

Research Article

Iranian COVID-19 Publications in LitCovid: Text Mining and Topic Modeling

Meisam Dastani ¹ and Farshid Danesh ²

¹Infectious Diseases Research Center, Gonabad University of Medical Sciences, Gonabad, Iran

²Information Management Department, Regional Information Center for Science and Technology (RICeST), Shiraz, Iran

Correspondence should be addressed to Farshid Danesh; farshiddanesh@ricest.ac.ir

Received 27 April 2021; Revised 21 July 2021; Accepted 27 August 2021; Published 16 September 2021

Academic Editor: Shah Nazir

Copyright © 2021 Meisam Dastani and Farshid Danesh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

COVID-19 is a threat to the lives of people all over the world. As a result of the new and unknown nature of COVID-19, much research has been conducted recently. In order to increase and enhance the growth rate of Iranian publications on COVID-19, this article aims to analyze these publications in LitCovid to identify the topical and content structure and topic modeling of scientific publications in the mentioned subject area. The present article is applied research performed by using an analytical approach as well as text mining techniques. The statistical population is all the publications of Iranian researchers in LitCovid. Latent Dirichlet Allocation (LDA) and Python were used to analyze the data and implement text mining and topic modeling algorithms. Data analysis shows that the percentage of Iranian publications in the eight topical groups in LitCovid is as follows: prevention (39.57%), treatment (18.99%), diagnosis (18.99%), forecasting (7.83%), case report (6.52%), mechanism (3.91%), transmission (3.62%), and general (0.58%). The results indicate that patient, pandemic, outbreak, case, Iranian, model, care, health, coronavirus, and disease are the most important words in the publications of Iranian researchers in LitCovid. Six topics for prevention; four topics for treatment and case report and forecasting; three topics for diagnosis, mechanism, and transmission in general have been obtained by implementing the topic modeling algorithm. Most of the Iranian publications in LitCovid are related to the topic “pandemic status,” with 22.47% in the prevention category, and the lowest number of publications is related to the topic “environment,” with 11.11% in the transmission category. The present study indicates a better understanding of essential and strategic issues of Iranian publications in LitCovid. The results reveal that many Iranian studies on COVID-19 were primarily on the issues related to prevention, management, and control. These findings provided a structured and research-based viewpoint of COVID-19 in Iran to guide researchers and policymakers.

1. Introduction

In December 2019, an epidemic with mild respiratory infections was first reported to the World Health Organization (WHO) in Wuhan, China’s largest metropolitan area in the Hubei Province. Since it was impossible to identify the causative agent, people who suffered from this disease were classified as “pneumonia of unknown etiology.” The CDC (Center for Disease Control and Prevention) in China organized an extensive research program to study the extent and prevalence of the disease. The etiology of this disease is currently attributed to a new virus named SARS-CoV-2, belonging to the coronavirus family and causing COVID-19 [1]. The Director-

General of the WHO, “Tedros Adhanom Ghebreyesus,” on February 11, 2020, announced that the disease is caused by this new coronavirus and named it “COVID-19.” According to studies and evidence, COVID-19 is highly contagious and has a high global prevalence rate. At the International Health Regulations (IHR) meeting on January 30, 2020, the outbreak of COVID-19 was identified as a global health threat, since by that time, it was reported that it had spread to 18 countries by human-to-human contact [2].

Because of the increase in the COVID-19 pandemic, the WHO stated on January 30, 2020, that the new coronavirus is the sixth most common public health emergency worldwide, threatening China and all countries [2]. The

disease is also being rapidly spread. By August 06, 2021, global statistics on the coronavirus have indicated that more than 200 countries from all continents were struggling with COVID-19. The Islamic Republic of Iran, located in Southwest Asia, reported the first case of COVID-19 on February 19, 2020, in Qom city [3]. According to the website of Johns Hopkins University Coronavirus Resource Center (<https://coronavirus.jhu.edu/map.html>), by August 16, 2021, the total COVID-19 cases in Iran were 4467015 (ranked 11th in the world) and total deaths were 98483 (ranked 13th in the world).

To identify the various aspects of COVID-19, the publication and dissemination of results of all scientific activities in COVID-19 have become of particular importance [4], and significant publications in this field have been indexed in international citation databases. The investigations indicated that such a volume of scientific publications had not been observed in any scientific field and in such a short time [5]. In addition to the high number of COVID-19 publications, they are also being increased. However, medical scientists are still faced with many ambiguities; hence, extensive research is being conducted on various aspects of COVID-19 in many countries. Scientists worldwide are conducting significant studies on the aspects and methods of prevention, treatment, and the development of effective vaccines and medications to fight and eradicate COVID-19, and the results of their research projects are published in peer-reviewed journals [6]. The increase in global publications on COVID-19 has been influenced by researchers' events and efforts to fully understand the pandemic circumstances [6].

The reports revealed that scientists and researchers in China, the United States, the United Kingdom, and Italy had published the highest number of publications on COVID-19 [7, 8]. Because of the high volume of reputable publications on COVID-19, many studies with a macro and exploratory approach for policy-making and strategic planning of COVID-19 research have been done [8]. Text mining is one of the methods applied to analyze scientific publications on COVID-19 and involves extraction of information and latent knowledge in texts, especially scientific texts, and converting tacit knowledge into explicit knowledge [9, 10]. The purpose of text mining is to extract knowledge from textual data and has many applications in processing and analyzing scientific publications [11]. Therefore, text mining and text extraction techniques are used to analyze scientific texts and discover latent knowledge and topics discussed in a set of these texts [12–15]. Topic modeling is a statistical method and an essential algorithm in text mining, applied to identify latent topics in a set of textual documents [16, 17]. The results of this algorithm are extensively used for research policy-making and strategic planning [18] and allow analysts and researchers to comprehend better the relationships and variations of the provided topics [19].

After the COVID-19 pandemic, the “Information Explosion” phenomenon is witnessed regarding scientific publications on COVID-19. A review of PubMed revealed that more than 150,000 records related to COVID-19 had been indexed in this medical science database so far. Three researchers named Qingyu Chen, Alexis Allot, and Zhiyong

Lu from the United States, who work at the National Center for Biotechnology Information (NCBI), introduced the LitCovid as a specialized COVID-19 research database, which was designed with the support of the US National Institutes of Health's Intramural Research Program (IRP) and its data are being updated daily (Figure 1-Appendix). More than 150,000 records are indexed on LitCovid; this number is increasing and PubMed is the primary data source for LitCovid (Figure 1-Appendix). In LitCovid, research on COVID-19 has been divided into eight categories: general, mechanism, transmission, diagnosis, treatment, prevention, case report, and forecasting [20, 21].

According to the world meter database (<https://www.worldometers.info/coronavirus/country/iran/>), the I.R. Iran, with a population of 85196159 people, total cases of 4467015, total deaths of 98483, and total recovered cases of 3757157, among Asian countries, ranks third in terms of total deaths after India and Indonesia and ranks third in terms of total cases and total recovered cases after India and Turkey. Iran ranks 11th in terms of total recovered and total cases and 13th in terms of total deaths compared to other countries. According to the COVID-19 vaccination tracker provided by the New York Times, Iran's data are as follows: 16.2M doses of the COVID-19 vaccines have been given to Iranians. 3.15 million people (3.8%) received two doses of the vaccine and were fully vaccinated (<https://www.nytimes.com/interactive/2021/world/Covid-vaccinations-tracker.html>). All the mentioned data regarding COVID-19 statistics and vaccination were extracted on August 16, 2021.

