Health Information Technology during the COVID-19 Epidemic: A Review via Text Mining

Meisam Dastani¹, Alireza Atarodi²*

¹ Social Determinants of Health Research Center, Gonabad University of Medical Sciences, Gonabad, Iran.
² Department of Knowledge and Information Science, Paramedical College and Social Development & Health Promotion Research Center, Gonabad University of Medical Sciences, Gonabad, Iran.

Abstract

Background: Due to the prevalence of the COVID-19 epidemic in all countries of the world, the need to apply health information technology is of great importance. Hence, the study has identified the role of health information technology during the period of COVID-19 epidemic.

Methods: The present research is a review study by employing text-mining techniques. Therefore, 941 published documents related to health information technology’s role during COVID-19 epidemic were extracted by keyword searching in the Web of Science database. In order to analyze the data and implement the text mining and topic modeling algorithms, Python programming language was applied.

Results: The results indicated that the highest number of publications related to the role of health information technology in the period of COVID-19 epidemic was respectively on the following topics: “Models and smart systems,” “Telemedicine,” “Health care,” “Health information technology,” “Evidence-based medicine,” “Big data and statistic analysis.”

Conclusion: Health information technology has been extensively used during COVID-19 epidemic. Therefore, different communities could apply these technologies, considering the conditions and facilities to manage COVID-19 epidemic better.

Keywords: Health information technology, COVID-19, HIT, Text Mining, Topic modeling.

DOI: 10.5210/ojphi.v14i1.11090

*Correspondence: Alireza Atarodi

Copyright ©2022 the author(s)

This is an Open Access article. Authors own copyright of their articles appearing in the Online Journal of Public Health Informatics. Readers may copy articles without permission of the copyright owner(s), as long as the author and OJPHI are acknowledged in the copy and the copy is used for educational, not-for-profit purposes.

Introduction

Coronaviruses are large groups of viruses causing mild to severe diseases from the common cold to Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). In December 2019, an emerging infectious outbreak was found in Wuhan, Hubei Province, China, which was caused by the novel coronavirus (2019-nCoV) [1]. The pandemic was spreading across the whole world [2]. There were human-to-human and healthcare worker transmissions; however, the source of the coronavirus disease [COVID-19] has not been found,
and the route of pandemic transmission has not been fully understood. It is also probable that this virus to continue its mutations. The 2019-nCoV has a long incubation period and strong infectivity; therefore, the prevention and control of COVID-19 pandemic have faced great challenges. Compared with the Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), COVID-19 has some new and different features; it has spread more rapidly due to increased globalization, a longer incubation period, and hidden symptoms [3].

Due to the present circumstances, the use of technologies can help resolve the current crisis and make it simpler to manage [4]. As a center for disease control and prevention, the CDC also considers the use of information technology to manage COVID-19 to be necessary [5].

Health information technology is one of the new and interdisciplinary scientific areas in medical sciences that has attracted researchers' attention from various fields. Health information technology includes different types of information and communication technologies to collect, transmit, display, and store patient data [6], that includes an extensive range of products, technologies, and services such as telehealth technology, cloud-based services, medical devices, remote monitoring devices, and sensors [7,8]. The use of information technology in health is called electronic health and has been extensively applied in the health care system for many years [9, 10].

The governments and health organizations can use smart approaches to overcome this epidemic due to the development and advancement of existing technology infrastructures; hence, the use of technologies to fight against this epidemic has increased during the COVID-19 crisis [11].

At present, in line with the challenging and global spread of COVID-19 (new coronavirus 2019), medical researchers are conducting many studies on the prevention and treatment of this disease, and the results of their studies are presented at conferences and are published in credible scientific journals [12]. Moreover, much research has been carried out in health information technology and COVID-19 [13, 14]. Therefore, the rapid growth and variety of topics discussed in health information technology on one hand and the participation of experts of other fields in the studies related to health information technology, on the other hand, the analysis of published topics related to the application of information technology and COVID-19 is of particular importance for medical professionals, researchers, and policymakers.

