Shiraz E-Med J. 2022 November; 23(11):e123803. https://doi.org/10.5812/semj-123803.

Published online 2022 November 5. **Research Article**

Identifying the Trends of Global Publications in Health Information Technology Using Text-mining Techniques

Meisam Dastani [1], Hamideh Ehtesham [2], Zohreh Javanmard [3], Azam Sabahi [2] and Fateme Bahador [3, *]

[1] Infectious Diseases Research Center, Gonabad University of Medical Sciences, Gonabad, Iran
[2] Department of Health Information Technology, Ferdows School of Paramedical and Health, Birjand University of Medical Sciences, Birjand, Iran
[3] Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran

[*] *Corresponding author*: Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran. Email: fatemebahador@gmail.com

Received 2022 March 05; **Revised** 2022 September 03; **Accepted** 2022 October 04.

Abstract

Background: Due to the increased publication of articles in various scientific fields, analyzing the published topics in specialized journals is important and necessary.
Objectives: This research has identified the published topics in global publications in the health information technology (HIT) field.
Methods: This study analyzed articles in the field of HIT using text-mining techniques. For this purpose, 162,994 documents were extracted from PubMed and Scopus databases from 2000 to 2019 using the appropriate search strategy. Text mining techniques and the Latent Dirichlet Allocation (LDA) topic modeling algorithm were used to identify the published topics. Python programming language has also been used to run text-mining algorithms.
Results: This study categorized the subject of HIT-related published articles into 16 topics, the most important of which were Telemedicine and telehealth, Adoption of HIT, Radiotherapy planning techniques, Medical image analysis, and Evidence-based medicine.
Conclusions: The results of the trends of subjects of HIT-related published articles represented the thematic extent and the interdisciplinary nature of this field. The publication of various topics in this scientific field has shown a growing trend in recent years.

Keywords: Health Information Technology, Text Mining, Scientific Publications, Trend, Health Information Management

1. Background

Health information technology (HIT) is one of the young and interdisciplinary scientific fields in medical sciences that has attracted researchers' attention from different fields. This scientific field includes various types of information and communication technologies (1) including a wide range of products, technologies, and services such as telehealth technology, cloud-based services, medical devices, remote monitoring devices, and sensors (2), for patient data collection, transfer, display, and storage (1). HIT also leads to the safe and confidential delivery of health information to patients, increases the efficiency of information exchange between physicians and patients, increases the quality of care, and reduces operating costs (3-5). Analyzing published texts in this field is necessary due to the rapid growth and variety of HIT topics and medical professionals' participation in various fields.

Identification of patterns and extraction of potential knowledge from textual data is important. Therefore, such a volume of scientific texts should be analyzed using automated tools and techniques. Generally, the extraction process of knowledge and practical and important information from various texts is called text mining (6, 7). The purpose of text mining research is to identify the topics of scientific texts, the evolution of these topics, using visualization tools for each current topic, and the relationship between them to help users identify topics (7).

Various researchers have analyzed the scientific documents and articles published in credible journals and citation databases by using text mining techniques, including topics related to insomnia and disorders (8), analysis and evolution of Mobile Health research (9), identifying published research trends in the anesthesia field (10), analysis of health information articles (11), medical library and information science (12), medical and biomedical education (13), and personality disorders (14).

In the field of HIT and medical informatics, Kim and

Delen identified the medical informatics research trend to understand the current state, the progressive paths, and the existing limitations of medical informatics. For this purpose, they analyzed 26,307 article titles and abstracts of 23 selected journals in the HIT field from 2002 to 2013 using text mining techniques (15). Shokouhian et al. also analyzed 6863 articles in the field of health records between 2009 and 2019 using text mining and bibliometric techniques. According to their results, the topics published in this scientific field have shown a growing trend in recent years, and the topics of electronic health records, health care, and healthcare systems have been of great importance (16). Feng et al. analyzed 62,340 articles in the medical informatics field using text-mining techniques and topic-modeling algorithms. Based on their results, topics such as new health technologies, drug safety, algorithm optimization, and medical information standardization are new and emerging in medical informatics (17).

According to the previous research review, the analysis and evaluation of scientific publications are crucial methods for researchers, organizations, and scientific policymakers at the international level. A growing trend in scientific publications has created a large amount of research data. The solution to this challenge and managing such a massive volume of publications is topic modeling and analyzing keywords in articles using automated text mining. Text mining is a statistical method that examines publications and documents to identify their contents and topics (18, 19).

Topic modeling is one of the most potent text mining techniques for exploring textual data exploration, embedded patterns in texts, and textual document connections (20), which reveals hidden topics in a set of documents (21). Researchers believe that topic modeling is a useful tool for information extraction from textual data and performs better in information retrieval than many traditional approaches (22).

Briefly, it can be acknowledged that topic modeling is a reliable and practical tool for scientometric studies that can respond to researchers' needs to explore a huge collection of scientific texts. It also introduces a structured and automated method to identify texts' content topics, a process beyond the power of traditional methods (23). The results of topic modeling can analyze the relationship of topics with each other and the evolution of topics over time (19). It discovers important and relevant topics and terms, track them over time, and provides an opportunity for analysts to better understand connections and changes (24).

2. Objectives

At the macro level, it also has many applications and effectiveness in strategic decisions and scientific and research policy-making in the health system (22). As a result of the importance of analyzing scientific texts, this study investigated the published topics in the HIT field and their trends over time through text-mining techniques and topic-modeling algorithms.

3. Methods

3.1. Sample and Data

The study was conducted using text mining and analytical methods. The population was all HIT-related publications in English from 2000 to 2019, which were indexed in PubMed and Scopus citation databases. The Mesh Browser was used to determine the search strategy, and it was consulted with HIT experts to determine the main keywords in the search strategy design. Only the three topics listed (Medical Informatics, Telemedicine, and Health Information Management) are relevant to the HIT domain in the mesh tree hierarchy, and no other items are left out. (Table 1).