The number of total cases, deaths, and fully vaccinated population in proportion to the total population of Iran (2.8% fully vaccinated) has made this country one of the most important countries globally in terms of the COVID-19 pandemic. Therefore, it is necessary to conduct a study by text mining methods and topic modeling to determine the topics of Iranian publications on COVID-19. It is necessary to answer the following questions to achieve the main objective of this paper:

- (i) RQ1: What are the percentages of global and Iranian publications based on the eight-topic division of LitCovid?
- (ii) RQ2: What are the most important words applied based on the repetition rate in Iranian publications in LitCovid?
- (iii) RQ3: What are the most important words applied based on TF-IDF weighting in Iranian publications in LitCovid?
- (iv) RQ4: What are the main topics and subtopics of Iranian publications on COVID-19 in LitCovid based on the topic modeling algorithm?
- (v) RQ5: What are the percentages of Iranian publications on COVID-19 in the eight topics of LitCovid and sub-topics extract from text mining?

2. Related Works

The studies indicate that the COVID-19 pandemic has negatively affected various aspects of the mental and physical

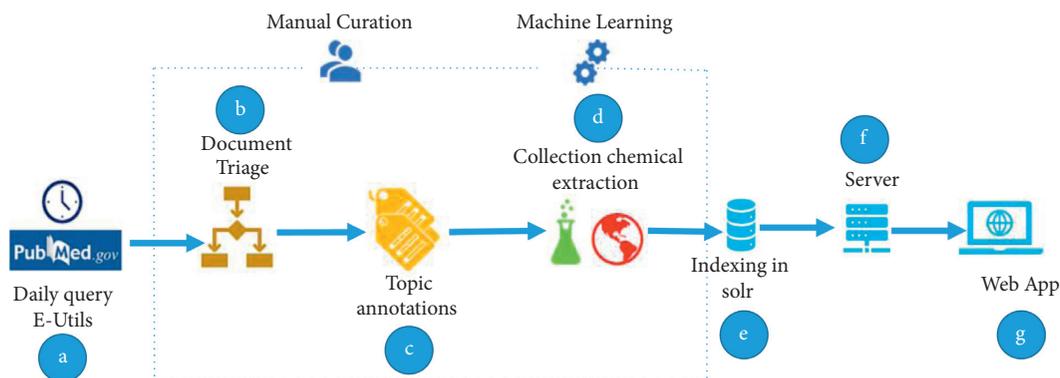


FIGURE 1: Daily workflow of LitCovid [20].

health of people in the community and has so far killed more than 4 million people worldwide. However, in recent waves of the disease, its adverse effects have diminished, but new disease variants (such as the Delta variant) are still emerging, and the whole world is still in a state of emergency [22]. Undoubtedly, the advancement of medical knowledge and related technologies will improve people's health [23]. The advancement of science is directly related to social needs and cultural elements. Scientists conduct research projects to solve social problems and anomalies.

Many studies have analyzed the publications of medical sciences using the text mining method and its topic modeling algorithms, including studying topic areas of biomedicine [12], personality disorders [24], clinical research [15], AIDS and economic evaluation [25], HIV prevention [26], health information [27], medical informatics [28], and health and cybersecurity [29]. The objective of this paper was the text mining of Iranian publications on COVID-19 in LitCovid. Hence, the text-mining-related works on COVID-19 have been reviewed in the following. It is noteworthy that some studies have analyzed and evaluated COVID-19 articles using scientometrics methods and indicators [6, 30–32]. The authors also applied the text mining method in the present paper.

Tran et al. have used text mining to evaluate the COVID-19 publications by April 23, 2020. In the present paper, countries were classified based on gross domestic product and World Bank report, and ten highly cited documents of each country were analyzed. The most important research topics were “Guidelines for emergency care and surgery,” “viral pathogenesis,” and “global responses in the COVID-19 pandemic” [33]. In line with previous research, Radanliev et al. also used WOSCC data to analyze articles related to mortality, immunity, and vaccine in the topic area of COVID-19 using data mining techniques. They identified the relationships between concepts and keywords in each scientific topic and collaborations between countries regarding scientific publications [34].

In a study conducted by Dong et al., the publishing features of COVID-19 in CORD-19 (COVID-19 Open Research Dataset) were analyzed by applying topic modeling. They identified important topics in COVID-19 publications and highlighted topics. The results of their study identified eight main topics for COVID-19 publications and

indicated that researchers should conduct further investigations on diagnostics, therapeutics, vaccines, viral genomics, and pathogenesis. In another research, the topics of global publications on COVID-19 in the CORD-19 were identified using text mining and topic modeling techniques; this study revealed that the COVID-19 publications have focused more on and have paid less attention to which topics [35]. In another study, the publications related to SARS, MERS, and COVID-19 were evaluated; in this article, the text mining of topics was performed separately, the scientific publications related to each of the viruses were modeled independently, and finally, the results were reviewed using an analytical-comparative approach [36]. In another paper, COVID-19 publications were extracted in the first six months of the pandemic using PubMed, and the topics of COVID-19 articles and their publication trends were extracted using topic modeling and text mining techniques [37].

A literature review indicated that text mining techniques were applied in text analysis and knowledge discovery and extraction in a large volume of texts and analysis of texts published in citation databases and identifying research trends in various scientific fields. According to the objectives and the research population, special text mining techniques have been used in each reviewed literature. Identifying the high-frequency keywords and the essential topics using special text mining algorithms was one of the essential techniques applied in scientific texts. The topic modeling algorithm, because of its high accuracy, was one of the most important algorithms employed in previous research to determine clusters and topics. The searches conducted in national and international citation databases and a review of foreign and domestic literature indicated that no similar research was observed in the topic area of Iranian publications on COVID-19 in LitCovid.

3. Materials and Methods

The present article deals with applied research conducted by an analytical approach and also the text mining technique. The statistical population was all Iranian publications of COVID-19, which were extracted from LitCovid in February 2021. LitCovid has been designed to track and rapidly access COVID-19 publications. Given that LitCovid is updated

daily and extracts its data from PubMed, it allows researchers to review the latest and most reputable scientific publications in the field of COVID-19 [20, 21]. All articles indexed in LitCovid are classified into eight categories: general, mechanism, transmission, diagnosis, treatment, prevention, case report, and forecasting. In the present research, text mining and topic modeling of titles, abstracts, and keywords of Iranian publications on COVID-19 have been performed separately in each of the eight categories of LitCovid.