Due to the increasing number of scientific papers and the considerable volume of published articles, evaluating and reviewing the articles one by one and manually extracting information and knowledge from this huge volume of texts is cumbersome or impossible. However, identifying patterns and extracting potential knowledge in large volumes of textual data is an important issue in various scientific fields [15]. The way to review these scientific texts quickly is through topic modeling and keyword analysis of the articles using automated text mining. Topic modeling and text mining are statistical techniques that assess the publications and documents to identify their topic [16, 17].

Topic modeling is a type of statistical modeling that explores latent patterns in texts and discovers the connections in a set of textual documents using machine learning [18, 19]. This type of modeling also provides a method and framework for smart review and exploratory analysis to researchers. Scientists can apply this approach to review publications and documents to make a transparent and highly reliable analysis of a large volume of publications and documents in the shortest possible time [20].
Thus, the present study has evaluated the publication texts related to the use of information technology during the COVID-19 epidemic by applying text mining techniques and identifying the published topics.

**Methods**

This section consists of three steps: data collection, topic modeling, and topic analysis.

**- Data Collection**

The statistical population is all publications related to health information technology and COVID-19, which has been indexed in the citation database of WOSCC (Web of Science Core Collection). The WOSCC advanced search was then applied to retrieve related publications. Since the Web of Science is the most authoritative, extensively used, and the oldest citation database in the world [21], the validity and reliability of the retrieved data are valid. In the next step, the designed search strategy was searched in the advanced search of WOSCC on July 31, 2020. The search strategy applied in the present research was as follows:

\[
\text{TS}=\left(\text{(Telemedicine) OR (Tele-medicine) OR (Telehealth) OR (Tele-health) OR (Mobile applications) OR (Mobile Apps) OR (M-health) OR (mHealth) OR (Mobile health) OR (eHealth) OR (Geographic information systems) OR (Geographic information system) OR (GIS) OR (Global Positioning Systems) OR (Global Positioning System) OR (GPS) OR (Registries) OR (Registry) OR (machine learning) OR (deep learning) OR (artificial intelligence) OR (Medical Order Entry Systems) OR (CPOE) OR (Computerized Provider Order Entry) OR (Computerized Physician Order Entry) OR (Medication Alert Systems) OR (Decision Support Systems, Clinical) OR (Clinical Decision Support Systems) OR (CDSS) OR (Clinical Decision Support) OR (CDS) OR (information technology)}\right)
\]

\[
\text{AND}
\]

\[
\]

The result of the advanced search was the retrieval of 941 articles; then, the title, abstract, and keywords of these articles were extracted for text mining analysis.

**- Topic Modeling**

The extracted documents were then investigated by the text mining method and topic-modeling algorithm. As one of the most popular text mining methods, topic modeling is an efficient approach to analyzing many documents [22]; topic modeling is also applied in some review studies [20,23].

Topic modeling is a machine learning approach to discover patterns or topics within a set of documents. Latent Dirichlet Analysis (LDA) is one of the implementation methods in topic modeling [24, 25]. LDA is one of the best and extensively used algorithms and highly effective in identifying related semantic issues in scientific texts [26], and outperforms many other algorithms [27]. In contrast to its advantages, the LDA algorithm has a limitation in predicting the number of topics. In this study, the number of predicted topics and the LDA limit were
eliminated using the logarithmic (log) criterion (UMass Coherence) [28]. Moreover, in the present research, six topics were selected for articles related to health information technology and COVID-19 using UMass criteria. It is noteworthy that selecting an excessive number of topics will lead to a large number of small and considerably similar topics [29, 30]. A higher number of topics also leads to the unavailability of no additional topic data. Furthermore, due to the dispersion of keywords between topics, the interpretation of topics becomes more difficult [31] and the lower number of topics facilitates the interpretation of results [32]. The python programming language and Gensim library have been applied to implement the topic-modeling algorithm [33]. The Gensim library is an open-source topic-modeling tool that is compact and versatile, possesses a simple syntax, is easy to develop, and provides various libraries for working with texts [33]. Numerous studies have also applied Gensim to implement LDA [34-36].

