these institutions exert important influence in the research area of emerging technologies in cancer.

Table 1. Institutions and the number of articles published (70 or more).

Institution	No. of published articles	Local Citation Score (LCS)	Global Citation Score (GCS)
Harvard University	112	213	5307
Stanford University	90	145	2692
National Cancer Institute	89	191	3172
Johns Hopkins University	86	174	2732
University Michigan	79	130	2496
University Penn	79	120	1614
University Toronto	77	37	1228
University Texas MD Anderson Cancer Center	76	93	1049
Yonsei University	71	125	1093

Research collaboration is an important means of enhancing overall research strength because it allows researchers to complement one another's advantages and share experience and knowledge. The level of research collaboration is an important index to assess the state of research in a specific research field. We set relevant parameters with CiteSpace and drew an

institutional collaboration network, as shown in Fig. 4, to study institutional collaboration in emerging technologies in cancer research. $N = 417$, $E = 553$, and the density is 0.0064. N represents the number of network nodes, E represents the number of connections, and Density represents the Density of the network. These values indicate that some collaboration occurs across the institutions. Therefore, much institutional cooperation occurs in emerging technologies in cancer, and the research groups are relatively concentrated. Fig. 4 shows that the number of articles published from the major institutes in this research is increasing, and the cooperation is very close. Overall, emerging technologies in cancer research are still in the early but fast developing stage.

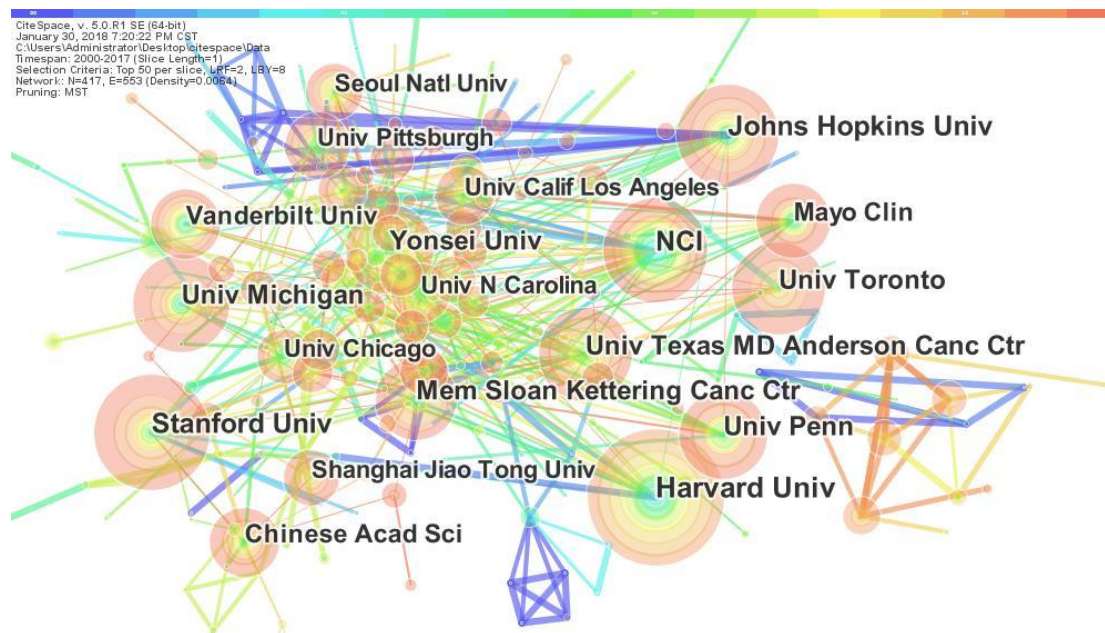


Fig. 4. Institutional collaboration network.

3.2.2 National/regional distribution

In terms of the total number of articles published (Table 2), *the United States* tops the list with 2711 articles, followed by *China* and *the United Kingdom* with 756 and 569 articles, respectively. In terms of centrality, *the United Kingdom* ranks first with a ratio of 0.31, and *the United States* ranks second with a ratio of 0.25. *Australia* and *Spain* are tied for third with a ratio of 0.14. Centrality describes the importance of a node among other nodes. In terms of the *Local Citation Score (LCS)* [29], *the United States* ranks first with a citation frequency of 4352, and *the United Kingdom* follows with 754. The average number of citations in China is relatively low. The revelation from this analysis is that although China has a large number of publication in cancer field, its actual influence is relatively low. There is still a need for further improvement in the quality of academic research achievements.