3.1. Text Mining and Topic Modeling. Text mining analysis included three main steps: preprocessing, text mining, and post-processing, generally after collecting textual data [38]. Preprocessing texts included the selection of documents, extraction of the words used in texts, unification of texts using the manual analysis of articles' keywords and unification of synonymous words, and removing meaningless words and stop-words. In the present study, the most important words were identified by the Term Frequency–Inverse Document Frequency (TF-IDF) method, which indicated the importance of the word in a document or a set of documents. The main objective of this statistical method was to show the importance of the word in the text [28, 39]. In order to perform the preprocessing mentioned on the data and before extracting the most important words based on TF-IDF, the words used in the text of the articles were stemmed by Porter stemmer [40, 41]. In the next step, the frequency and weighting of the words, implementation of topic modeling, and visualization were done through different text mining techniques, and in the final step, the knowledge was extracted and interpreted. The LDA algorithm has been applied for topic modeling in all categories of LitCovid in the present paper.

In the present research and the mentioned preprocessing on the data and implementing the LDA topic modeling algorithm, the Bigrams and Trigrams have also been extracted from the texts to extract valuable data [42]. The Bigram was a set of two adjacent words, and Trigram was three adjacent words. For instance, Machine and Learning were two separate words, but their Bigram is extracted as Machine Learning [36].

Figure 2 indicates the idea behind the LDA logic. This method assumed that the set of documents consists of many topics, each containing many words (left Figure 2). Hence, it can be imagined that each document is created as follows: first, distribution of topics is selected (Figure 2), then a topic assignment (colored circles in the figure) is selected for each word, and finally, the desired word is picked from the relevant topic. It should be noted that the topics and topic assignments shown in Figure 2 were only presented as illustrative examples and were not derived by applying them to real data [17].

The basic idea of this process was that a combination of topics models each document, and each topic was a discrete probable dispersion that determined how the probability of each word appeared in the topic. These topic probabilities provided an accurate representation of the document. A

“document” was a “bag of words” with no structure beyond word and topic statistics.

Since the LDA algorithm did not determine the number of desirable topics, the logarithmic (log) UMass Coherence criterion has been applied in the present study to determine the number of desirable topics [43]. The UMass logarithm proposed different values of topics. In the present article, a different number of specific topics have been extracted for each of the topic categories of Iranian COVID-19 publications in the LitCovid. Guidance has been taken from medical sciences specialists to achieve the desired number of topics in each category. It is important to note that the number of topics should be selected proportionally because a large number of topics will lead to a significant quantity of small and considerably similar topics [38, 39]. Also, interpretation of topics becomes more challenging due to the dispersion of keywords between topics [40]. Afterward, the topics resulting from the implementation of the LDA algorithm were interpreted using the most important words and publications of each topic. Python and the libraries related to text mining, including Gensim, NLTK, and Spacy, were applied to implement text mining algorithms such as word frequency determination, TF-IDF, and topic modeling algorithms [41]. Python 3.9.5 and the libraries related to text mining, including Gensim, NLTK, and Spacy, were applied to implement text mining algorithms such as word frequency determination, TF-IDF, and topic modeling algorithms [44]. Python is an open-source, compact, and versatile programming language with a simple syntax. It is also straightforward to develop and provide various libraries for working with texts [44].

4. Results

In this section, the collected data were analyzed based on the methodology described in the previous section. The research questions are answered in the following.

4.1. The Percentages of COVID-19 Global and Iranian Publications Based on the Eight-topic Division in LitCovid. Figure 3 indicates the percentage of global and Iranian publications based on the eight-topic division in LitCovid. It should be mentioned that the percentage of Iranian publications in each of the eight categories is measured based on the total number of Iranian publications, and then the percentage of each topic is obtained. The same method has been applied to calculate the percentage of global publications. It was evident that the highest percentage of global and Iranian publications on COVID-19 was in the category of “Prevention”; 39.57% and 32.86% of the publications of Iranian and global researchers in LitCovid, respectively, were in the category of prevention. Moreover, the analysis of the data in Figure 3 revealed that the lowest percentage of Iranian research projects (0.58%) is in the “general” category, and the lowest percentage of global publications (1.94%) is related to the “Forecasting” category.

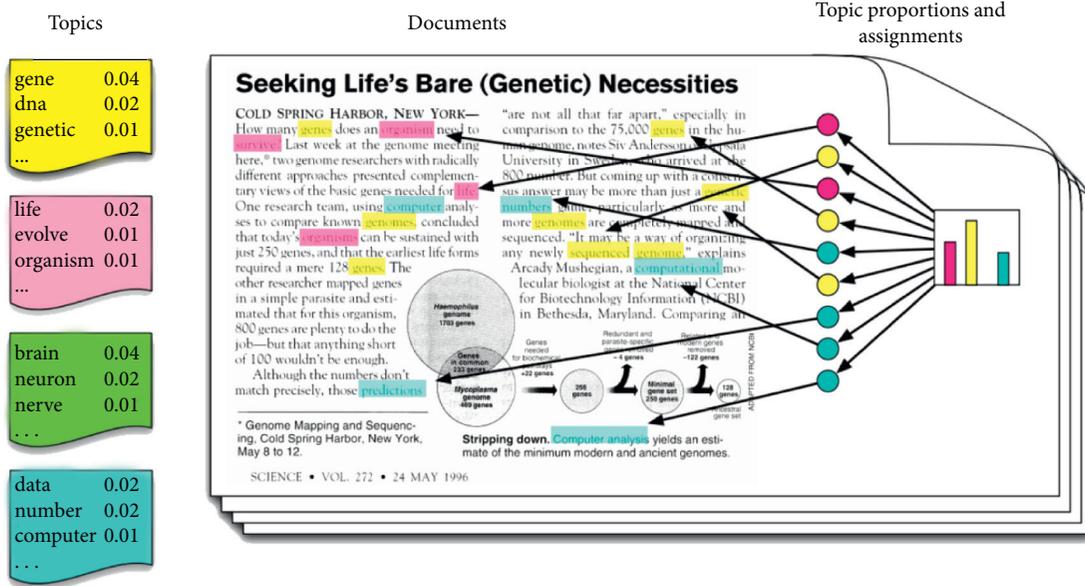


FIGURE 2: The idea behind the Latent Dirichlet Allocation (LDA) algorithm [17].

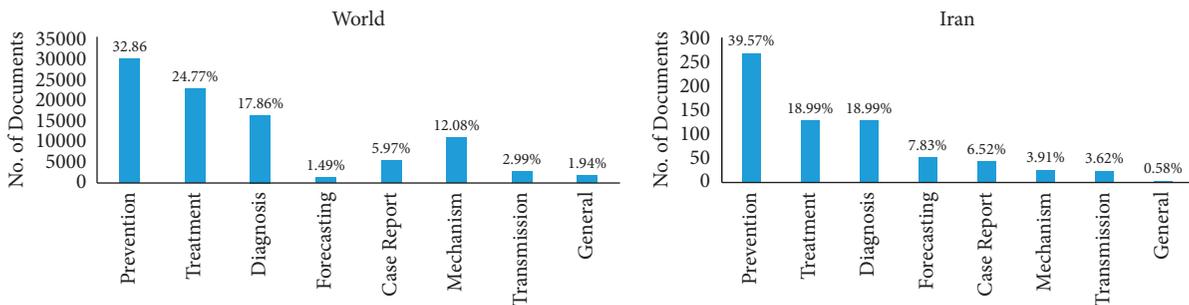


FIGURE 3: Percentage of COVID-19 global and Iranian publications based on the eight-topic division in LitCovid.