A total of 208235 publications (Scopus: 67386 and Pubmed: 140849) related to HIT were extracted from the databases. Then after removing duplicates, 162,994 documents were selected; the titles and abstracts were analyzed using the text mining method taken from the designed framework by Zhang et al. (25), which includes three stages: text preprocessing, text mining operations, and interpretation of results.

3.2. Data Preprocessing

Titles and abstracts of retrieved publications data were merged in a text field to accomplish the text mining process. Then, the data was followed by preprocessing and data cleaning operations.

Preprocessing and data cleaning operations were applied to the text data to increase data quality and validity of patterns and extracted relationships, and it increases data quality, pattern validity, and extracted relationships and maintained only the required textual data (3).

In this research, the preprocessing included the following steps:

- eliminating unimportant characters, extra empty spaces, text formatting tags, and removing non-alphabetic characters such as deleting punctuation marks or numbers from the text).

- breaking text components into words (Tokenization).

- converting uppercase to lowercase letters for text unification

Table 1. The Search Strategy

Database	Search Strategies
PubMed	(((((("Medical Informatics/adverse effects"[Mesh] OR "Medical Informatics/classification"[Mesh] OR "Medical Informatics/economics"[Mesh] OR "Medical Informatics/education"[Mesh] OR "Medical Informatics/instrumentation"[Mesh] OR "Medical Informatics/methods"[Mesh] OR "Medical Informatics/nursing"[Mesh] OR "Medical Informatics/organization and administration"[Mesh] OR "Medical Informatics/pharmacology"[Mesh] OR "Medical Informatics/psychology"[Mesh] OR "Medical Informatics/rehabilitation"[Mesh] OR "Medical Informatics/standards"[Mesh] OR "Medical Informatics/statistics and numerical data"[Mesh] OR "Medical Informatics/supply and distribution"[Mesh] OR "Medical Informatics/therapeutic use"[Mesh] OR "Medical Informatics/therapy"[Mesh] OR "Medical Informatics/trends"[Mesh] OR "Medical Informatics/veterinary"[Mesh])) OR ("Telemedicine/classification"[Mesh] OR "Telemedicine/economics"[Mesh] OR "Telemedicine/ethics"[Mesh] OR "Telemedicine/history"[Mesh] OR "Telemedicine/instrumentation"[Mesh] OR "Telemedicine/legislation and jurisprudence"[Mesh] OR "Telemedicine/methods"[Mesh] OR "Telemedicine/organization and administration"[Mesh] OR "Telemedicine/standards"[Mesh] OR "Telemedicine/statistics and numerical data"[Mesh] OR "Telemedicine/trends"[Mesh])) OR "Health Information Management"[Mesh]) AND ("2000/01/01"[PDat] : "2020/12/31"[PDat]) AND English[lang]) Sort by: Title Filters: Publication date from 2000/01/01 to 2019/12/31; English
Scopus	TITLE-ABS-KEY (health AND information AND technology) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (PUBYEAR , 2002-2019) AND (LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "HEAL") OR LIMIT-TO (SUBJAREA , "NURS") OR LIMIT-TO (SUBJAREA , "PHAR") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "NEUR") OR LIMIT-TO (SUBJAREA , "MULT") OR LIMIT-TO (SUBJAREA , "DENT") OR LIMIT-TO (SUBJAREA , "VETE") OR LIMIT-TO (SUBJAREA , "Undefined"))

The stop words and other words that are not valuable for retrieving or analyzing documents, such as conjunctions and suffixes (and, the, of, for), were eliminated. As a result of evaluating the textual data, other words that were repeated and did not have a specific meaning were also added to the list of stop words.

The unification of synonyms was performed manually. Lemmatization was also used to create different forms of uniform words. Words or their primary dictionary forms were replaced with conjugated forms through lemmatization.

3.3. Topic Modeling

The next step was to perform the topic modeling algorithm and visualization with different text mining techniques, and the results of text mining operations were interpreted in the final step. In this research, topic modeling algorithms and patterns were used as machine learning techniques to discover the hidden topics in HIT global publications.

One of the topic modeling implementation methods is the Latent Dirichlet Allocation (LDA) (26), which is a very effective method for the identification of related semantic topics in scientific texts (27). Along with the LDA algorithm's strengths, it also has limitations in predicting the number of topics. Therefore, logarithmic (log) UMass Coherence was used to predict the number of topics and eliminate LDA limitations (28). Sixteen topics were selected for HIT publications using the UMass criteria. It should be noted that voluminous topics will lead to trivial and similar topics in the topic modeling algorithm (29, 30); thus, topic interpretation will be more difficult due to the dispersion of keywords in topics (31). Therefore, the number of topics should be determined logically, with thematic expert advice, and according to the extent of the topic matter (32). The LDA algorithm also determines the optimal number of topics, the frequency distribution of documents in the selected topics, and the keywords list related to each topic; however, it cannot do automatic labeling, and hence, topic labels are defined and specified non-automatically (33). Therefore, the topics resulting from the LDA algorithm are labeled and interpreted using each topic's most important words and articles.

In the present article, in the beginning, the number of topics was obtained for all data of HIT publications, and then, the signification of the number of topics and topic modeling were completed for publications related to each of the obtained topics.

Furthermore, after the words were identified, the publications related to each topic were labeled for that topic by consultation with a topical expert.

Python and Genism Library are used for the topic modeling algorithm, and Genism Library is used for open source topic modeling; it has a simple syntax, small size, and multipurpose use (34).

4. Results

The publication trend of HIT articles

Figure 1 shows the trend of global publications in the Scopus database of the HIT field over time.

The data in Figure 1 show that the trend of global HIT publications has shown an upward growth since 2000, and the highest number of articles in this field was published in 2015.

4.1. Identifying the Most Important Words Used in HIT Global Publications

Table 2 shows the ten most frequent words used in HIT publications. The results indicated that HIT publications mostly repeated the word "patients" (170,055 repetitions).