**Topic Analysis**

The LDA algorithm determines the optimal number of topics, the frequency distribution of each document in the selected topics, and the list of keywords related to each topic. However, it is not capable of automatic labeling; hence, topic labels are defined and specified manually [25, 37]. Accordingly, the topics resulting from the implementation of the LDA algorithm were labeled and interpreted using the most important words and articles of each topic and consulting health information technology professionals.

**Results**

The results of implementing the LDA topic modeling algorithm are presented in Table 1, in which the six obtained topics are shown along with the most important words and related articles of each topic.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Label</th>
<th>Top Keywords</th>
<th>Top Relevant Articles</th>
</tr>
</thead>
</table>
| **Topic 0** | Health information technology and management | health, information, response, practice, visit, outbreak, experience, child, technology, lock-down | A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact[38]  
Cyber Security Responsibilization: An Evaluation of the Intervention Approaches Adopted by the Five Eyes Countries and China [39]  
Effects of the COVID-19 crisis on survey fieldwork: Experience and lessons from two major supplements to the US Panel Study of Income Dynamics[40] |
<p>| <strong>Topic 1</strong> | Models and smart systems | model, base, disease, system, health, case, | A modeling framework to assess the likely effectiveness of facemasks in combination with 'lock-down' in managing the COVID-19 pandemic[41] |</p>
<table>
<thead>
<tr>
<th>Topic</th>
<th>Domain</th>
<th>Keywords</th>
<th>Articles</th>
</tr>
</thead>
</table>
| 2     | Big data and Statistic analysis | diagnosis, test, medical, spread | Tracking the Covid zones through geo-fencing technique [42]  
|       |        |          | Multi-tiered screening and diagnosis strategy for COVID-19: a model for sustainable testing capacity in response to pandemic [43] |
| 3     | Health care | case, analysis, factor, infection, cluster, identify, risk, mortality, predict, death | Interdependence assessing for networked readiness index economic and social informative factors [44]  
|       |        |          | Main factors influencing recovery in MERS Co-V patients using machine learning [45]  
|       |        |          | Spatiotemporal Clustering of Middle East Respiratory Syndrome Coronavirus [MERS-CoV] Incidence in Saudi Arabia, 2012-2019 [46] |
| 4     | Telemedicine | care, health, mental, risk, service, disease, disorder, infection, face, current | Ensuring mental health care during the SARS-CoV-2 epidemic in France: A narrative review [47]  
|       |        |          | The Silver Lining to COVID-19: Avoiding Diabetic Ketoacidosis Admissions with Telehealth [48]  
|       |        |          | Cardiac patients and COVID-19: what the general practitioner should know [49] |
| 5     | Evidence-based medicine | trial, clinical, report, participant, treatment, risk, evidence, drug, therapy, infection | Navigating telemedicine for facial trauma during the COVID-19 pandemic [50]  
|       |        |          | Telemedicine in the Time of Coronavirus [51]  
|       |        |          | Telehealth transformation: COVID-19 and the rise of virtual care [52] |
|       |        |          | Convalescent plasma or hyperimmune immunoglobulin for people with COVID-19: a rapid review [53]  
|       |        |          | Effect of hydroxychloroquine on prevention of COVID-19 virus infection among healthcare professionals: a structured summary of a study protocol for a randomized controlled trial [54]  
|       |        |          | Impact of pantoprazole on absorption and disposition of hydroxychloroquine, a drug used in Corona Virus Disease-19 (Covid-19): A structured summary of a study protocol for a randomized controlled trial [55] |
Figure 1 also illustrates the ten most important words of each topic in the form of word clouds. In a word cloud, the words with larger fonts are more important and useful in the related topic.

