**DIGITAL
HEALTH**

Text classification technique for discovering country-based publications from international COVID-19 publications

Farshid Danesh[1] and Meisam Dastani[2] (iD)

Abstract

Objective: The significant increase in the number of COVID-19 publications, on the one hand, and the strategic importance of this subject area for research and treatment systems in the health field, on the other hand, reveals the need for text-mining research more than ever. The main objective of the present paper is to discover country-based publications from international COVID-19 publications with text classification techniques.

Methods: The present paper is applied research that has been performed using text-mining techniques such as clustering and text classification. The statistical population is all COVID-19 publications from PubMed Central® (PMC), extracted from November 2019 to June 2021. Latent Dirichlet allocation (LDA) was used for clustering, and support vector machine (SVM), scikit-learn library, and Python programming language were used for text classification. Text classification was applied to discover the consistency of Iranian and international topics.

Results: The findings showed that seven topics were extracted using the LDA algorithm for international and Iranian publications on COVID-19. Moreover, the COVID-19 publications show the largest share in the subject area of "Social and Technology in COVID-19" at the international (April 2021) and national (February 2021) levels with 50.61% and 39.44%, respectively. The highest rate of publications at international and national levels was in April 2021 and February 2021, respectively.

Conclusion: One of the most important results of this study was discovering a common trend and consistency of Iranian and international publications on COVID-19. Accordingly, in the topic category "Covid-19 Proteins: Vaccine and Antibody Response," Iranian publications have a common publishing and research trend with international ones.

Keywords

Publication, text mining, text classification, artificial intelligence, COVID-19, python, machine learning

Submission date: 3 November 2022; Acceptance date: 15 June 2023

Introduction

"Data science" has emerged as a powerful instrument for collecting, storing, managing, and analyzing big data.[1] Textual data are also an instance of big unstructured data; therefore, analyzing the huge volume of textual data, especially in scientific publications, is among researchers' critical challenges.[2] Text mining involves techniques used to manage textual data and benefits from artificial intelligence (AI) and natural language processing (NLP) to convert texts to analytic data using machine learning techniques. Moreover, text mining allows researchers to automatically analyze and categorize large collections of texts that are impossible to analyze manually.[3]

[1]Information Management Department, ISC, Shiraz, Iran
[2]Statistics and Information Technology Department, Gonabad University of Medical Science, Gonabad, Iran

Corresponding author:
Meisam Dastani, Gonabad University of Medical Sciences, Asiaie Roadside, Gonabad 9691793718, Iran.
Email: meisam.dastani@gmail.com

The fields of digital health and text mining have the potential to revolutionize the healthcare system as well. Digital health involves using technology to improve health outcomes,[4] while text mining involves extracting useful information from large amounts of unstructured data, such as medical records and scientific literature.[3] It is apparent that the combination of digital health and text mining has tremendous potential for accelerating medical research, improving patient outcomes, and reducing healthcare expenditures. The use of text mining, for instance, can provide insights into disease management by identifying patterns in large data sets that may not be apparent to human analysts. By utilizing these insights, more effective treatments and interventions can be developed.[5,6]

From late 2019, the challenging and global prevalence of the COVID-19 pandemic and the emergence of new variants such as *Delta* and *Omicron* have led scientists to conduct extensive research on various dimensions of COVID-19 to produce new vaccines, drugs, and treatments as well as publish the results as articles in peer-reviewed journals in the Web of Science Core Collection (WOSCC) and Scopus citation databases.[7] Research on the rapid spread of the pandemic and the significant economic, social, and psychological consequences of COVID-19 has been published as part of scientific publications to struggle and reduce the pandemic.[8] Researchers have conducted extensive research and published numerous papers on the pandemic's diagnosis, vaccine production, treatment, and general management.[9] The rapid increase in COVID-19 publications during the pandemic has played a key role in informing the scientific community and updating their professional knowledge.[10] It has shaped the current knowledge of the scientific community regarding COVID-19.[11] Accordingly, many scientific texts are indexed and published in citation databases.