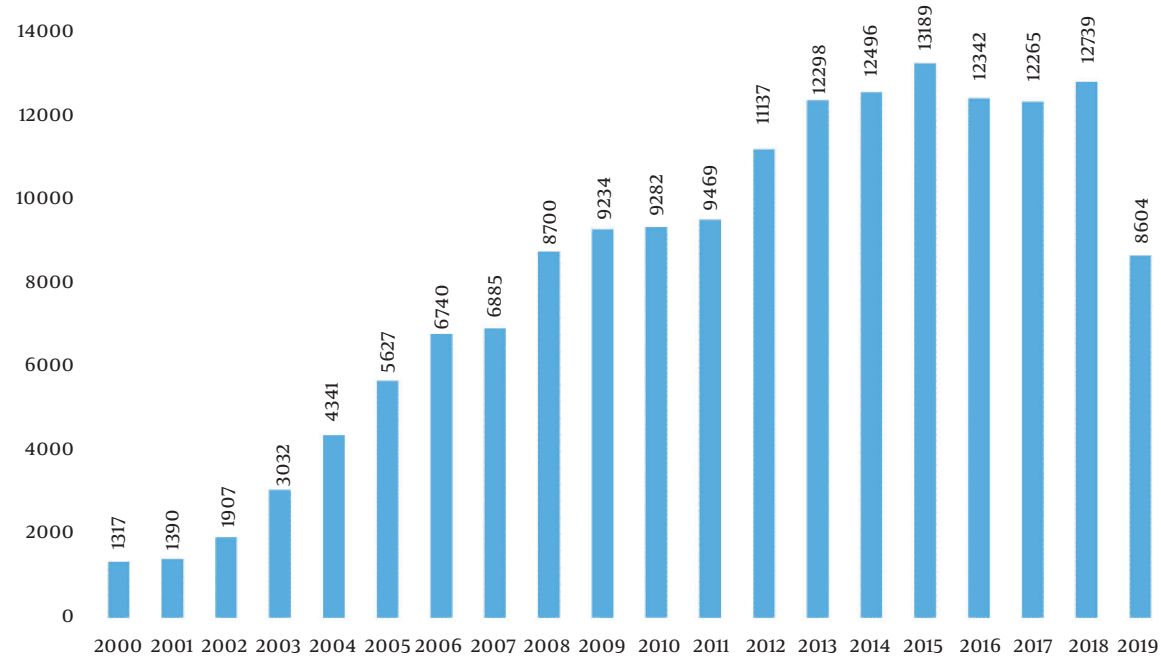


Figure 1. The trend of published health information technology articles

The Word cloud in Figure 2 shows the 50 most frequent words used in the HIT global publications.

Table 2. The Most Frequently Words Used in Health Information Technology Global Publications

Word	Frequency
Patients	170055
Health	148104
Based	122379
Information	115508
System	94482
Care	82092
Clinical	80006
Analysis	73382
Dose	72739
Treatment	68667

In Figure 2, words with larger fonts are more important and useful. Accordingly, Figure 2 shows that the words patient, health, image, and system have been used the most.

Figure 3 shows the most frequent words used in HIT publications during 2000 - 2004, 2005 - 2009, 2010 - 2014, and 2015 - 2019.

The data in Figure 3 indicate that the terms Patients, Dose, System in the years 2000 to 2004, Patients, Information, System 2005 - 2009, Patients, Health, and Information in 2010 - 2014 and 2015 - 2019 had the highest usage and repetition.

The results of implementing the topic modeling algorithm, 16 topics, along with the most important words of each topic, are shown in Table 3.

The data in Table 3 indicate that "telemedicine and telehealth," "HIT Adoption," and "Medical image analysis" have the highest numbers of publications releases, while "web-based technology in health" has the least number.

Figure 4 also shows 50 important words of each topic in the word cloud form. Accordingly, the data in Figure 4 show what the most important and widely used words were in each topic.

4.2. The Trend of Topics in HIT Publications

The heat map chart in Figure 5 shows the trend of various topics in publications in the HIT field from 2000 to 2019. Most publications were related to "telemedicine and telehealth" in 2018, with 3109 documents; also, the growth of this topic has been from 2007 onwards. "HIT Adoption" was another topic with a high publication rate among other topics, and the highest number of articles was published in 2015 with 1642 documents; the beginning of the growth of this issue was from 2007 onwards.