Table 2. Countries/regions with 200 or more published articles.

Nation or region	No. of published articles	LCS	Centrality
USA	2711	4352	0.25

Mainland China	756	357	0.05
UK	569	754	0.31
India	393	118	0.13
Germany	392	503	0.01
Italy	374	459	0.05
Canada	330	277	0.09
South Korea	270	299	0.02
France	262	267	0.11
Spain	236	331	0.14
Taiwan(region)	220	299	0.02
Australia	213	236	0.14

To study international collaboration distribution in this field, we set the relevant parameters using *CiteSpace III* and formed an international collaboration network (Fig. 5). A total of 87 countries with 376 collaboration edges are found with a density of 0.1005. Therefore, international research cooperation between countries or regions is relatively close. According to the connection between various nodes, most countries have a cooperative relationship with the United States, indicating that the United States attaches great importance to exchanges and cooperation in the academic community. This also explains to some extent why the United States has such a high output of research results.

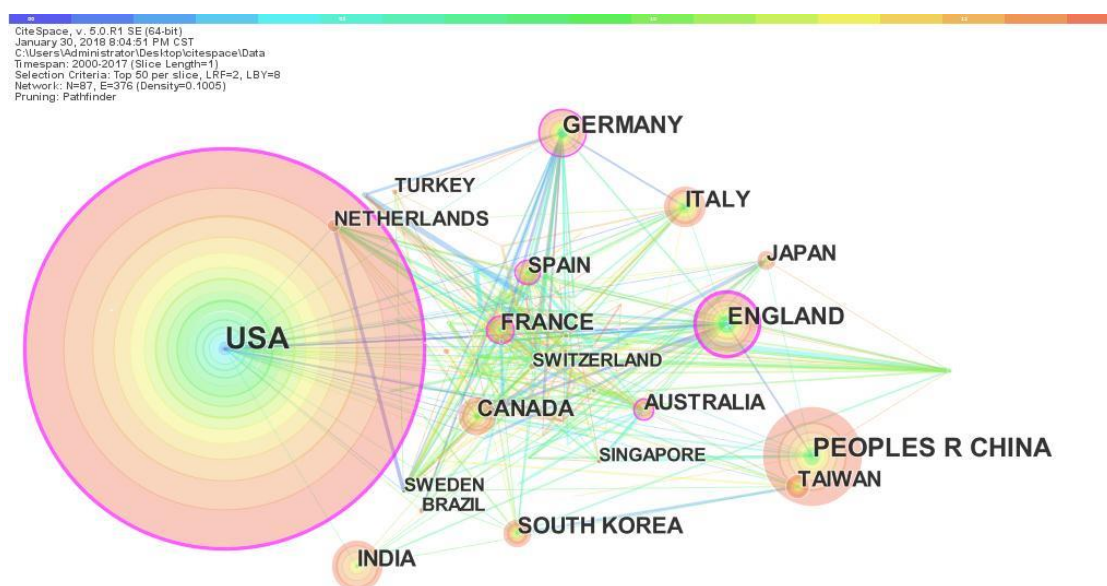


Fig. 5. National collaboration network.

3.2.3 Author distribution

By using the *HistCite* and *Excel* tools, we analyzed 7,136 related articles with a total of 26,020 authors. The ratio of authors with only one publication is as high as 77.99%. Notably, the vast majority of researchers in this field do not have continuous academic research and are only “passers-off.” Table 3 lists the researchers who contributed greatly in this field.

These researchers are the core authors and major contributors in this field. Menon M not only ranks first in the number of articles published but also presents a considerably high LCS. Therefore, he is a very important scholar in this research area. Li CF, Wang J, and Madabhushi A rank second, third, and fourth with 35, 33, and 29 postings, respectively, with obviously lower LCS than Patel VR. The spread rate of the articles published by Patel VR is higher than that of all other authors. His articles are likely to be with academic innovation.

Table 3. Top 10 authors.