Since reviewing all the texts is difficult and time-consuming, applying AI technology and text-mining techniques is essential in analyzing textual data related to COVID-19 publications. Therefore, text mining, one of AI's technologies, has provided various techniques and algorithms for analyzing scientific texts. Due to the high volume of COVID-19 publications and the necessity for analysis, the present paper applies text classification and clustering to classify and compare the publications at the international and national levels.

Text classification is used for the topic classification of scientific texts. Text classification refers to identifying the main topics of textual documents and engaging them in predefined collections.[12,13] However, this technique is primarily applied to a set of documents and is considered a supervised method used as a training model. The main objective of classification is based on known samples, based on which the unknown samples are automatically classified. This technique classifies text documents in a class with a predefined label.

Moreover, clustering refers to grouping similar text documents based on content without containing a predefined set.[12,13] The technique is unsupervised, and the input and output patterns are not predefined. Furthermore, there are no predefined labels and classes; instead, the amount of similarity is employed between different topics.

This study applies clustering and classification to text-mining scientific publications on COVID-19. As a first step, international publications on COVID-19 were clustered using text-mining techniques; then, Iranian publications on COVID-19 were classified according to the obtained clusters. This study uses two approaches unsupervised and supervised classification algorithms. Considering the large volume of international publications, it is undoubtedly not possible to categorize them into similar groups, so by using unsupervised text clustering algorithms, textual data are automatically and mechanically categorized into different groups, and then based on the obtained clusters, algorithms with classification supervision were used to obtain thematic clusters of Iranian publications based on thematic clusters of global publications. The use of clustering and classification methods in text mining is very effective and useful and can be applied more effectively to analyzing and using different data types. Using these methods makes it possible to access data more quickly and accurately and conduct various analyses.

Research questions

RQ1. What is the topic modeling of international COVID-19 publications?
RQ2. What is the classification of Iranian publications based on the topics of international COVID-19 publications?
RQ3. According to text classification, which of the subject categories of Iranian COVID-19 publications is consistent with international ones?

Literature review

Numerous studies use text-mining techniques and classification and clustering algorithms to analyze the published texts related to COVID-19, including the texts published by users on social networks and scientific publications indexed in databases. Liu et al. collected media reports on COVID-19 and examined media-based health communication patterns and the role of media in this crisis in China. WiseSearch was used to extract news articles. The data were analyzed by Python and the Python package Jieba. This article used latent Dirichlet allocation (LDA) for topic modeling. Based on coherence value, 20 topics were selected, and their keywords and themes were generated. Based on the topic visualization figure, the topics were divided into nine main categories. The most prominent

themes were prevention and control procedures, medical treatment and research, and global or local social and economic influences. Comparing the number of news articles for each day and the development of the outbreak, it was found that mass media news reports in China lag behind the development of COVID-19.[14] In line with previous research, Ghasiya and Okamura analyzed the database of more than 100,000 titles and news articles on COVID-19 using (topic modeling) and (classification and sentiment analysis). Top2vec for topic modeling and RoBERTa for sentiment analysis were used. Topic modeling results showed that education, the economy, the USA, and sports are some of the most widespread topics in the UK, India, Japan, and South Korea. In addition, sentiment analysis achieved a validation accuracy of 90%. The study showed that the UK has the highest percentage of negative sentiments and South Korea has the highest rate of positive sentiments.[15] Danesh et al., using text mining and LDA clustering algorithm, clustered 50 years of publications on coronavirus. The highest numbers of publications were respectively on the following topics: "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."[16]

Moreover, Dastani and Danesh analyzed topic clusters on Iranian COVID-19 publications in the LitCovid database using text mining and topic modeling techniques. 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. The results reveal that many Iranian studies on COVID-19 were primarily on prevention, management, and control issues.[17] In addition, Gupta et al. studied the publication trend of COVID-19 by applying NLP and LDA. They used the LDA algorithm to analyze 25 topics from publications indexed in PubMed by NLPs techniques. They demonstrate that research on "masks" and "personal protective equipment (PPE)" is skewed toward clinical applications, with a lack of population-based epidemiological studies.[18] Anderson analyzed and clustered COVID-19 publications using the singular value decomposition (SVD) and the expectation–maximization (EM) algorithms. The findings of this study identified 25 topic clusters for COVID-19 publications. The findings suggest that text clustering can detect hidden research themes in the published literature relating to COVID-19 and reduce the number of articles researchers must search to find relevant material.[3]

Koh et al. collected and analyzed 12,399 media news reports. LDA was applied to examine developments in old Korean media related to the launch of the COVID-19 vaccine. International trends in vaccination and social policies continue to be influenced by traditional media.