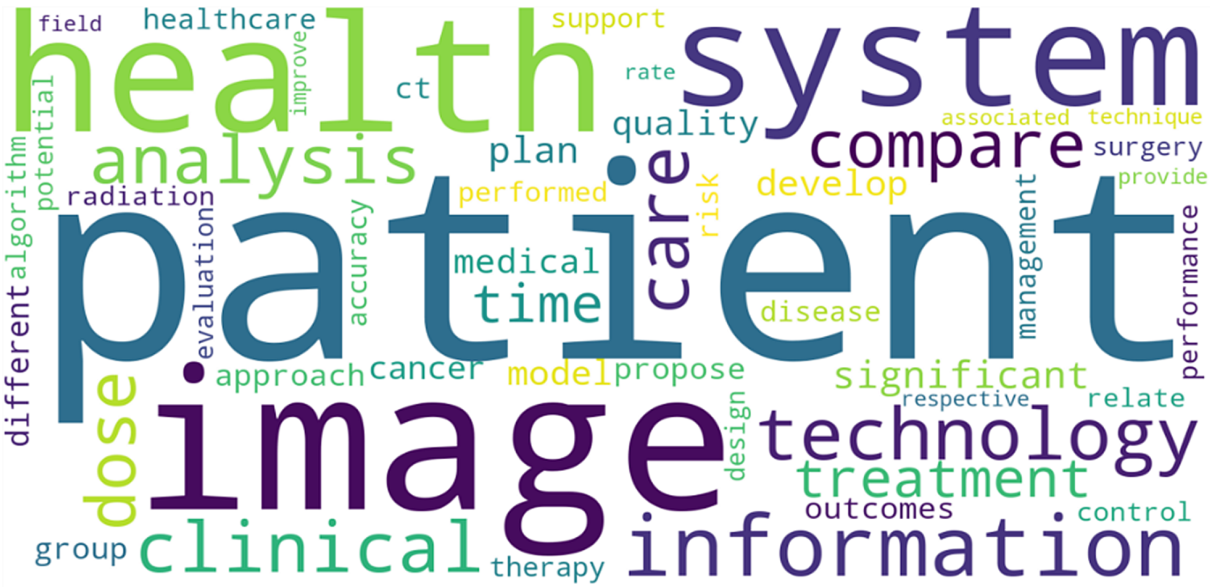


Figure 2. The most frequently used words in the HIT global publications

"Robotic Surgery," "Evidence-Based Medicine," and "Radiotherapy planning techniques" have been some of the most popular topics in recent years, and the other publication of topics has shown the same trend, and it has been growing, Although in the last five years, the topics "Models and algorithm performance" and "Electroenceph Alogram Analysis" there has been a declining trend in publishing.

5. Discussion

The results of this study showed that "patients," "information," "health," and "system" were the most important words used in the global publications of HIT, showing the general view of these publications in this context. Other studies have shown the most important words are used based on their repetition rates and weights (35-37).

The results of topic modeling for articles in the HIT field showed a clear picture of the published topics in this field from 2000 onwards. The results of this research indicate that the publications of this scientific field have shown a growing trend since 2000, with increasing interest in publishing various topics in this scientific field.

The topic modeling results identified 16 main topics for articles in the HIT field. The "telemedicine and telehealth" topic has been popular in recent years and has the most publications in the HIT field, with a publication growth from 2016 onwards. A study by Feng et al. showed that "telemedicine" was one of the emerging topics in medical informatics research, which received great interest in this research area (17). The growth of this topic has been reported in other studies (17, 38, 39). Considering that the goal of Telemedicine and telehealth is equal access to medicine regardless of geographical location (40), it is of interest to healthcare organizations worldwide. The continued advancement of internet-based audio and video communication technologies, combined with patients' desire for easier and more efficient care, has led to the growth of Telemedicine and telehealth. Over the past two decades, telemedicine functions, including teleconsultation, intensive care services, mental health monitoring, and chronic disease management, have grown significantly as a complement or alternative to physician visits (40).

HIT Adoption was another important topic in the HIT field and has grown significantly in the last decade. Kim and Delen also considered HIT Adoption as one of the main clusters of medical informatics research, including patient information electronic systems, electronic prescriptions, data sharing, and electronic reminders for health services (15). A study by Sadoughi and Ebrahimi showed that the HIT topic areas and health information systems were the raised topics among articles in the health information management and informatics fields, and the topic of electronic health records was the most important one in this field (41).

Radiotherapy planning techniques, medical imaging technology, robotic surgery, and virtual surgical planning are other topics identified in this study. These topics have shown a slow publication trend over time. The publica-

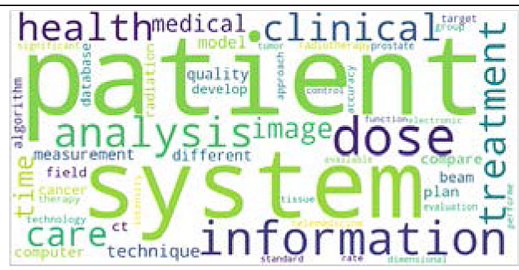
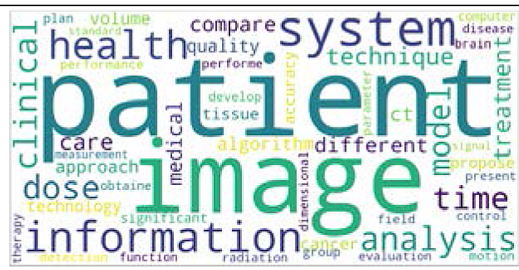
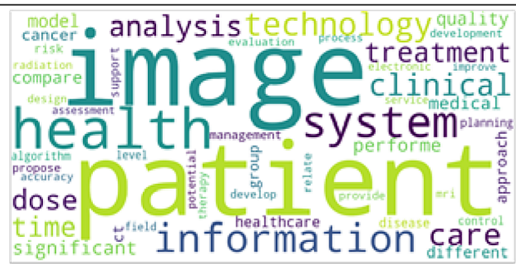
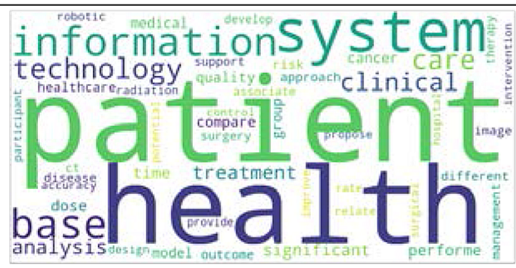
Time period	Words (Frequency)	Word cloud
2000-2004	patient (15120), system (10895), dose (9057), information (7163), analysis (7111), treatment (5890), health (5780), clinical (5583), care (4756), image (4274)	
2005-2009	patient (40682), image (32498), system (29747), information (20559), health (17374), analysis (15792), dose (15521), time (14768), clinical (14613), model (13767)	
2010-2014	image (57026), patient (54973), health (52375), system (50626), information (41201), technology (30710), care (28755), clinical (27447), analysis (24222), dose (23890)	
2015-2019	patient (77821), health (72575), system (49606), information (46585), base (45929), technology (41484), care (36609), image (33775), clinical (32363), analysis (28964)	

Figure 3. The most frequent terms used in global HIT published in five years

tion of these topics indicates the interdisciplinary nature of this scientific field. For example, 3D printing technology in surgical simulator construction for training and preoperative surgery practice has been researchers' focus (42, 43). Feng also considered these as the main and emerging topics in the medical informatics field (17).