Authors	Recs	LCS	GCS
Menon M	44	395	3792
Li CF	35	69	232
Wang J	33	39	252
Madabhushi A	29	38	363
Patel VR	28	217	1916
Kim S	27	10	212
Li J	25	23	518
Liu Y	24	14	161
Hemal AK	22	65	474
Wang Y	22	47	429

Fig. 6 shows the co-author network knowledge map, which describes the cooperation relationship among researchers in the field of emerging technologies in cancer. A large number of nodes are found in the figure, which indicates that the research field has received extensive attention from researchers and that the number of co-authored publications in this research field has tremendously and continuously increased. These collaborations have occurred within and across countries or regions. Co-authorship ties in this research field are generally weak and inconsistent. They consist of many collaborative groups of varying size; some are interconnected, and others are isolated from all others. Some researchers within these various groupings play a primary role, whereas others have a secondary one. Obviously, academic exchanges and research cooperation should be further reinforced in this field.

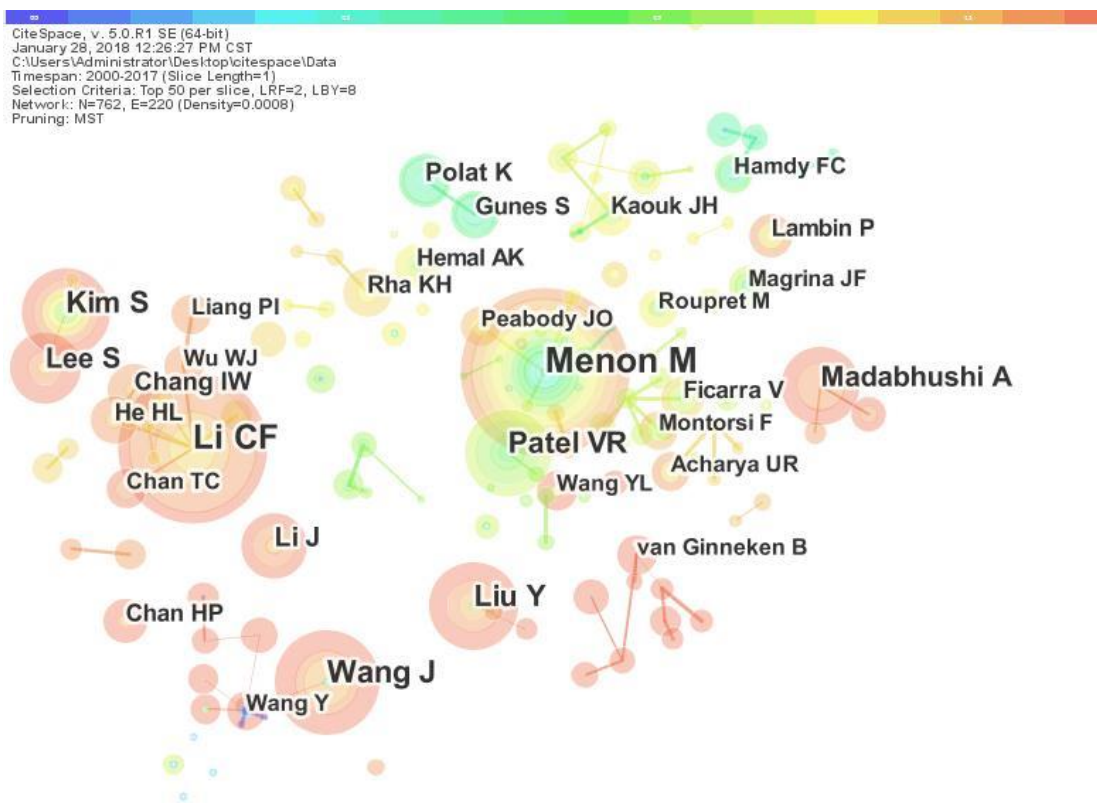


Fig. 6. Co-author network in emerging technologies in cancer literature.

3.2.4 Journal distribution

By analyzing the distribution of journals, scholars can easily understand the core and marginal journals, the spatial distribution of documents, and the preferences of journals published in the field, which will provide academic researchers with a reliable reference source [30]. The core journals are a type of intensive information source. They have important reference value for scholars to determine in which journals they should mainly focus and read articles published and to where they should submit their research achievements [31].