Furthermore, the media contributes to health-related communication in the public by reporting on and forming public discourse.[19] In another study, Silalahi, Arini, and Mulyani identify the most critical topics and identify research gaps using LDA during the COVID-19 pandemic. The topics most discussed were handling the case (lockdown and airport closure), conspiracy issues, fake news, daily cases reported, the importance of preventing COVID-19, the COVID-19 vaccination policy, economic recession, transportation systems, learning systems, and new procedures. Results indicate that reviewed articles discussed COVID-19 pandemic modeling or analysis. It can be acknowledged that LDA is an effective approach to topic modeling.[20]

Following the previous research, Amores, Blanco-Herrero, and Arcila-Calderón have analyzed the public conversation around COVID-19 on Twitter. Tweets were mostly in English and originated from the USA and the UK. Word frequency distribution, topic modeling, and sentiment analysis were used to analyze the data. Twitter was also a more useful platform for facilitating conversations. Topic modeling, sentiment analysis, and word frequency reveal that Twitter users also discussed ways to deal with the pandemic.[21] Also in another paper, Dehghani and Ebrahimi modeled Persian language publications about COVID-19. They extracted 815 Persian articles from Magiran and used the LDA with ParsBERT for modeling. The results indicate 10 main topics. Medical articles had the highest number of clusters, while engineering articles and religious publications had the lowest number. The topics found in the created clusters have structural relationships. Clusters such as these demonstrate how COVID-19 affects society on all levels, from individuals to families to communities. The results in the field of humanities indicated that most disorder is related to education and learning. In this study, Data collected from the Persian database – Magiran – were used to model scientific communication in COVID-19.[22] In line with previous research, the article by Chin et al. examined topics related to COVID-19 that were discussed online conversations of 19,782 users from five countries, the USA, the UK, Canada, Malaysia, and the Philippines, conducted with a SimSimi chatbot.[23] The analysis results indicated that the work provides insights into people's informational and emotional needs during a global health crisis. Users sought health-related information and shared emotional messages with the chatbot, indicating the potential use of chatbots to provide accurate health information and emotional support.

Previous studies have emphasized the importance of using text-mining techniques in analyzing and categorizing large volumes of text. In particular, no study uses a combination of clustering and classification techniques in the analysis and classification of scientific publications to reveal the degree of adaptation of national research based on the study of the subject classes of international publications.

Clustering, an unsupervised learning method, and classification, a supervised learning method, were used in this study.[24] This study aims to automatically categorize Iranian publications according to the subject categories extracted from COVID-19 international scientific publications without supervision. In fact, the results of subject categories obtained from the clustering method have been used to train the classification method in this study. In this study, unsupervised and supervised learning methods were employed to identify and discover the level of consistency between Iranian scientific publications and international scientific publications on COVID-19.

In other words, the present paper has combined clustering and text classification techniques to analyze international publications on COVID-19, particularly Iranian publications in this field. Due to the necessity and importance of the issue raised in this article and not observing similar literature, publishing this article felt more than ever.

Section 1 is the introduction of the article. The trend of the COVID-19 pandemic is briefly mentioned. In the continuation of the introduction, the research questions have been raised, and in the last part of the introduction, the literature has been reviewed. In part 2, the methodology of the paper and the steps of research implementation and methods of data analysis, models, and software are explained. Section 3 is the research findings. This section answers the research questions, and the analyzed data are reported as tables and figures. In Section 4, the results obtained are compared with the results of previous research, and the examines related to each of the findings are given. Section 5 of the article also contains the general conclusion and suggestions for future research.