This study identified other topics like medical image analysis, using a computer to diagnose cancer, models and algorithm performance, electroencephalogram analysis, and radiographic evaluation. This subject indicates a close connection between these fields and computer engineering. In these studies, researchers extract images, collect data in more detail, and categorize those with algorithms and neural networks to provide a more accurate diagnosis of diseases. Computer images are also used to diagnose tumors, cancer, bone diseases, and special dental treatments (44). Researchers also use machine learning algorithms, image recognition, semantic analysis, and other technologies and methods to conduct in-depth investigations about information systems, decision support,

Figure 4. The results of the modeling algorithm on the topics of HIT publications

Topic/ Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
HIT Adoption	243	266	304	349	388	395	496	511	914	1003	1140	1260	1371	1414	1536	1642	1478	1440	1473	1129
Robotic surgery	24	32	64	94	123	131	150	191	255	326	276	396	482	618	758	983	1086	1250	1258	745
Medical image analysis	40	41	87	192	473	729	969	1086	993	728	770	774	1012	998	940	834	611	598	806	484
Models and algorithm performance	31	23	54	138	326	479	552	483	537	379	337	269	296	332	288	248	192	132	121	64
Telemedicine and telehealth	269	274	315	340	492	461	563	566	1322	1800	1883	2163	2267	2499	2611	2820	2911	3010	3109	2784
Evidence based medicine	125	142	138	212	254	275	308	388	550	609	678	710	866	961	999	1105	1075	1152	1220	659
Virtual surgical planning	31	40	79	93	92	184	202	230	245	266	285	290	282	363	356	368	365	350	381	222
Medical and Biomedical database	66	95	219	398	432	418	473	544	512	391	539	416	517	548	525	566	447	394	479	255
Electroencephalogram Analysis	12	14	36	102	242	433	493	428	467	565	352	301	397	494	439	422	354	265	211	109
Disease risk factors	18	29	24	63	87	111	147	171	197	242	250	281	362	442	451	478	480	471	532	278
Medical imaging technology	20	18	43	159	334	597	748	651	734	739	600	430	679	765	720	662	526	425	359	176
Radiotherapy planning techniques	256	255	284	441	476	503	594	561	631	717	758	764	855	912	940	985	962	872	944	525
Radiographic evaluation	66	65	106	233	349	541	571	524	606	611	528	471	660	620	628	646	528	452	429	236
Using a computer to diagnose cancer	50	50	59	122	148	191	243	303	371	429	387	307	463	628	627	637	580	537	531	319
Healthcare monitoring system	42	25	54	57	81	131	164	185	240	309	313	412	384	401	378	401	387	417	363	324
Web based technology in health	24	21	41	39	44	48	67	63	126	120	186	225	244	303	300	392	360	500	523	295

Figure 5. The trend of topics in HIT publications

Table 3. Topics Result of the Health Information Technology Articles Topic Modeling Algorithm

Topic	Name	Top Keywords	The Articles Published, No. (%)
Topic 0	HIT Adoption	care, patient, health, system, information, medical, service, hospital, clinical, support	18752 (11.50)
Topic 1	Robotic surgery	patient, group, robotic, follow, month, year, surgery, outcome, cancer, time	9242 (5.67)
Topic 2	Medical image analysis	base, image, feature, segmentation, classification, algorithm, set, performance, detection, show	13165 (8.08)
Topic 3	Models and algorithm performance	model, base, simulation, estimate, function, present, distribution, parameter, source, estimation	5281 (3.24)
Topic 4	Telemedicine and telehealth	health, technology, information, development, new, need, user, provide, application, challenge	32459 (19.91)
Topic 5	Evidence-based medicine	clinical, review, report, cost, include, treatment, drug, identify, evidence, case	12426 (7.62)
Topic 6	Virtual surgical planning	registration, image, guide, surgical, procedure, system, surgery, technique, tracking, bone	4724 (2.90)
Topic 7	Medical and Biomedical database	database, information, tool, web, user, gene, protein, provide, analysis, search	8234 (5.05)
Topic 8	Electroencephalogram Analysis	brain, tissue, signal, time, change, subject, volume, cell, flow, heart	6136 (3.76)
Topic 9	Disease risk factors	risk, age, factor, high, analysis, increase, effect, level, population, child	5114 (3.14)
Topic 10	Medical imaging technology	image, high resolution, imaging, reconstruction, phase, quality, technique, noise, phantom	9385 (5.76)
Topic 11	Radiotherapy planning techniques	dose, treatment, plan, radiation, patient, volume, target, planning, therapy, radiotherapy	13235 (8.12)
Topic 12	Radiographic evaluation	measurement, measure, error, degree, accuracy, motion, compare, position, difference, surface	8870 (5.44)
Topic 13	Using a computer to diagnose cancer	cancer, tumor, image, lesion, breast, diagnosis, value, diagnostic, patient, detection	6982 (4.28)
Topic 14	healthcare monitoring system	the system, device, sensor, monitoring, time, base, monitor, network, application, real	5068 (3.11)
Topic 15	Web-based technology in health	intervention, group, participant, base, self, program, online, training, face, treatment	3921 (2.415)

medical imaging, biomedicine, and so on (45-48). Furthermore, algorithm performance optimization is increasingly used in medicine due to the increasing amount of health-related information (49-51). According to Feng (17), these are the main, interdisciplinary, and emerging topics in medical informatics. Kim and Delen also stated the application of algorithms for medical informatics research. Their study considered algorithms, neural networks, and computational technology for grouping/categorizing diseases and symptoms, the most important subset of topics in this cluster, so that anomalies are detectable by evaluating patterns (15). Evidence-based medicine is a topic in HIT that refers to systematic review articles, provides test reports and evidence-based studies for better treatment and healthcare, and is used to improve clinical outcomes (52). For medical informatics research, Kim and Delen stated knowledge representation as a topic, and this cluster includes medical text classification to report vaccination side effects, semantic disease classification, the use of medical tips in the EHR system to evaluate heart disease symptoms, a better understanding of disease diagnosis methods, management of clinical documents, report classification, medical text analysis knowledge discovery, and knowledge reuse (15). Feng has also identified this topic in medical informatics research and states that this issue is less important than other medical informatics topics (17). Similar to the evidence-based medicine subject, disease risk factors provide reports and analysis of disease trends, and this issue is declining in medical informatics (17).