4.2. *The Most Important Words Applied Based on the Repetition rate in Iranian COVID-19 Publications in the LitCovid.* Figure 4 indicates the 20 words with the highest frequencies in Iranian publications. Figure 5 also illustrated the word cloud of 100 words with the highest frequency in Iranian publications on COVID-19. The data analysis showed that COVID, patient, and Iran had the highest frequencies in Iranian publications on COVID-19 in LitCovid with 2415, 1162, and 887, respectively (Figures 4 and 5).

4.3. *The Most Important Words Applied Based on TF-IDF Weighting in Iranian COVID-19 Publications in the LitCovid.* The data in Figure 6 included 20 important words based on TF-IDF weighting. Figure 7 also shows the word cloud of 50 top words based on TF-IDF weighting. The analysis of data revealed that “patient,” “pandemic,” and “outbreak” with the weights of 6.46, 4.57, and 4.29, respectively, are the most important words of Iranian publications in LitCovid based on TF-IDF weighting.

4.4. *The Main Topics and Subtopics of Iranian Publications on COVID-19 in the LitCovid Based on the Topic Modeling*

Algorithm. The data in Table 1 indicates the implementation of the topic modeling algorithm on Iranian publications on COVID-19 based on the eight topic category of the LitCovid. The tags subtopics have been identified based on the most important words and a review of the most significant articles of each subtopic. The number of documents for each of the main topics was also mentioned, e.g., in the topic “general” with four documents, only one topic with the same general title has been identified. The topic “mechanism” involved 18 documents, and three topics named “characteristics,” “clinical features,” and “genomic sequence” have been identified.

4.5. *The Percentages of Iranian COVID-19 Publications in the Eight Topics of LitCovid and the Subtopics Extracted from Text Mining.* Figure 8 indicates the percentages of Iranian publications in each of the eight categories of LitCovid and subtopics resulting from text mining of Iranian publications. The analysis of data shown in this figure revealed that in the topic area of “mechanism,” the publications percentages of the topics “characteristics,” “genomic sequence,” and “clinical features” are equal to 38.89%, 38.89%, and 22.22%,

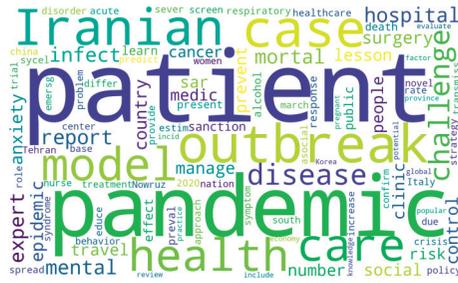


FIGURE 7: The word cloud of most important words based on the TF-IDF weighting of Iranian publications on COVID-19 in LitCovid.

TABLE 1: The topics extracted from the implementation of topic modeling algorithm on Iranian publications on Covid-19 in LitCovid.

LitCovid topic	Subtopics based topic modeling	Keywords
General	Topic1: General ¹	—
Mechanism	Topic1: Characteristics	Case, virus, people, smoking, potential, expression, disease, spread, find, level
	Topic2: Clinical features	Cell, infection, clinical, virus, case, acute, respiratory, severe, disease, lymphocyte
	Topic3: Genomic sequence	Sequence, respiratory, virus, phylogenetic, isolate, infection, mutation, strain, cause, acute
Transmission	Topic1: Environment	Rate, case, temperature, fatality, cluster, increase, spread, attack, correlation, size
	Topic2: Different areas	Transmission, disease, infection, rate, region, case, outbreak, report, measure, sample
	Topic3: Modes	Case, transmission, travel, cluster, age, early, global, report, individual, infection
Diagnosis	Topic1: Infection	Infection, case, disease, risk, chest, severe, clinical, symptom, evaluate, report
	Topic2: Risk factors	Mortality, clinical, outcome, risk, disease, diabetes, hospital, age, death, factor
	Topic3: Symptoms	Case, symptom, age, clinical, mortality, disease, positive, infection, test, confirm
Treatment	Topic1: Clinical features of mortality	Clinical, mortality, case, risk, hospital, factor, value, confirm, report, age
	Topic2: Clinical features of the disease	Disease, clinical, case, mortality, outcome, severe, infection, risk, age, rate
	Topic3: Drug	Control, trial, treatment, clinical, cell, drug, vitamin, receive, level, disease
	Topic4: Outcome	Trial, clinical, treatment, outcome, control, intervention, participant, arm, randomization, registration
Prevention	Topic1: Behaviors	Pandemic, health, disease, care, risk, control, outbreak, dental, infection, intervention
	Topic2: Management	Pandemic, hospital, health, stroke, knowledge, disease, management, base, care, infection
	Topic3: Policy	Pandemic, public, health, people, infection, policy, risk, disease, model, case
	Topic4: Control	Case, pandemic, disease, risk, control, spread, outbreak, infection, epidemic, people
	Topic5: Other diseases	Pandemic, outbreak, health, infection, report, care, trauma, increase, case, review
	Topic6: Pandemic status	Case, model, rate, death, spread, pandemic, measure, confirm, mortality, estimate
Case report	Topic1: Children	Case, clinical, infection, acute, test, child, report, syndrome, respiratory, neonate
	Topic2: New symptoms	Case, report, liver, manifestation, infection, symptom, clinical, cutaneous, disease, respiratory
	Topic3: Pregnant	Report, infection, case, old, pregnant woman, chest, treatment, refer, pandemic, hospitalize
	Topic4: Death	Report, case, disease, respiratory, acute, novel, rate, death, fetal, severe
Forecasting	Topic1: Estimate	Case, estimate, model, epidemic, death, predict, confirm, trend, outbreak, base
	Topic2: Modeling	Model, case, spread, rate, confirm, death, predict, pandemic, prediction, trend
	Topic3: Epidemic	Model, case, estimate, trend, base, policy, disease, infection, epidemic, prediction
	Topic4: Spread	Case, base, pandemic, spread, region, weather, mean, humidity, temperature, estimate

respectively. Moreover, the topic area of “transmission” includes three topics, “different areas,” “modes,” and “environment,” with publication percentages of 61.11%, 27.78%, and 11.11%, respectively. The topic area of “diagnosis” also involved three topics, “infection,” “risk factors,” and “symptoms,” with publication percentages of 29.55%, 32.95%, and 37.50%, respectively. The fourth topic area was “treatment,” which included four topics “clinical features of

mortality,” “clinical features of disease,” “outcome,” and “drug,” with publication percentages of 29.79%, 25.53%, 29.79%, and 14.89%, respectively. Moreover, the topic area of “prevention” had six topics: pandemic status (22.47%), management (20.26%), policy (16.74%), control (16.30%), behaviors (12.78%), and other diseases (11.45%). The “case report” was another area of eight topics, which involved four topics: new symptoms (31.58%), children (23.68%), death