Methods and materials

The present research is quantitative in nature. Big textual data were analyzed by text mining, clustering, and text classification techniques. Text mining with statistical methods and machine learning algorithms automatically extracts valuable and hidden information from textual data. Search, extraction, preprocessing, data analysis, and answering the questions in the findings section of this article started on 10 June 2021 and ended on 27 September 2021. The present research was conducted in the information management department at ISC – Shiraz – Iran.

The present paper is applied research conducted using text-mining techniques such as text classification and clustering with an analytical approach. The statistical population includes all COVID-19 publications indexed in PubMed Central® (PMC) from November 2019 to July 2021. The PMC has been used to collect the data, one of the most extensive and most reputable medical science databases globally regarding coverage.[25] A search was done on PMC on 10 June 2021, to extract the data related to international and Iranian COVID-19 publications.

Furthermore, data from international and Iranian COVID-19 publications were extracted separately. The data were extracted in File Medline format and converted to CSV using Science Space[1] for text analysis. Accordingly, 157,719 and 3143 records were extracted for international and Iranian publications on COVID-19, respectively. For all COVID-19 publications, the "title" and "abstract" fields were used to perform the text-mining process.

Text-mining techniques were used to analyze the extracted data. However, the used text-mining operations were performed in three stages: "text preprocessing," "text analysis," and "results discussion." First, each retrieved publication's title, abstract, and keywords were merged to perform the text-mining process. Then, preprocessing and data cleansing operations were applied to the data to increase the quality of data and the validity of patterns and the extracted relationships. The preprocessing steps included (a) elimination of unimportant characters such as extra empty spaces, text formatting tags, and nonalphabetic characters (removing punctuation or numbers from the text), (b) breaking text components into words and phrases, (c) converting uppercase letters to lowercase ones for text homogenization, (d) homogenization of synonyms to the preferred word, (e) homogenization of different forms of words by lemmatization method (replacing words or their basic form or vocabulary with the used conjugated ones),[26] and (f) removing stop words and those that are not valuable for retrieving or analyzing documents, such as conjunctions and suffixes (e.g. and, the, of, for). In this paper, the LDA topic modeling algorithm has been used for topic clustering. The LDA is one of the most important methods of implementing topic modeling[27] and allows the discovery of basic issues from large volumes of unstructured textual data.[28]

LDA is a statistical method that can automatically identify main topics and trends in large and unstructured collections of documents. It has been used in recent research to extract topics and discover research trends efficiently.[29,30] Also, LDA provides a powerful tool for finding hidden "topics" in large document sets and can be easily used for more complex applications.[31] Researchers have approved and accepted LDA to apply topic modeling.[32] In addition, LDA has obtained better results in both academic and nonacademic sectors in analyzing large-scale document sets.[33,34] Finally, the results of LDA are more interpretable than hierarchical Dirichlet process (HDP).[35]

LDA is a probabilistic method. For each document, the results give a mix of topics that make up that document. Each document receives a probability distribution over the k topics. Each word in the document is attributed to a particular topic with probability given by this distribution. Topics themselves are defined as probability distributions over the vocabulary. So, the results are two sets of probability distributions: the set of distributions of topics for each

document and the set of distributions of words for each topic. Each topic is itself a probability distribution over words in the vocabulary. LDA aims to obtain results indicating that each document consists of a small number of topics and each topic is primarily composed of a small number of main words.[27]

The original concept of LDA was that the text was regarded as a combination of several topics in which word distribution determined the features of a topic. In LDA, the word is called a term, the word collection is considered a document, and the document collection is considered a corpus. All words in the corpus are considered vocabulary. Moreover, LDA is a way of discovering the topic. A word distribution determines the characteristic of the subject.[27,36] An assumption applied in the LDA is that a document is regarded as a bag of words. Several procedures are required to identify the topic in the LDA by identifying the number of topics and providing random topic initialization in words.

For each corpus document, the following is a generative method: computing the probability of the word on a topic and the probability of the topic on the document to show the probability of a word on a topic and the predominance of the topic on the document and updating the topic on every single word according to the highest value of probability. The probabilistic LDA graph models in step 2 may be represented as a directed acyclic graph (DAG) in Figure 1.