The medical and biomedical database is also a topic that refers to medical and biomedical databases research such as gene and protein databases and Genome databases. The healthcare monitoring system also refers to technologies for the direct monitoring of patient's vital signs. Kim and Delen have also put these technologies in a biomedical (biomedical, medical engineering) cluster (15). Feng et al. have also addressed this topic in more detail and stated that some of these technologies in this topic, such as wearable healthcare technologies, are known as emerging. They have also stated the declining trend of some of the subset technologies of these topics, such as

blood pressure monitoring systems (17). The "web-based technology in health" topic seeks to achieve goals such as using the internet or online services for care quality improvement, patients' treatment, and using online resources to conduct medical research. This topic refers to internet-based research. Kim and Delen addressed this topic in their research (15). Feng also addressed more detailed topics, such as online health information and information integration in this area, and identified them as emerging topics in the medical informatics field (17).

Dastani et al. 2022 showed that the topic of Web-based treatment is one of the growing topics in the field of medical library and information (53); therefore, this topic is one of the interdisciplinary topics in the field of library and medical information and health information technology.

However, this topic has the lowest article publication rate, suggesting that internet studies generally include various fields and are the basis of the other technologies for HIT, such as Telemedicine, the internet, and its capabilities. Generally, previous studies emphasize that the medical informatics research fields are extensive and include a variety of topics, such as management information systems, mobile healthcare, telecare, social media, data mining, and machine learning (54, 55).

5.1. Conclusions

The medical and biomedical databases topic also refers to research into gene and protein databases and Genome databases. The healthcare monitoring system also refers to technologies for directly monitoring patients' vital signs. Kim and Delen have likewise put these technologies in a biomedical (biomedical, medical engineering) cluster (15). Feng et al. have also addressed this topic in more detail and stated that some of these technologies, such as wearable healthcare technologies, are known as emerging. They have also stated the declining trend of some of the subset technologies of these topics, such as blood pressure monitoring systems (17). The "web-based technology in health" topic seeks to achieve goals such as using the internet or online services for care quality improvement, patient treatment, and online resources to conduct medical research. This topic refers to internet-based research. Kim and Delen addressed this topic in their research (15). Feng also addressed more detailed topics, such as online health information and information integration in this area, and identified them as emerging topics in the medical informatics field (17).

Dastani et al. showed that the topic of web-based treatment is one of the growing topics in the field of medical library and information (53); therefore, this topic is one of the interdisciplinary topics in the field of library and medical information and health information technology.

Yet, this topic has the lowest article publication rate, suggesting that internet studies generally include a wide range of subjects and form the foundation for other technologies for HIT, such as Telemedicine, the internet, and its applications. Generally, previous studies emphasize that the medical informatics research fields are extensive and include a variety of topics, such as management information systems, mobile healthcare, telecare, social media, data mining, and machine learning (54, 55).

Acknowledgments

Researchers appreciate the Vice-Chancellor for Research and Technology of Birjand University of Medical Sciences for its financial support.

Footnotes

Authors' Contribution: Fateme Bahador and Meisam Dastani conceived of the presented idea. All authors discussed the search strategy. Fateme Bahador and Hamideh Ehtesham searched databases. Meisam Dastani verified the analytical methods and analyzed the search result. Hamideh Ehtesham and Zohreh Javanmard wrote the findings. Azam Sabahi and Fateme Bahador supervised the findings of this work. All authors discussed the results and contributed to the final manuscript. Meisam Dastani supervised the project.

Conflict of Interests: The authors declare that they have no conflicts of interest.

Ethical Approval: IR.BUMS.REC.1399.067.

Funding/Support: The present article is from a research project with code 5264, approved by the Research Council of Birjand University of Medical Sciences and implemented with the financial support of this university.

References

1. Sittig DF. *Electronic health records: Challenges in design and implementation*. CRC Press; 2013.
2. DeSalvo DD. *RE: Federal Health IT Strategic Plan 2015-2020*. Mc Kesson Corporation; 2015.
3. Devaraj S, Ow TT, Kohli R. Examining the impact of information technology and patient flow on healthcare performance: A Theory of Swift and Even Flow (TSEF) perspective. *J Oper Manag*. 2013;**31**(4):181–92. https://doi.org/10.1016/j.jom.2013.03.001.
4. Shekelle PG, Morton SC, Keeler EB. Costs and benefits of health information technology. *Evid Rep Technol Assess (Full Rep)*. 2006;(132):1–71. [PubMed ID: 17627328]. [PubMed Central ID: PMC4781594]. https://doi.org/10.23970/ahrqepcerta132.
5. Sherer S. Advocating for Action Design Research on IT Value Creation in Healthcare. *J Assoc Inf Syst*. 2014;**15**(12):860–78. https://doi.org/10.17705/1jais.00384.