Table 4 shows that 15 journals with the most published articles were listed and 1083 articles were published, accounting for 15.18% of total publications in this field. Among them, *BMC Bioinformatics*, *PLOS One*, *Journal of Urology*, *Scientific Reports*, and *Bioinformatics* are the top five journals. Their impact factors are relatively high, which indicates that their influence is relatively large [32]. They are core journals in the research field of emerging technologies in cancer [33].

Table 4. Top 15 journals with most published articles related to emerging technologies in cancer.

NO	Journal	Recs	LCS	GCS
1	BMC Bioinformatics	133	0	3872
2	PLOS One	128	0	1385
3	Journal of Urology	116	719	5873

4	Scientific Reports	83	0	291
5	Bioinformatics	80	528	5305
6	Expert Systems with Applications	78	298	1692
7	Journal of Biomedical Informatics	71	84	926
8	BJU International	67	241	1697
9	Oncotarget	65	3	112
10	European Urology	59	518	4565
11	Medical Physics	58	85	628
12	Artificial Intelligence in Medicine	56	270	1657
13	Computers in Biology and Medicine	46	95	498
14	International Journal of Medical Robotics and Computer Assisted Surgery	43	68	603
15	BMC Genomics	42	0	375

4. Research hotspot analysis

Keywords summarize the major content, academic thoughts, and principal research methods of the researchers, which are the core and essence of literature. Frequently appearing keywords are often used to identify major topics in a research field [34]. In order to understand the current knowledge structure and hot topics in the field of emerging technologies in cancer, this paper extracts and calculates the frequency of key words in 7,136 articles. Table 5 shows the TOP10 keywords and the co-occurrence frequency. The most frequent occurrence of emerging technologies in cancer is cancer. This result is consistent with the research object of this paper. It shows that the core of research in this field must be related to cancer. It can be seen from the table that high frequency keywords: breast cancer, prostate cancer, etc. have become the hot issues of emerging technologies in cancer. Among them, breast cancer is a prominent problem of cancer. Al-Hajj, M. etc. (2003) mentioned that Breast cancer is the most common malignancy in women, accounting for many deaths each year [35]. There is a relatively high co-occurrence frequency of keywords such as classification, robotics, machine learning, data mining, support vector machines, etc. They represent the main techniques or methods applying in the field of cancer knowledge discovery and management. Various cancer classifiers based on data analysis and mining are available now. Many of them have been successfully applied to solve complex issues, such as cancer early detection, diagnosis, treatment, prognosis, and even nursing care after operations [36]. In the first 17 years in this century, scholars worldwide have attached great importance to the development of robotics technologies [37]. In 2009, VG Mallapragada pointed out that the use of medical robots to assist real-time tumor treatment results in fewer traumas to patients and can improve the quality of surgical treatment [38]. In recent years, scientists and engineers in the industry have expanded robotics' applications in cancer. Machine learning plays a very

important role in the research field of emerging technologies in cancer. Cancers are among complex diseases, and machine learning plays an important role in their treatment and diagnosis [39]. Behrmann (2017) mentioned that tumor data are very complex and that machine learning plays an auxiliary role in the classification of tumor images and has a good clinical application prospect [40]. In 2001, WJ Kuo used data mining techniques to improve the accuracy of tumor diagnosis [41]. Therefore, data mining occupies an important position in the research field of emerging technologies in cancer.

Table 5. Co-occurrence frequency TOP10 keywords

Keywords	Frequency	Centrality
cancer	1378	0.44
classification	1088	0.49
breast cancer	800	0.09
robotics	784	0.43
machine learning	674	0.15
data mining	607	0.09
prediction	465	0.11
support vector machine	403	0.14
prostate cancer	380	0.04
feature selection	372	0.02

Fig. 7 shows the keyword co-occurrence network map in the research field of emerging technologies in cancer. Each node in the graph represents a keyword. The size of the node is proportional to the co-occurrence frequency. The connection between the nodes represents the co-occurrence relationship between the two keywords in the same document. Different colors indicate the keywords of the year appear together. The number of network nodes is 191, the number of connections between nodes is 666, and the density of the network is 0.0367. According to this figure, there are strong connections between the keywords, and the entire network is densely connected. It shows that the most of papers published in this research field are multi-topic research.