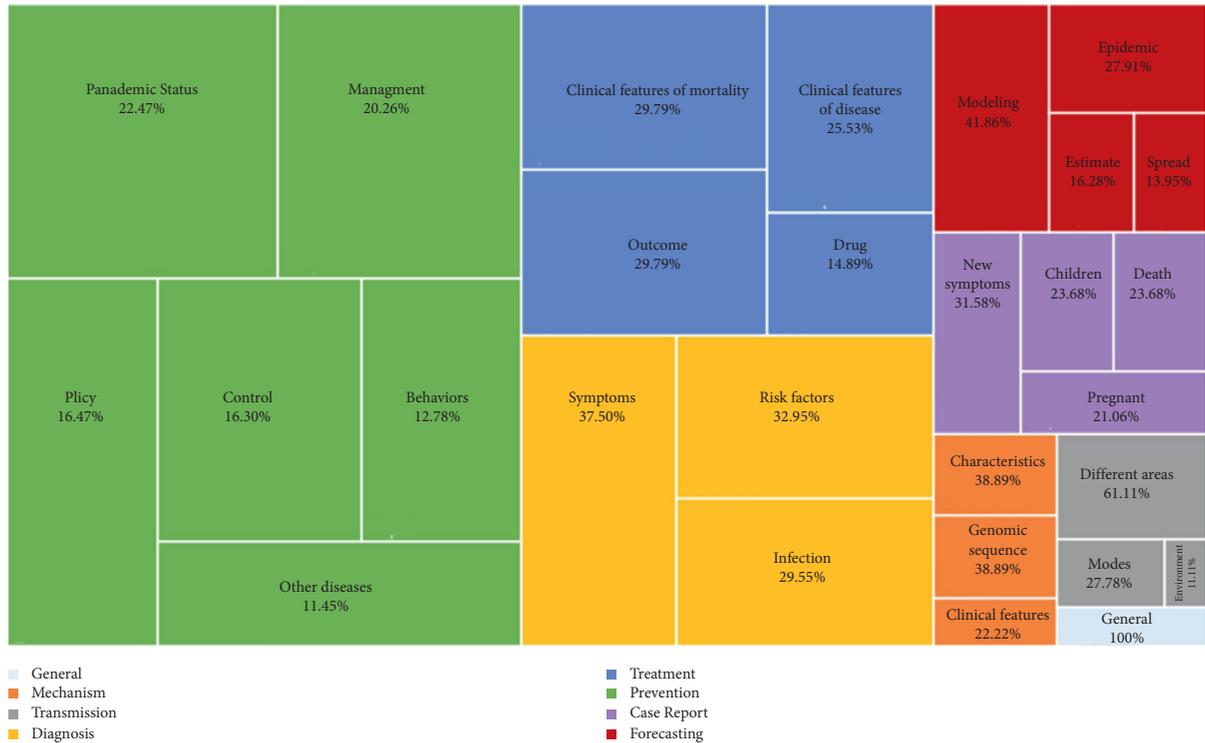


FIGURE 8: The percentages of Iranian COVID-19 publications in the eight-topic division of LitCovid and the topics extracted from text mining.

(23.68%), and pregnant (21.06%). The topic category of “forecasting” also included four topics: modeling (41.86%), epidemic (27.91%), estimate (16.28%), and spread (13.95%). The topic area “general” also had the lowest percentage of publications and addressed only one topic. The highest number of Iranian publications in LitCovid was related to the topic “pandemic status” in the “prevention” category, and the lowest number of publications was related to the topic “environment” in the “transmission” category.

5. Discussion

The management strategies for the health crisis that resulted from the COVID-19 pandemic are not the same in different countries, and unprecedented efforts have been made to fight the disease in all countries [45]. The global scientific community has responded to this crisis by applying all research resources and capacities to identify virus characteristics, mechanisms of its transmission, clinical aspects of the disease, and prevention and management strategies; so that the number of articles related to this crisis is being increased rapidly [46]. According to an increase in scientific publications, research centers have provided researchers different databases and software to immediately access and follow the trend of these publications; LitCovid is one of the most important and reputable databases [47].

LitCovid is an up-to-date set of all COVID-19 articles indexed in PubMed and categorizes COVID-19 publications into eight topic areas. The results of the present study indicated that the largest percentage of global and Iranian publications in LitCovid is related to the topic categories

“prevention,” “treatment,” and “diagnosis,” respectively. The lowest global and Iranian publications correspond to “forecasting” and “general,” respectively.

Scientific research also plays an essential role in controlling and preventing diseases, especially epidemics. The research findings and results significantly affect the identification of the virus variants, vaccine production, treatment protocols, preventive measures, and new medicines [48].

The most important words of Iranian publications on COVID-19 in LitCovid have been identified in the present study using text mining techniques. Moreover, the subtopics of these publications have been specified in each of the eight categories of LitCovid.

The results indicated that COVID, patient, Iran, case, disease, coronavirus, infect, pandemic, health, and hospital were the most important words, in terms of frequency, that have been used in Iranian publications in LitCovid. The words, patient, pandemic, outbreak, case, Iranian, model, care, health, coronavirus, and disease were the most significant words used in Iranian publications in LitCovid regarding TF-IDF weighting. In a study conducted by Doanvo et al., COVID, patient, pandemic, coronavirus, and case were the most frequently used words in COVID-19 publications in the COVID-19; this study is in line with the present investigation [35]. Moreover, in an article published by Cheng et al., data analysis indicated that the words, patient, case, number, infection, and study had the most repetitions in COVID-19 publications [36]. Dong et al. reported that infection, cell, protein, diseases, and patient were the most frequently applied words in COVID-19 publications [49]. In addition, the results of Hossain’s

research indicated that the keywords used in the COVID-19 publications showed the complexity and extent of this scientific field, which includes various disciplines such as virology, microbiology, infectious diseases, clinical medicine, public health, allied health sciences, social sciences, and other branches of knowledge [50].

Furthermore, in the present paper, the topics of each of the eight-topic categories of Iranian publications in LitCovid were identified using the LDA topic modeling algorithm. "Prevention," including status, management, policy, control, behaviors, and other diseases, involved some studies in the fields of control, response, and management strategies and had the largest percentage of Iranian publications in LitCovid.

Considering the nature of prevention studies, which is an important issue in the current pandemic situation, most studies and research projects are expected to be conducted in this regard. Colavizza concluded that the prevalence of SARS in 2003 had been associated with an increase in coronavirus and epidemic management [51]. Haghani et al. also believes that safety issues, such as physical safety and mental health of patients, constitute a large volume of the published knowledge in the field of COVID-19 [52]. In line with previous studies, Haleem et al. also believed that extensive research was needed to develop a vaccine against coronavirus infection. Besides, there was an urgent need to develop essential items and new therapies to cope with this disease [53].

Moreover, it was shown in some investigations that the countries with lower per capita income have more of a tendency toward the discovery and usage of disease prevention methods to prevent the spread of COVID-19. After "prevention," "treatment" was the next category with the highest percentage of Iranian publications in LitCovid. This topic area included the studies on treatment strategies, therapeutic procedures, and vaccine development, and involved the topics "clinical features of mortality," "outcome," "clinical features of patients," and "drug."