6. Rodriguez-Esteban R, Bundschus M. Text mining patents for biomedical knowledge. *Drug Discov Today*. 2016;**21**(6):997–1002. [PubMed ID: 27179985]. https://doi.org/10.1016/j.drudis.2016.05.002.
7. Salloum SA, Al-Emran M, Monem AA, Shaalan K. Using Text Mining Techniques for Extracting Information from Research Articles. *Intelligent natural language processing: Trends and Applications*. **740**. Springer; 2018. p. 373–97. https://doi.org/10.1007/978-3-319-67056-0_18.
8. Lam C, Lai FC, Wang CH, Lai MH, Hsu N, Chung MH. Text Mining of Journal Articles for Sleep Disorder Terminologies. *PLoS One*. 2016;**11**(5). e0156031. [PubMed ID: 27203858]. [PubMed Central ID: PMC4874549]. https://doi.org/10.1371/journal.pone.0156031.
9. Ozaydin B, Zengul F, Oner N, Delen D. Text-mining analysis of mHealth research. *Mhealth*. 2017;**3**:53. [PubMed ID: 29430456]. [PubMed Central ID: PMC5803006]. https://doi.org/10.21037/mhealth.2017.12.02.
10. Rusanov A, Miotto R, Weng C. Trends in anesthesiology research: a machine learning approach to theme discovery and summarization. *JAMIA Open*. 2018;**1**(2):283–93. [PubMed ID: 30474079]. [PubMed Central ID: PMC6241511]. https://doi.org/10.1093/jamiaopen/ooy009.
11. Saheb T, Saheb M. Analyzing and Visualizing Knowledge Structures of Health Informatics from 1974 to 2018: A Bibliometric and Social Network Analysis. *Healthc Inform Res*. 2019;**25**(2):61–72. [PubMed ID: 31131140]. [PubMed Central ID: PMC6517625]. https://doi.org/10.4258/hir.2019.25.2.61.
12. Dastani M, Mousavi Chelak A, Ziaei S, Delghandi F. Topic Analysis of Published Articles in Medical librarianship and Information Science in Iran Using Text Mining Techniques. *Depict Health*. 2020;**11**(4):355–67. https://doi.org/10.34172/doh.2020.43.
13. Dastani M, Atarodi A, Erfanpoor S. A Study of Research Trends in Iranian Journals of Medical Education. *Fut Med Educ J*. 2021;**11**(1):3–12.
14. Boyack KW, Newman D, Duhon RJ, Klavans R, Patek M, Biberstine JR, et al. Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches. *PLoS One*. 2011;**6**(3). e18029. [PubMed ID: 21437291]. [PubMed Central ID: PMC3060097]. https://doi.org/10.1371/journal.pone.0018029.
15. Kim YM, Delen D. Medical informatics research trend analysis: A text mining approach. *Health Informatics J*. 2018;**24**(4):432–52. [PubMed ID: 30376768]. https://doi.org/10.1177/1460458216678443.
16. Shokouhian M, Asemi A, Shabani A, Cheshme-Sohrabi M. [Combined bibliometric and text-mining analysis of scientific productions in PubMed database in the field of electronic health records]. *Health Inf Manag*. 2019;**16**(4):190–6. Persian. https://doi.org/10.22122/him.v16i4.3953.
17. Feng J, Mu X, Wang W, Xu Y. A topic analysis method based on a three-dimensional strategic diagram. *J Inf Sci*. 2020;**47**(6):770–82. https://doi.org/10.1177/0165551520930907.
18. Abramson D, Lees M, Krzhizhanovskaya VV, Dongarra J, Sloot PM. Big Data Meets Computational Science, Preface for ICCS 2014. *Procedia Comput Sci*. 2014;**29**:1–7. https://doi.org/10.1016/j.procs.2014.04.002.
19. Wang C, Blei D, Heckerman D. Continuous time dynamic topic models. *arXiv preprint arXiv:1206.3298*. 2012.
20. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimed Tools Appl*. 2018;**78**(11):15169–211. https://doi.org/10.1007/s11042-018-6894-4.
21. Yau C, Porter A, Newman N, Suominen A. Clustering scientific documents with topic modeling. *Scientometrics*. 2014;**100**(3):767–86. https://doi.org/10.1007/s11192-014-1321-8.
22. Cheng X, Shuai C, Liu J, Wang J, Liu Y, Li W, et al. Topic modelling of ecology, environment and poverty nexus: An integrated framework. *Agric Ecosyst Environ*. 2018;**267**:1–14. https://doi.org/10.1016/j.agee.2018.07.022.
23. Lamba M, Madhusudhan M. Mapping of topics in DESIDOC Journal of Library and Information Technology, India: a study. *Scientometrics*. 2019;**120**(2):477–505. https://doi.org/10.1007/s11192-019-03137-5.
24. O'Callaghan D, Greene D, Carthy J, Cunningham P. An analysis of the coherence of descriptors in topic modeling. *Expert Syst Appl*. 2015;**42**(13):5645–57. https://doi.org/10.1016/j.eswa.2015.02.055.
25. Zhang Y, Chen M, Liu L. A review on text mining. *6th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE; 2015. p. 681–5.
26. Blei DM. Probabilistic topic models. *Commun ACM*. 2012;**55**(4):77–84. https://doi.org/10.1145/2133806.2133826.
27. Griffiths TL, Steyvers M. Finding scientific topics. *Proc Natl Acad Sci U S A*. 2004;**101 Suppl 1**:5228–35. [PubMed ID: 14872004]. [PubMed Central ID: PMC387300]. https://doi.org/10.1073/pnas.0307752101.
28. Röder M, Both A, Hinneburg A. Exploring the Space of Topic Coherence Measures. *Proceedings of the eighth ACM international conference on Web search and data mining*. ACM Digital Library; 2015. p. 399–408.
29. Greene D, O'Callaghan D, Cunningham P. How Many Topics? Stability Analysis for Topic Models. *Joint European conference on machine learning and knowledge discovery in databases*. Springer; 2014. p. 498–513.
30. Greene D, Cross JP. Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. *Polit Anal*. 2017;**25**(1):77–94. https://doi.org/10.1017/pan.2016.7.
31. Wieczorek O, Schubert D. The Symbolic Power of the Research Excellence Framework. Evidence from a case study on the individual and collective adaptation of British Sociologists. *SocArXiv Papers*. 2020. https://doi.org/10.31235/osf.io/wda3j.
32. Colavizza G, Costas R, Traag VA, van Eck NJ, van Leeuwen T, Waltman L. A scientometric overview of CORD-19. *PLoS One*. 2021;**16**(1). e0244839. [PubMed ID: 33411846]. [PubMed Central ID: PMC7790270]. https://doi.org/10.1371/journal.pone.0244839.
33. Danesh F, Dastani M, Ghorbani M. Retrospective and prospective approaches of coronavirus publications in the last half-century: a Latent Dirichlet allocation analysis. *Library Hi Tech*. 2021;**39**(3):855–72. https://doi.org/10.1108/lht-09-2020-0216.
34. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. *In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*. Citeseer; 2010.
35. Funk ME. Our words, our story: a textual analysis of articles published in the Bulletin of the Medical Library Association/Journal of the Medical Library Association from 1961 to 2010. *J Med Libr Assoc*. 2013;**101**(1):12–20. [PubMed ID: 23405042]. [PubMed Central ID: PMC3543134]. https://doi.org/10.3163/1536-5050.101.1.003.
36. Bhui T, Sahoo S. Trend of public library research in India: A bibliometric study. *Libr Philos Pract*. 2018.
37. Dastani M, Mousavi Chelak A, Ziaei S, Delghandi F. Identifying Emerging Trends in Scientific Texts Using TF-IDF Algorithm: A Case Study of Medical Librarianship and Information Articles. *Health Technol Assess Action*. 2021;**4**(2). https://doi.org/10.18502/htaa.v4i2.6231.
38. Fatehi F, Wootton R. Telemedicine, telehealth or e-health? A bibliometric analysis of the trends in the use of these terms. *J Telemed Telecare*. 2012;**18**(8):460–4. [PubMed ID: 23209265]. https://doi.org/10.1258/jtt.2012.GTH108.
39. Yang YT, Iqbal U, Ching JH, Ting JB, Chiu HT, Tamashiro H, et al. Trends in the growth of literature of telemedicine: A bibliometric analysis. *Comput Methods Programs Biomed*. 2015;**122**(3):471–9. [PubMed ID: 26415760]. https://doi.org/10.1016/j.cmpb.2015.09.008.
40. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemed J E Health*. 2007;**13**(5):573–90. [PubMed ID: 17999619]. https://doi.org/10.1089/tmj.2006.0073.
41. Sadoughi F, Ebrahimi K. [Trend Analysis of Health Information Management and Informatics in Web of Science Journals]. *Health Inf Manag*. 2014;**11**(5):581–92. Persian.
42. Cui H, Yu Y, Li X, Sun Z, Ruan J, Wu Z, et al. Direct 3D printing of a tough hydrogel incorporated with carbon nanotubes for bone regeneration. *J Mater Chem B*. 2019;**7**(45):7207–17. [PubMed ID: 31663588]. https://doi.org/10.1039/c9tb01494b.
43. Idram I, Bintara RD, Lai J, Essomba T, Lee P. Development of Mesh-Defect Removal Algorithm to Enhance the Fitting of 3D-Printed Parts