The corpus level parameters α, β, and θ denote a variable at the document level, and Z and W imply the variables in the term (word) level. The variable is an object symbol capable of being filled with changing content; however, distribution occurs in this type, and the Z and W contents can be changed in this case. A variable includes particular parameters because of value alterations.[18,36]

LDA is a generative probabilistic model, including a three-level structure with word, topic, and document. In LDA, documents are viewed as a distribution over topics, while each topic is distributed over words. To generate a document, LDA firstly samples a document-specific multinomial distribution over topics from a Dirichlet distribution. Then it repeatedly samples the words from these topics.[37]
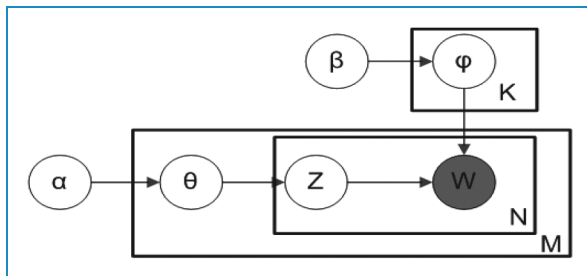

Figure 1. Probabilistic graph model of LDA.

The equation which governs the working of LDA is shown in Equation (1):

$$P(W, Z, \theta, \varphi; \alpha, \beta) = \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{i=1}^{K} P(\varphi_i; \beta)$$
$$\prod_{t=1}^{N} P(Z_{j,t}|\theta_j)P(W_{j,t}|\varphi_{j,t}) \tag{1}$$

where P (W, Z, θ, φ_i, α, β) is the probability that a specific document is produced from a hypothetical LDA machine, the settings of which will aid in knowing the latent topics in the document. α and β are the Dirichlet, while θ and φ are multinomial distributions, respectively; Z is the list of topics, and W is the corpus of Words. The value of P relies on four independent probabilities as mentioned in the right side of Equation (1) where K is one of the three model hyperparameters, i.e. number of topics, M and N denote the number of documents and word count in a given document, respectively.[38]

Moreover, the "CV coherence" was used to determine the appropriate number of topics. CV coherence is an index that measures the co-occurrence of the words extracted by the topic model. If those words from the same topic often co-occur (i.e. the *CV* coherence is high), the model is considered well-performed.[39]

In the final step, text classification algorithms were used for the topic classification of Iranian COVID-19 publications. This paper employs the support vector machine (SVM) algorithm as the classification algorithm. The SVM is one of the most important and successful algorithms in text mining and is an efficient and accurate method. SVMs were introduced by Vapnik et al.[40] for classification tasks, which adhere to the structural risk minimization principle to construct an optimal hyperplane with the widest possible margin to separate a set of data points that consist of positive and negative data examples.[41]

SVM is a powerful classification algorithm that divides the output data into categories using vector space. In simple mode, SVM separates the data using a line called hyperplane. The algorithm aims to find the best hyperplane for separating different categories.[42] Studies have shown that this algorithm performs well and appropriately classifies scientific documents.[43]

Similar to other machine learning algorithms, SVM works by supervising two stages of training and classification.[44] To implement the SVM algorithm, the extracted topics were first used as a supervised learning set in the topic modeling stage of international COVID-19 publications. Therefore, each set of keywords extracted from the topic modeling representing a topic, with the specific weights for each word, was given to the SVM as supervised learning (Figure 2).[45] Consequently, the learning process was shaped by supervision through a set of words, and each document was assigned to a topic based on the similarity of the subject word set. Ultimately, the specified tag was used as the "topic document tag."[46,47]
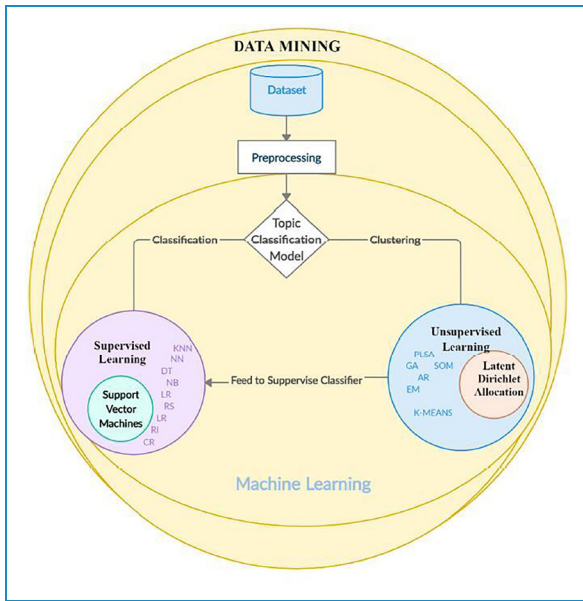
Figure 2. Stages of classification of COVID-19 international and Iranian publications.[45]