With the emergence of the COVID-19 pandemic, its rapid geographical spread (more than 200 countries), and the announcement of its prevalence as a public health emergency and international concern by the WHO, there is an urgent need for new diagnoses, vaccines, and treatments for this new and serious threat that has been affecting human life for more than one year. However, considering limited knowledge about COVID-19, there is no comprehensive prevention and treatment strategy, and all medical scientists around the world are working on new strategies to reduce virus infections, outbreaks, and the rate of mortality in high-risk environments. However, countries are implementing some measures to prevent the spread of the disease and create more time to produce appropriate vaccines and treatments. Moreover, pharmacotherapy options are consistently used in clinical trials to reduce the severity of infection, morbidity, and mortality [54]. Different drugs and treatments are currently being used to reduce the severity of the disease, and extensive research is underway for COVID-19 treatment [55]. Research on the development of COVID-19 vaccines and discovering new methods to improve

human safety against COVID-19 are the most important and cited topics under investigation since the discovery of the vaccine would enhance immunity and protection of people against the COVID-19 virus. They would also be effective in reducing the anxiety of the common populace.

"Diagnosis" was also ranked third of the highest percentage of Iranian publications on COVID-19, which involved studies on disease assessment through symptoms, test results, and radiological features. "Symptoms," "risk factors," and "infection" are the topics extracted from text mining in the present study.

In the diagnosis of COVID-19, different symptoms appear in the patient ranging from asymptomatic and mild symptomatic to severe symptomatic infections. Pneumonia is the most common serious manifestation of infection, characterized by fever, cough, shortness of breath, and bilateral pulmonary infiltrates on a chest X-ray [56]. Moreover, the results regarding the severity of the symptoms of this disease are scattered and incomplete. Besides, unexplained or unusual symptoms make COVID-19 challenging to diagnose and complicate appropriate treatment to patients. Lack of vaccines and effective treatment protocols increases the importance of early and definitive diagnosis of this disease [57].

The "forecasting" category includes some studies related to modeling and forecasting the trend of COVID-19 spread and is in the third rank of the highest percentage of publications. This category involves the topics "modeling," "epidemic," "estimate," and "region," respectively, from the highest to the lowest number of publications. Predicting the prevalence trend of this disease can help health authorities determine the characteristics of the virus transmission and develop appropriate strategies for preventing and controlling the disease. Researchers have applied conventional epidemic models, such as Susceptible-Exposed-Infective-Recovery (SEIR), or machine-learning models, such as logistic regression, to predict these trends in COVID-19 disease [58]. The "case report" category included studies in descriptions of specific patient cases, which is in the next rank in terms of the highest number of publications, and included the topics "new symptoms," "children," "other," and "pregnant."

The "children" play critical role in cases reported in the publications, and they are of particular importance in this disease; the course of this disease in children is mild, and infection caused by this virus in children has a better prognosis. Moreover, due to the mild clinical symptoms in children, many are not diagnosed in the early stages of COVID-19. Since this infection can be transmitted to others, children can play an essential role in transmitting this virus in the family, even from asymptomatic cases. In other words, children can lead to infection clusters in the home environment [59]. Given the predominance of gastrointestinal symptoms in infants and children, the spread of the virus through the feces continues for several weeks after diagnosis in children, and thus COVID-19 can spread in kindergartens or elementary schools [60]. Pregnant women are one of the most vulnerable groups and are prone to infectious diseases due to weakened immune systems, and their infection and

the risk of transmission to the fetus has become a significant concern [61].

The “transmission” category involved studies on characteristics and modes of COVID-19 transmissions, such as human-to-human type. This category included the topics “different areas,” “modes,” and “environment,” respectively, from the highest to the lowest number of publications. The “mechanism” category included studies on the underlying cause(s) of COVID-19 infections and the transmission and possible drug mechanism of action. In this category, “characteristics,” “genomic sequence,” and “clinical features” are the topics extracted from text mining. The result of the topic modeling indicated that “general” is one of the eight Categories. General information on the web and news on websites are the essential sub-topics. Similar results were obtained from the analysis of the scientific texts and publications on Coronavirus and COVID-19.

Danesh et al. identified the topics of coronavirus scientific publications in the last fifty years using text mining techniques and topic modeling algorithms. The results of their study provided eight topics for the global publications on coronavirus. These topics were “structure and proteomics,” “cell signaling and immune response,” “clinical presentation and detection,” “gene sequence and genomics,” “diagnosis tests,” “vaccine and immune response and outbreak,” “epidemiology and transmission,” and “gastrointestinal tissue” [62].

Continuing previous research, Colavizza indicated that the topics in the COVID-19 focus on specific topics such as coronaviruses (SARS, MERS, and COVID-19), public health and epidemics, molecular biology, influenza and the family of viruses, immunology and antiviruses, and methodology (test, diagnosis, and clinical trials) [51]. Another study in line with the previous investigation showed that the publications in the field of coronavirus have initially focused on public health and epidemic control; the chemical structure of the virus; and the studies related to treatment, vaccines, and clinical care [63].

Dehghanbanadaki believed that COVID-19 researchers focus on various aspects of this infection, such as pathogenesis, epidemiology, transmission, diagnosis, treatment, prevention, and complications [64]. Pal also showed that studies on COVID-19 had been published in an extensive range of disciplines such as medicine, biochemistry, molecular biology, immunology, microbiology, social sciences, nursing, pharmacology, neuroscience, environmental sciences, health care, and multidisciplinary [31]. In another article compiled by the topic modeling method, COVID-19 publications were categorized into eight topics: clinical characterization, pathogenesis research, therapeutics research, epidemiological study, virus transmission, vaccine research, virus diagnostics, and viral genomics [49]. In another investigation entitled “Text mining of global COVID-19 publications,” the results indicated that the focus in global COVID-19 publications is on clinical management, viral pathogenesis, and public health responses, and little attention has been paid to psychosocial problems or the impacts of COVID-19 on different vulnerable populations [33].

A study carried out by Älgå et al. during the first six months of the COVID-19 epidemic concluded that health

care response, clinical manifestations, and psychological impact were the most important published topics [37]. Moreover, Amiri et al. identified three topic clusters of health research, basic science research, and clinical research for COVID-19 publications in Scopus. In the health research cluster, the epidemiological aspects of the disease, public health, and prevention and control of infection were the main focus of the research. The main focus was on the virus’s virological, immunological, and genetic aspects in the basic science research cluster. Finally, clinical signs of disease, treatment methods, and diagnostic imaging were the primary focus in the clinical research cluster [8]. In another study conducted “discovering associations in COVID-19 related research paper,” the authors used the text mining method on COVID-19 data. They concluded that, the scientific publications related to various aspects of coronavirus were distinguished with different types of virus (e.g., RNA), clinical manifestations (e.g., pneumonia), consequences (e.g., quarantine), acquaintance (e.g., H7N9), and virus description (e.g., pathogen) [65].

6. Conclusion and Suggestions for Future

The findings of the present article have raised structural viewpoints of COVID-19 documents in Iran to guide and provide practical solutions to researchers, planners, and policymakers, and demonstrated different aspects of Iranian publications on COVID-19. A suggestion is that in an independent study, text mining method to be applied to all LitCovid publications, the topics of global publications in each of the eight topic areas to be obtained using topic modeling algorithms, and the results of text mining of global publications are to be compared with the present paper.

We suggested that researchers identify the most significant countries and institutions in the publication of each category in LitCovid. According to the limitations in LitCovid, it is also suggested that LitCovid designers design a mechanism wherein in the section of bibliographic data download, in addition to article titles, journal names, and PMID, other bibliographic data including the authors’ names, partner countries, organizational affiliation, publisher as well as the exact publishing date are downloadable and exportable to the statistical, text mining, and scientometrics software.