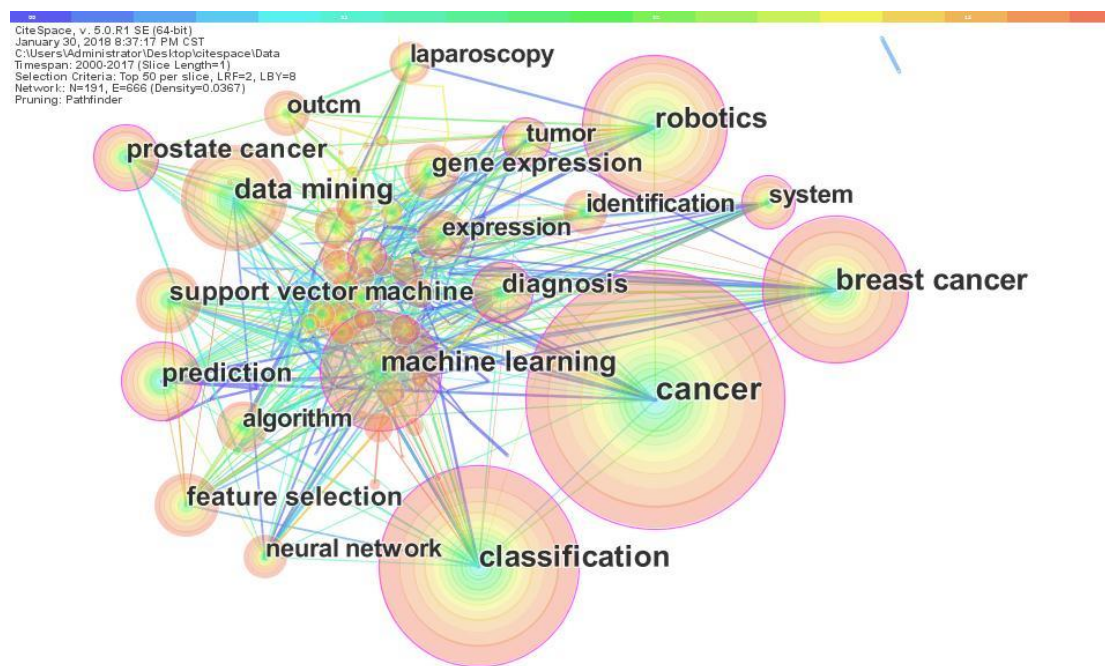


Fig. 7. Keywords co-occurrence network

5. Conclusions and future trends

5.1 Concluding remarks

The main work and findings of this study are as follows: (a) the trend of research output and the changes in co-author numbers in papers in the research field of emerging technologies in cancer is obtained using a time distribution map analysis; (b) the distribution of research results worldwide is determined using a national collaboration network; (c) core literature and authors in this field are determined using an article co-citation network; (d) scholars in medical informatics and e-health community are provided with an overview of hotspots of emerging technologies in cancer. Our results will provide these researchers with an important background for their future research.

In terms of the time distribution map of research, the number of articles and annual authors is increasing. Emerging technologies in cancer research are in the early but fast developing stage. From the average number of co-authors per article, we can find a trend of collaboration among authors in the field is increasing. However Co-authorship ties in this research field are still generally weak and inconsistent at present. The mutual cooperation between the cooperative groups of authors should be further strengthened in order to using resources more completely.

In terms of number of published journal articles the top academic institutions are Harvard University, Stanford University, National Cancer Institute, Johns Hopkins University, University Michigan and etc. Much institutional cooperation occurs in this field, and research groups are relatively concentrated. In terms of national distribution, the United States ranks first. China should focus highly on improving the quality of research articles. The core authors and major contributors in this field are Menon M, Li CF, Wang J, Madabhushi A, Patel VR,

Kim S, and etc. BMC Bioinformatics, PLOS One, Journal of Urology, Scientific Reports, and Bioinformatics are the top five journals in the research field of emerging technologies in cancer.

From hotspot analysis and keyword co-word network evolution diagram, the research field of emerging technologies in cancer is found to present diversified topics. Disease hotspots (e.g., cancer, breast cancer, and prostate cancer) reflect the major diseases that have been of concern to scholars in this field worldwide in this century [42,43]. Hotspots in technologies, such as categorization, robotics, machine learning, data mining, prediction, support vector machines, and feature selection, reflect that the research field of emerging technologies in cancer has been prominent since the beginning of this century.