for Comminuted Bone Fractures. *J Med Biol Eng*. 2019;**39**(6):855–73. https://doi.org/10.1007/s40846-019-00477-8.
44. Keshri AK, Das BN, Mallick DK, Sinha RK. Parallel algorithm to analyze the brain signals: application on epileptic spikes. *J Med Syst*. 2011;**35**(1):93–104. [PubMed ID: 20703581]. https://doi.org/10.1007/s10916-009-9345-y.
45. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 2019;**380**(14):1347–58. [PubMed ID: 30943338]. https://doi.org/10.1056/NEJMra1814259.
46. Blasiak A, Khong J, Kee T. CURATE.AI: Optimizing Personalized Medicine with Artificial Intelligence. *SLAS Technol*. 2020;**25**(2):95–105. [PubMed ID: 31771394]. https://doi.org/10.1177/2472630319890316.
47. Yanase J, Triantaphyllou E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst Appl*. 2019;**138**:112821. https://doi.org/10.1016/j.eswa.2019.112821.
48. Zarinabad N, Meeus EM, Manias K, Foster K, Peet A. Automated Modular Magnetic Resonance Imaging Clinical Decision Support System (MIROR): An Application in Pediatric Cancer Diagnosis. *JMIR Med Inform*. 2018;**6**(2). e30. [PubMed ID: 29720361]. [PubMed Central ID: PMC5956158]. https://doi.org/10.2196/medinform.9171.
49. Medlock S, Wyatt JC. Health Behaviour Theory in Health Informatics: Support for Positive Change. *Stud Health Technol Inform*. 2019;**263**:146–58. [PubMed ID: 31411160]. https://doi.org/10.3233/SHTI190119.
50. Wyatt JC. The Need for Theory to Inform Clinical Information Systems and Professionalise the Health Informatics Discipline. *Stud Health Technol Inform*. 2019;**263**:1–8. [PubMed ID: 31411148]. https://doi.org/10.3233/SHTI190105.
51. Schaffer JL, Rasmussen PA, Faiman MR. The Emergence of Distance Health Technologies. *J Arthroplasty*. 2018;**33**(8):2345–51. [PubMed ID: 29887358]. https://doi.org/10.1016/j.arth.2018.04.017.
52. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Inform Assoc*. 2001;**8**(6):527–34. [PubMed ID: 11687560]. [PubMed Central ID: PMC130063]. https://doi.org/10.1136/jamia.2001.0080527.
53. Dastani M, Ziaei S, Delghandi F. Discovering Research Topics from Medical Librarianship and Information using Text Mining. *Int J Inf Sci Manag*. 2022;**20**(2).
54. Deng H, Wang J, Liu X, Liu B, Lei J. Evaluating the outcomes of medical informatics development as a discipline in China: A publication perspective. *Comput Methods Programs Biomed*. 2018;**164**:75–85. [PubMed ID: 30195433]. https://doi.org/10.1016/j.cmpb.2018.07.001.
55. Liu S, Liu J, Zheng T. Current Status and Trends in Health Informatics Research: A Bibliometric Analysis by Health Technology and Informatics. *Stud Health Technol Inform*. 2019;**264**:1960–1. [PubMed ID: 31438428]. https://doi.org/10.3233/SHTI190734.