Appendix

Data Availability

The text mining data used to support the findings of this study are available from the corresponding author upon request.

Disclosure

The present article was extracted from a research project with code A-10-1263-5 and research ethics ID IR.G-MU.REC.1400.002, approved by the Infectious Disease

Research Center and implemented with the financial support of this research center.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

- [1] L. S. Wang, Y. R. Wang, D. W. Ye, and Q. Q. Liu, "A review of the 2019 Novel Coronavirus (COVID-19) based on current evidence," *International Journal of Antimicrobial Agents*, vol. 55, no. 6, 2020.
- [2] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, *Features, evaluation and treatment coronavirus (COVID-19)* StatPearls, treasure island, FL, USA, 2020.
- [3] A. Shamsi, M. J. Mansourzadeh, A. Ghazbani, K. Khalagi, N. Fahimfar, and A. Ostovar, "Contribution of Iran in COVID-19 studies: a bibliometrics analysis," *Journal of Diabetes and Metabolic Disorders*, vol. 19, no. 2, pp. 1–10, 2020.
- [4] P. Song and T. Karako, "COVID-19: real-time dissemination of scientific information to fight a public health emergency of international concern," *Bioscience trends*, vol. 14, no. 1, pp. 1–2, 2020.
- [5] J. A. Teixeira da Silva, P. Tsigaris, and M. Erfanmanesh, "Publishing volumes in major databases related to COVID-19," *Scientometrics*, vol. 126, no. 1, pp. 831–842, 2021.
- [6] F. Danesh, S. Ghavidel, and S. Ghavidel, "Coronavirus: scientometrics of 50 Years of global scientific productions," *Iranian Journal of Medical Microbiology*, vol. 14, no. 1, pp. 1–16, 2020.
- [7] B. Harsanto, "The first-three-month review of research on COVID-19: a scientometrics analysis," in *Proceedings of the 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1–6, IEEE, Cardiff, UK, June 2020.
- [8] M. Amiri, N. Taha, and P. Mahdizadeh, "Subjects clustering analysis and science mapping on COVID-19 researches in scopus database," *Journal Mil Med*, vol. 22, no. 6, pp. 663–669, 2020.
- [9] X. Liu, Yu Shi, F. Janssens, W. Glänzel, Y. Moreau, and B. De Moor, "Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 6, pp. 1105–1119, 2010.
- [10] Y. Zhang, A. L. Porter, and D. Chiavetta, "Scientometrics for tech mining: an introduction," *Scientometrics*, vol. 111, no. 3, pp. 1875–1878, 2017.
- [11] A. K. Choudhary, P. I. Oluikpe, J. A. Harding, and P. M. Carrillo, "The needs and benefits of text mining applications on post-project reviews," *Computers in Industry*, vol. 60, no. 9, pp. 728–740, 2009.
- [12] K. W. Boyack, D. Newman, R. J. Duhon et al., "Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches," *PLoS One*, vol. 6, no. 3, Article ID 18029, 2011.
- [13] J. Feng, X. Mu, W. Wang, and Y. Xu, "A topic analysis method based on a three-dimensional strategic diagram," *Journal of Information Science*, 2020.
- [14] B. Nie and S. Sun, "Using text mining techniques to identify research trends: a case study of design research," *Applied Sciences*, vol. 7, no. 4, p. 401, 2017.
- [15] J. R. Rogers, H. Mills, L. V. Grossman, A. Goldstein, and C. Weng, "Understanding the nature and scope of clinical research commentaries in PubMed," *Journal of the American Medical Informatics Association*, vol. 27, no. 3, pp. 449–456, 2020.
- [16] D. Abramson, M. Lees, V. V. Krzhizhanovskaya, J. Dongarra, and P. M. A. Sloot, "Big Data Meets Computational Science, Preface for ICCS2014," *Procedia Computer Science*, vol. 29, pp. 1–7, 2014.
- [17] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.
- [18] X. Cheng, C. Shuai, J. Liu et al., "Topic modelling of ecology, environment and poverty nexus: an integrated framework," *Agriculture, Ecosystems & Environment*, vol. 267, pp. 1–14, 2018.
- [19] H. Jelodar, Y. Wang, C. Yuan et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, 2019.
- [20] Q. Chen, A. Allot, and Z. Lu, "LitCovid: an open database of COVID-19 literature," *Nucleic Acids Research*, vol. 49, no. D1, pp. D1534–D1540, 2021.
- [21] Q. Chen, A. Allot, and Z. Lu, "Keep up with the latest coronavirus research," *Nature*, vol. 579, no. 7798, p. 193, 2020.
- [22] M. Coccia, "The impact of first and second wave of the COVID-19 pandemic in society: comparative analysis to support control measures to cope with negative effects of future infectious diseases," *Environmental Research*, vol. 197, Article ID 111099, 2021.
- [23] P. Howitt, A. Darzi, G.-Z. Yang et al., "Technologies for global health," *The Lancet*, vol. 380, no. 9840, pp. 507–535, 2012.
- [24] R. Sperandeo, G. Messina, D. Iennaco et al., "What does personality mean in the context of mental health? A topic modeling approach based on abstracts published in PubMed over the last 5 years," *Frontiers in Psychiatry*, vol. 10, p. 938, 2020.
- [25] X. Barn, L. H. Nguyen, H. C. Turner et al., "Economic evaluation studies in the field of HIV/AIDS: bibliometric analysis on research development and scopes (GAPRESEARCH)," *BMC Health Services Research*, vol. 19, no. 1, pp. 834–912, 2019.
- [26] B. X. Tran, H. T. Phan, Q. N. Nguyen et al., "Pre-exposure prophylaxis in HIV research: a latent dirichlet allocation analysis (gapresearch)," *AIDS Reviews*, vol. 22, no. 2, pp. 103–111, 2020.
- [27] T. Saheb and M. Saheb, "Analyzing and visualizing knowledge structures of health informatics from 1974 to 2018: a bibliometric and social network analysis," *Healthcare informatics research*, vol. 25, no. 2, p. 61, 2019.
- [28] Y.-M. Kim and D. Delen, "Medical informatics research trend analysis: a text mining approach," *Health Informatics Journal*, vol. 24, no. 4, pp. 432–452, 2018.
- [29] M. S. Jalali, S. Razak, W. Gordon, E. Perakslis, and S. Madnick, "Health care and cybersecurity: bibliometric analysis of the literature," *Journal of Medical Internet Research*, vol. 21, no. 2, Article ID 12644, 2019.
- [30] F. Danesh, S. Ghavidel, and V. Piranfar, "Coronavirus: discover the structure of global knowledge, hidden patterns and emerging events," *Journal of Advances in Medical and Biomedical Research*, vol. 28, no. 130, pp. 253–264, 2020.
- [31] J. K. Pal, "Visualizing the knowledge outburst in global research on COVID-19," *Scientometrics*, vol. 126, no. 5, pp. 4173–4193, 2021.