5.2 Future trends

The rise of emerging information technologies has provided many opportunities and technical ways for overcoming the cancer problem [44]. Fully utilizing emerging information technologies to solve cancer-related issues is attracting increasing interest from scholars. Here, several possible directions of future research are suggested for scholars in e-health and medical informatics.

First, cloud-based medical decision support systems for cancer diagnosis, treatment, and personalized nursing can be developed. Currently, no clear and authoritative definition of cloud medical services is available. Most definitions refer to medical services on the basis of new technologies, such as cloud computing, IoT, and decision support system technology, by improving the efficiency of diagnosis and treatment and the efficiency of resource utilization [45]. By using cloud medical decision support care, hospitals can conduct early detections for inpatients and remote diagnosis or provide personalized rehabilitation advices for patients. All data can be shared during the hospitalization of patients. During their rehabilitation or after discharge, necessary medical instruments can be installed in their residence or community health service agencies, and the monitoring data can be transmitted to hospitals. Doctors can review the data and provide patients with real-time healthcare service by analyzing results from a cloud-based medical decision support platform [46]. Research on cloud medicine, which is a potential important direction in the future, is lacking.

Second, big data-driven deep learning can be applied in cancer research field. Deep learning is a new field in machine learning research. It simulates neural networks of the human brain for analytical learning and imitates the mechanism of the human brain to interpret data, such as images, sounds, and texts [47]. Its advantage is that big data and enhanced algorithms are used to improve the performance of learning algorithms. The results of hotspot analysis in the previous section show that deep learning is an important issue in this field. The rapid development of big data science and technology makes big data-driven deep learning a vital research direction in the future.

Third, broad data sets can be used to provide accurate survival prediction for cancer patients. Prediction of survival periods is generally difficult because of the lack of sufficient and comprehensive data. Nowadays, medical data and survey data after treatments can be acquired and combined with patients' electronic medical records [48, 49, 50]. In the past, emerging information technologies have been applied in diagnosis and treatment, but application in the prognosis after treatments, such as the prediction of survival, is rare. Big

data-driven prognosis, such as survival prediction for cancer patients, is also expected to be a major research direction in the future.

Fourth, emerging technologies can be applied in the nursing care of cancer patients in urban and rural places. Comprehensive and scientific care is an important guarantee for successful treatment. Application of emerging technologies, such as early warning systems and mobile wearable devices [51], in the nursing care of cancer patients can improve work efficiency and can help in scheduling configuration of nursing resources. This approach can also remarkably decrease nurses' errors in work. The rapid advancement of smart hospital construction makes smart nursing care interesting for scholars in health informatics and e-health management.

Fifth, Knowledge graph will be applied to represent the health knowledge of cancer data and knowledge reasoning based on knowledge graph will become more popular in the future [52]. Semantic searching based on knowledge graph of patients with cancer will make the knowledge discovery be more accurate.

Finally, strengthen international cooperation. The complexity and comprehensiveness of scientific research have led academic researchers to gradually abandon the traditional individual combat mode, and the trend of scientific research cooperation is inevitable and increasingly strengthened [53, 54]. Only by continuously promoting active cooperation among academic talents in various countries can we solve more academic problems and promote the development of scientific research. In the context of the continuous development of emerging information technologies, countries should work together to address the challenges of cancer to human health.

Acknowledgements

We would like to thank Ms Jingjing Li for her insightful comments on an earlier version of this manuscript. This research is partially supported in the collection, analysis and interpretation of data by the National Natural Science Foundation of China under Grant Nos. 71771077, 71771075, 71501054, and 71573071.

Author Contributions: Xiaoyu Wang and Dongxiao Gu conceived the research, Jingjing Guo and Xiaoyu Wang wrote the draft; Jingjing Guo collected the data; Jingjing Guo and Xuejie Yang conducted data analysis; Ying Yang and Keyu Zhu provided some guiding suggestions and revised the paper.

Competing interests: The authors declare that they have no competing interests.