Word clouds provide a unique way to summarize the content of text documents. The word's size in a word cloud is proportional to its importance and application in the whole text collection [56].
Discussion

The topic modeling acts as a text mining tool for processing, organizing, managing, and extracting knowledge, and is commonly applied to identify basic “topics” in texts [55] and provides a practical and useful representation of a very large collection of documents, publications, and the relationships between them [18].

The results obtained from topic modeling have identified six main topics for articles in health information technology. The titles of each topic in order of most publications of each topic include “Models and smart systems,” “Telemedicine,” “Health care,” “Health information technology and management,” “Evidence-based medicine,” “Big data and Statistic analysis.”

The topics obtained from the present investigation show the role and application of information technology during the COVID-19 epidemic. The topic of “models and smart systems” had the highest number of published documents, which is about the application of models, algorithms, and neural networks in patients with COVID-19.

In this topic, medical data and images are categorized in more detail to provide a more accurate diagnosis of the disease. Researchers also use machine learning models and algorithms, image recognition, semantic analysis, and other technologies and methods to conduct in-depth research on information systems, decision-making support, medical imaging, biomedicine, etc. [58-61]. Apart from these cases, the optimization of algorithm performance is used increasingly in medicine due to the increasing volume of health-related information [62-64]. Feng also stated that these topics were among the primary, interdisciplinary, and innovative topics in medical informatics [65]. Kim and Delen also considered the use of algorithms, neural networks, and computational technology for categorization/classification of diseases and symptoms as the most important issues identified in the topic of models, so that the possibility
of detecting anomalies by evaluating patterns has been shown [66]. Moreover, Amiri indicated that researchers were interested in carrying out studies in the field of smart systems and their awareness of the application and power of these related algorithms [13].

Due to the impact of employing artificial intelligence models in detecting many diseases [67-69], this issue's importance becomes more apparent and is an important reason for researchers to be interested in this topic.

“Telemedicine” is another topic with a considerably high number of publications. Feng et al. also showed “Telemedicine” as one of the innovative topics in medical informatics that has attracted the researchers' interests [65]. Other studies have also reported the growth of the number of publications on this topic [70-72]. Given that the purpose of telemedicine is to provide equal access to medical care regardless of geographical location [73], it is of interest to health organizations around the world. The continued advancement of Internet-based audio and video communication technologies along with patients' desire for easier and more efficient ways to receive health care has led to the significant expansion of telemedicine functions over the past two decades, including teleconsultation, intensive care services, mental health monitoring as well as chronic disease management, as a supplement or an alternative to visiting doctor's office [73]. Telemedicine was also a successful technology in epidemic diseases [74]. The present study also demonstrated the importance of telemedicine in COVID-19 disease.

“Health care” is another topic that has been identified in the present study. Since no specific treatment has been identified for COVID-19, the need for self-care and self-control to prevent the spread of the disease is of great importance. People and individuals in the affected communities must learn to protect themselves from the potential dangers and harms of the outbreak of this mysterious and unknown new virus. Therefore, the topics related to health information technology and health care are also important about COVID-19 disease.

Health care is also an important issue in health information technology, and special attention has been paid to which in health information technology in different countries (75).

“Health information technology and management” is another topic that has been identified in the present study and refers to the application of health information technology in COVID-19. Health information technology provides the facilities to the medical personnel for managing numerous activities such as prescriptions for the patients, creation of electronic health records, testing and analysis data, etc. (76).

The use of health information technology is also useful in the control and management of COVID-19 (77).

Sedoghi et al. have also indicated that health information technology and health information systems were among the topics discussed in the articles in the field of information management and health informatics; the most central topic in this issue is dedicated to electronic health records [78].