- [32] S. Belli, R. Mugnaini, J. Baltà, and E. Abadal, "Coronavirus mapping in scientific publications: when science advances rapidly and collectively, is access to this knowledge open to society?" *Scientometrics*, vol. 124, no. 3, pp. 2661–2685, 2020.
- [33] B. X. Tran, G. H. Ha, L. H. Nguyen et al., "Studies of novel coronavirus disease 19 (COVID-19) pandemic: a global analysis of literature," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, p. 4095, 2020.
- [34] P. Radanliev, D. De Roure, and R. Walton, "Data mining and analysis of scientific research data records on COVID-19 mortality, immunity, and vaccine development - in the first wave of the COVID-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research Reviews*, vol. 14, no. 5, pp. 1121–1132, 2020.
- [35] A. Doanvo, X. Qian, D. Ramjee, H. Piontkivska, A. Desai, and M. Majumder, "Machine learning maps research needs in COVID-19 literature," *Patterns*, vol. 1, no. 9, Article ID 100123, 2020.
- [36] X. Cheng, Q. Cao, and S. L. Stephen, "An overview of the literature on COVID-19, MERS, and SARS: using text mining and latent dirichlet allocation," *Journal of Information Science*, 2020.
- [37] A. Älgå, O. Eriksson, and M. Nordberg, "Analysis of scientific publications during the early phase of the COVID-19 pandemic: topic modeling study," *Journal of Medical Internet Research*, vol. 22, no. 11, Article ID 21559, 2020.
- [38] J. Žižka, F. Dařena, and A. Svoboda, *Text Mining with Machine Learning: Principles and Techniques*, CRC Press, Boca Raton, FL, USA, 2019.
- [39] W. B. Frakes, *Information Retrieval: Data Structures & Algorithms*, Prentice-Hall, Upper Saddle River, NJ, USA, 2004.
- [40] T. M. Abuhay, S. V. Kovalchuk, K. O. Bochenina et al., "Analysis of computational science papers from iccs 2001-2016 using topic modeling and graph theory," *Procedia Computer Science*, vol. 108, pp. 7–17, 2017.
- [41] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*, Vol. 39, Cambridge University Press, Cambridge, UK, 2008.
- [42] W. Wang, Y. Feng, and W. Dai, "Topic analysis of online reviews for two competitive products using latent Dirichlet allocation," *Electronic Commerce Research and Applications*, vol. 29, pp. 142–156, 2018.
- [43] A. Ebadi, P. Xi, S. Tremblay, B. Spencer, R. Pall, and A. Wong, "Understanding the temporal evolution of COVID-19 research through machine learning and natural language processing," *Scientometrics*, vol. 126, no. 1, pp. 725–739, 2021.
- [44] R. Radim and S. Petr, "Software framework for topic modelling with large corpora," in *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*, Valletta, Malta, May 2010.
- [45] J. Lau, "Coronavirus crisis inspiring unprecedented global research effort," *Higher Education*, vol. 25, 2020.
- [46] A. R. A. H. Hamid, "Social responsibility of medical journal: a concern for COVID-19 pandemic," *Medical Journal of Indonesia*, vol. 29, no. 1, pp. 1–3, 2020.
- [47] L. Škorić, A. Glasnović, and J. Petrak, "A publishing pandemic during the COVID-19 pandemic: how challenging can it become?" *Croatian Medical Journal*, vol. 61, no. 2, p. 79, 2020.
- [48] J. Lou, Si-J. Tian, Sheg-mei niu et al., "Coronavirus disease 2019: a bibliometric analysis and review," *European Review for Medical and Pharmacological Sciences*, vol. 24, no. 6, pp. 3411–3421, 2020.
- [49] M. Dong, X. Cao, M. Liang, L. Li, H. Liang, and G. Liu, "Understand research hotspots surrounding COVID-19 and other coronavirus infections using topic modeling," *medRxiv*, vol. 23, 2020.
- [50] M. M. Hossain, "Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping," *A Bibliometric Analysis and Knowledge Mapping*, vol. 9, p. 374, 2020.
- [51] G. Colavizza, R. Costas, V. A. Traag, N. J. Van Eck, T. van Leeuwen, and L. Waltman, "A scientometric overview of COVID-19," *PloS One*, vol. 16, no. 1, Article ID 0244839, 2021.
- [52] M. Haghani, M. C. J. Bliemer, F. Goerlandt, and J. Li, "The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: a scientometric analysis and scoping review," *Safety Science*, vol. 129, Article ID 104806, 2020.
- [53] A. Haleem, M. Javaid, R. Vaishya, and S. G. Deshmukh, "Areas of academic research with the impact of COVID-19," *The American Journal of Emergency Medicine*, vol. 38, no. 7, pp. 1524–1526, 2020.
- [54] V. K. Singh, A. Mishra, S. Singh et al., "Emerging prevention and treatment strategies to control COVID-19," *Pathogens*, vol. 9, no. 6, p. 501, 2020.
- [55] Y.-J. Han, Z.-G. Ren, X.-X. Li et al., "Advances and challenges in the prevention and treatment of COVID-19," *International Journal of Medical Sciences*, vol. 17, no. 12, pp. 1803–1810, 2020.
- [56] C. Jiehao, X. Jin, L. Daojiong et al., "A case series of children with 2019 novel coronavirus infection: clinical and epidemiological features," *Clinical Infectious Diseases*, vol. 71, no. 6, pp. 1547–1551, 2020.
- [57] J. Jeon, G. Baruah, S. Sarabadani, and A. Palanica, "Identification of risk factors and symptoms of COVID-19: analysis of biomedical literature and social media data," *Journal of Medical Internet Research*, vol. 22, no. 10, Article ID 20509, 2020.
- [58] J. Sun, X. Chen, Z. Zhang et al., "Forecasting the long-term trend of COVID-19 epidemic using a dynamic model," *Scientific Reports*, vol. 10, no. 1, Article ID 12112, 2020.
- [59] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges," *International Journal of Antimicrobial Agents*, vol. 55, no. 3, Article ID 105924, 2020.
- [60] P. Zimmermann and N. Curtis, "Coronavirus infections in children including COVID-19," *The Pediatric Infectious Disease Journal*, vol. 39, no. 5, pp. 355–368, 2020.
- [61] Y. Luo and K. Yin, "Management of pregnant women infected with COVID-19," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 513–514, 2020.
- [62] F. Danesh, M. Dastani, and M. Ghorbani, *Retrospective and Prospective Approaches of Coronavirus Publications in the Last half-century: A Latent Dirichlet Allocation Analysis*, Library Hi Tech, hongkong, China, 2021.
- [63] M. Haghani and M. C. J. Bliemer, "COVID-19 pandemic and the unprecedented mobilisation of scholarly efforts prompted by a health crisis: scientometric comparisons across SARS, MERS and 2019-nCoV literature," *Scientometrics*, vol. 125, no. 3, pp. 2695–2726, 2020.
- [64] H. Dehghanbanadaki, F. Seif, Y. Vahidi et al., "Bibliometric analysis of global scientific research on coronavirus (COVID-19)," *Medical Journal of the Islamic Republic of Iran*, vol. 34, p. 51, 2020.
- [65] I. Fister, K. Fister, and I. Fister, "Discovering associations in COVID-19 related research papers," 2020, <https://arxiv.org/abs/2004.03397>.