REFERENCES

1. Qian B Z, Pollard J W. Macrophage diversity enhances tumor progression and metastasis. *Cell* 2010; 141(1): 39-51.
2. Akaza H, Kawahara N, Masui T, et al. Union for International Cancer Control International Session: Healthcare Economics: The significance of the UN summit non - communicable diseases political declaration in Asia. *Cancer Science* 2013; 104(6): 773-778.
3. De Vries E, De Poll - Franse V, Louwman W J, et al. Predictions of skin cancer incidence in the Netherlands up to 2015. *British Journal of Dermatology* 2005; 152(3): 481-488.
4. Sullivan R, Badwe R A, Rath G K, et al. Cancer research in India: national priorities,

- global results. *The Lancet Oncology* 2014; 15(6): e213-e222.
5. Lynch C. Big data: How do your data grow? *Nature* 2008; 455(7209): 28.
 6. Armbrust M, Fox A, Griffith R, et al. Above the clouds: A Berkeley view of cloud computing. Technical Report UCB/EECS-2009-28, *EECS Department University of California Berkeley* 2009.
 7. Atzori L, Iera A, Morabito G. The internet of things: A survey. *Computer Networks* 2010; 54(15): 2787-2805.
 8. Scheffler M, Hirt E. Wearable devices for telemedicine applications. *Journal of Telemedicine and Telecare* 2005; 11(1_suppl): 11-14.
 9. Rotolo D, Rafols I, Hopkins M M, et al. Strategic intelligence on emerging technologies: Scientometric overlay mapping. *Journal of the Association for Information Science and Technology* 2017; 68(1): 214-233.
 10. Restifo N P. A "big data" view of the tumor "immunome". *Immunity* 2013;39(4): 631-632.
 11. Parkin D M, Bray F, Ferlay J, et al. Global cancer statistics, 2002. *CA: a Cancer Journal for Clinicians* 2005;55(2): 74-108.
 12. Lin F P Y, Pokorny A, Teng C, et al. Computational prediction of multidisciplinary team decision-making for adjuvant breast cancer drug therapies: a machine learning approach. *BMC Cancer* 2016; 16(1): 929.
 13. Jurca G, Addam O, Aksac A, et al. Integrating text mining, data mining, and network analysis for identifying genetic breast cancer trends. *BMC Research Notes* 2016; 9(1): 236.
 14. Jeong Y K, Heo G E, Kang K Y, et al. Trajectory analysis of drug-research trends in pancreatic cancer on PubMed and ClinicalTrials.gov. *Journal of Informetrics* 2016;10(1): 273-285.
 15. Zhu Y, Song M, Yan E. Identifying liver cancer and its relations with diseases, drugs, and genes: a literature-based approach. *PLOS One* 2016; 11(5): e0156091.
 16. Aggarwal A, Lewison G, Rodin D, et al. Radiation therapy research: a global analysis 2001-2015. *International Journal of Radiation Oncology• Biology• Physics* 2018.
 17. Chiang J K, Lin C W, Wang C L, et al. Cancer studies based on secondary data analysis of the Taiwan's National Health Insurance Research Database: A computational text analysis and visualization study. *Medicine* 2017; 96(17).
 18. Chen X, Xie H, Wang F L, et al. A bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making* 2018;18(1): 14.
 19. Gu D, Li J, Li X, et al. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International Journal of Medical Informatics* 2017; 98: 22-32.
 20. Falagas M E, Pitsouni E I, Malietzis G A, et al. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB Journal* 2008;22(2): 338-342.
 21. Song M, Kim S Y, Zhang G, et al. Productivity and influence in bioinformatics: A bibliometric analysis using PubMed central. *Journal of the Association for Information Science and Technology* 2014;65(2): 352-371.
 22. Heo G E, Kang K Y, Song M, et al. Analyzing the field of bioinformatics with the