Kim and Delen have also identified "adoption of hit" as one of the main clusters of medical informatics research, which has included topics such as electronic systems for recording patients' information, electronic prescriptions, data sharing, and electronic reminder system for health services [66].
“Evidence-based medicine” was the next topic identified in the present study; this concept refers to the presentation of experimental reports and evidence-based studies for better treatment and health care. Evidence-based medicine is extensively promoted as a tool to enhance clinical outcomes, which refers to medical operations based on the best scientific evidence. The scientific literature is the main source of evidence for evidence-based medicine, although evidence-based literature should be completed by local evidence, the practice-based operation for individual and site-specific clinical decision-making [79]. For medical informatics research, Kim and Delen also discussed a topic named knowledge presentation which included a classification of medical texts in reporting vaccination side effects, a semantic classification of diseases, the use of medical notes provided in the EHR system for evaluating the symptoms of heart disease, better understanding of the methods of early diagnosis of diseases, management of clinical records, classification of medical report texts and analysis of texts, the discovery of knowledge and its reuse [66]. Feng has identified the topic of “evidence-based medicine” in medical informatics studies and has stated that it is less important than other topics on medical informatics [65].

“Big data and Statistic analysis” was the topic with the least number of publications in the present study suggesting that researchers apply existing data and its statistical analysis to diagnose, control, and make related predictions.

Undoubtedly, after the global outbreak of COVID-19, a great deal of information about this virus, such as the number of people infected and or killed by COVID-19, epidemiological reports, the structure and activity of the coronavirus, the way of transmission of the virus, most-at-risk population, infection clusters, and symptoms of infected people have been available on official websites, scientific databases, hospitals, and other resources. Researchers can analyze these data to extract different models of the virus for identifying unknown aspects of the virus [14].

Conclusion

The COVID-19 epidemic is currently showing a global pandemic trend, and our understanding of the new coronavirus is deepening. The practitioners of global health information technology should be proactive and use their professional skills to respond to the COVID-19 epidemic.

The results indicate that health information technology has been extensively used during the period of the COVID-19 epidemic. Therefore, different communities can use these technologies to manage the COVID-19 epidemic better, considering the circumstances and facilities. The present investigation also showed that using models and smart systems and telemedicine in COVID-19 disease were the most important topics published in articles related to health information technology.

Limitations

The present study has categorized scientific publications related to health information technology and COVID-19 using text mining and topic modeling techniques. The data were obtained by searching for keywords related to COVID-19 from the Web of Science database. For this purpose, the keywords related to health information technology have been selected based on the most important and popular applications of health information technology; however, some new or less widely used technologies in these keywords might not be searched.
Since this study has identified six main topics and applications of health information technology in COVID-19 automatically and by applying text mining techniques and according to published scientific texts; therefore, each topic may have several sub-topics that have not been mentioned in this study. Therefore, it is suggested that each of the main issues be specifically evaluated and the sub-topics related to them be identified and analyzed in future studies.

Given that the data of this study is related to the first eight months of the COVID-19 epidemic, there may be other applications of information technology associated with this disease at other times, and researchers need to do similar studies for these situations.

Due to the fact that the data of this study was extracted only from the Web of Science database, some scientific publications that were published in other databases may not have been included.

The data of this study have been reported without separation of countries and geographical regions; here, there it is needed to conduct some studies on the application of health information technology by countries and geographical areas.

Researchers can use a combination of text mining techniques and conduct some types of reviews (e.g., systematic review or narrative review) on scientific publications to provide further details of their results.

Acknowledgements: The researchers thank the Social Determinants of Health Research Center of Gonabad University for their support.

Conflicts of Interest: None declared.

Financial Disclosures: None.

References


6. Sittig DF. Electronic Health Records: Challenges in Design and Implementation: CRC Press; 2013. URL: https://books.google.com/books?hl=en&lr=&id=oH5cAgAAQBAJ&oi=fnd&pg=PP1&ots=Z5s4_AcztS&sig=TVny4CE0eYx2ZtBGxIiBq66Y5G8#v=onepage&q&f=false


https://doi.org/10.1007/s11042-018-6894-4


Health Information Technology during COVID-19 Epidemic: A Review via Text Mining

https://doi.org/10.1136/jamia.2001.0080527