- multi-faceted topic modeling technique. *BMC Bioinformatics* 2017; 18(7): 251.
23. White H D, McCain K W. Bibliometrics. *Annual Review of Information Science and Technology* 1989;24: 119-186.
 24. Chen X, Xie H, Wang F L, et al. A bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making* 2018; 18(1): 14.
 25. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the Association for Information Science and Technology* 2006;57(3): 359-377.
 26. Gomez I, Teresa Fernández M, Sebastian J. Analysis of the structure of international scientific cooperation networks through bibliometric indicators. *Scientometrics* 1999, 44(3): 441-457.
 27. Zuckerman H A. Patterns of name ordering among authors of scientific papers: A study of social symbolism and its ambiguity. *American Journal of Sociology* 1968, 74(3): 276-291.
 28. Glänzel W, Czerwon H J. A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics* 1996, 37(2): 195-221.
 29. Cronin B. Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information science* 2001;27(1): 1-7.
 30. Zhuang Y, Jiang N, Wu Z, et al. Efficient and robust large medical image retrieval in mobile cloud computing environment. *Information Sciences* 2014; 263: 60-86.
 31. Smith D R. Citation analysis and impact factor trends of 5 core journals in occupational medicine, 1985-2006. *Archives of Environmental & Occupational Health* 2008; 63(3): 114-122.
 32. Merigó J M, Núñez A. Influential journals in health research: a bibliometric study. *Globalization and Health* 2016; 12(1): 46.
 33. Blanc X, Collet T H, Auer R, et al. Publication trends of shared decision making in 15 high impact medical journals: a full-text review with bibliometric analysis. *BMC Medical Informatics and Decision Making* 2014; 14(1): 71.
 34. Li H, An H, Wang Y, et al. Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. *Physica A: Statistical Mechanics and its Applications* 2016; 450: 657-669.
 35. Al-Hajj M, Wicha M S, Benito-Hernandez A, et al. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences*, 2003, 100(7): 3983-3988.
 36. Singletary S E, Greene F L, Sobin L H. Classification of isolated tumor cells. *Cancer* 2003;98(12): 2740-2741.
 37. Yang-Sheng X. Intelligent robots leading high-tech development. *China Science Daily* 2010.
 38. Mallapragada V G, Sarkar N, Podder T K. Robot-assisted real-time tumor manipulation for breast biopsy. *IEEE Transactions on Robotics* 2009, 25(2): 316-324.
 39. Drozdov I, Kidd M, Nadler B, et al. Predicting neuroendocrine tumor (carcinoid) neoplasia using gene expression profiling and supervised machine learning. *Cancer* 2009; 115(8): 1638-1650.

40. Behrmann J, Etmann C, Boskamp T, et al. Deep learning for tumor classification in imaging mass spectrometry. *Bioinformatics* 2017; 1: 9.
41. Kuo W J, Chang R F, Chen D R, et al. Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. *Breast Cancer Research and Treatment* 2001; 66(1): 51-57.
42. Moodley J, Singh V, Kagina B M, et al. A bibliometric analysis of cancer research in South Africa: study protocol. *BMJ Open* 2015; 5(2): e006913.
43. Wang Y, Zhai X, Liu C, et al. Trends of triple negative breast cancer research (2007–2015): A bibliometric study. *Medicine* 2016; 95(46).
44. Gu D, Liang C, Zhao H. A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. *Artificial Intelligence in Medicine*, 2017; 77: 31-47.
45. Kagadis G C, Kloukinas C, Moore K, et al. Cloud computing in medical imaging. *Medical Physics* 2013; 40(7).
46. Zhuang Y, Jiang N, Wu Z, et al. Efficient and robust large medical image retrieval in mobile cloud computing environment. *Information Sciences* 2014; 263: 60-86.
47. Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *nature* 1986; 323(6088): 533.
48. Wu Y, Zou J, Murillo P, et al. SU - F - J - 207: Non - Small Cell Lung Cancer Patient Survival Prediction with Quantitative Tumor Textures Analysis in Baseline CT. *Medical Physics* 2016; 43(6Part12): 3456-3456.
49. Kohli R, Tan S S L. Electronic health records: how can IS researchers contribute to transforming healthcare? *MIS Quarterly* 2016; 40(3): 553-573.
50. Gu D, Liang C, Zhao H. A case-based reasoning system based on weighted heterogeneous value distance metric for breast cancer diagnosis. *Artificial Intelligence in Medicine*, 2017;77: 31-47.
51. Zheng Y L, Ding X R, Poon C C Y, et al. Unobtrusive sensing and wearable devices for health informatics. *IEEE Transactions on Biomedical Engineering* 2014; 61(5): 1538-1554.
52. Bakal G, Talari P, Kakani E V, et al. Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations.. *Journal of Biomedical Informatics*, 2018.
53. Eguchi M, Koizumi S, Kumahara K. Dialogues on Cancer Survivorship: A New Model of International Cooperation Introduction. *Cancer*, 2013, 119(S11):2083-2085.
54. Gu D, Li J, Li X, et al. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International Journal of Medical Informatics*, 2017; 98:22-32.