The SVM was performed based on the supervised learning of data obtained from the topic modeling stage for text classification based on the performed analysis, and it was found that each of the subject classes of Iranian COVID-19 publications establishes a relationship, consistency, and common trend with the subject classes of international publications. In this regard, the topics of Iranian COVID-19 publications were determined based on the seven topics of international publications, and a diagram was drawn for each of the seven topics of international COVID-19 publications. In this diagram, based on the weighted average of international and Iranian COVID-19 publications topics per month, the publishing trend was reviewed and compared in seven topics.

Text-mining algorithms can be implemented in Python using the scikit-learn library, and Natural Language Toolkit (NLTK).[42,48] The Skit-learn library is an open-source project in which many experts worldwide participated its development and promotion, which was developed to perform machine learning algorithms with high standards.[49]

To perform the above tasks, python programming language and libraries related to text mining, such as Gensim, NLTK, Spacy, and scikit-learn, were used.[26] Python is an open-source, simple syntax, compact, and versatile programming language that easily develops and provides various libraries for working with texts for users.[50] The main framework of data analysis in this paper is shown in Figure 3.

Results

It is necessary to emphasize two important points regarding the novelty of this study before reporting the results. First,

national (Iranian) and international publications were text-mined, and based on the results from the topic modeling of international publications, text classification was done on national publications. The second innovation is using text classification to check the consistency of national publications with international ones. With this finding, it can be concluded to what extent the Iranian researchers of COVID-19 have conducted their research in both directions with international researchers.

Topic modeling of international COVID-19 publications

Table 1 presents the topics from implementing the LDA topic modeling algorithm on international COVID-19 publications. This table indicates the topic number, the subject expert's selected name, and the most important words of each topic. It is worth mentioning that the name of each topic has been determined based on a review of the most important words and the most relevant titles of the publications of each topic by consulting with topic experts.

Table 1 shows the results of implementing the topic modeling algorithm on international COVID-19 publications. For example, Table 1 shows that 10 of the most important words of topic 0 were performance, test, training, sample, image, propose, detect, base, content, and technique, which was named "diagnostic tests." Also, the most important words of topic 1 were virus, model, human, infection, show, gene, analysis, protein, mouse, and identify, which is named "COVID-19 Proteins: Vaccine and Antibody Response."

The following seven topics were used: (1) Diagnostic Tests; (2) COVID Proteins: Vaccine and Antibody Response; (3) Vaccine Immunogenicity; (4) Other; (5) Social and Technology in COVID-19; (6) COVID-19 Complication; and (7) COVID-19 and Immune System.

Classification of Iranian publications based on the topics of international COVID-19 publications

Figure 3 shows the (share) of international COVID-19 publications in the seven topics separately.

Figure 4 indicates that Topic 4 (Social and Technology in COVID-19), with 50.61%, provides the largest share of publications among the seven topics. Moreover, Topic 5 (COVID-19 Complication), with 15.94%, and Topic 2 (Vaccine Immunogenicity), with 13.92%, received the second and third places, respectively. Furthermore, Topic 3 (Other), with 0.58%, indicates the lowest share of international COVID-19 publications of the seven topics.

Figure 5 shows the results of classifying Iranian COVID-19 publications based on the topics of international ones.

Figure 5 indicates that the topic "Social and Technology in COVID-19," with 39.44%, shows the largest share of Iranian publications. "COVID-19 Complication" and
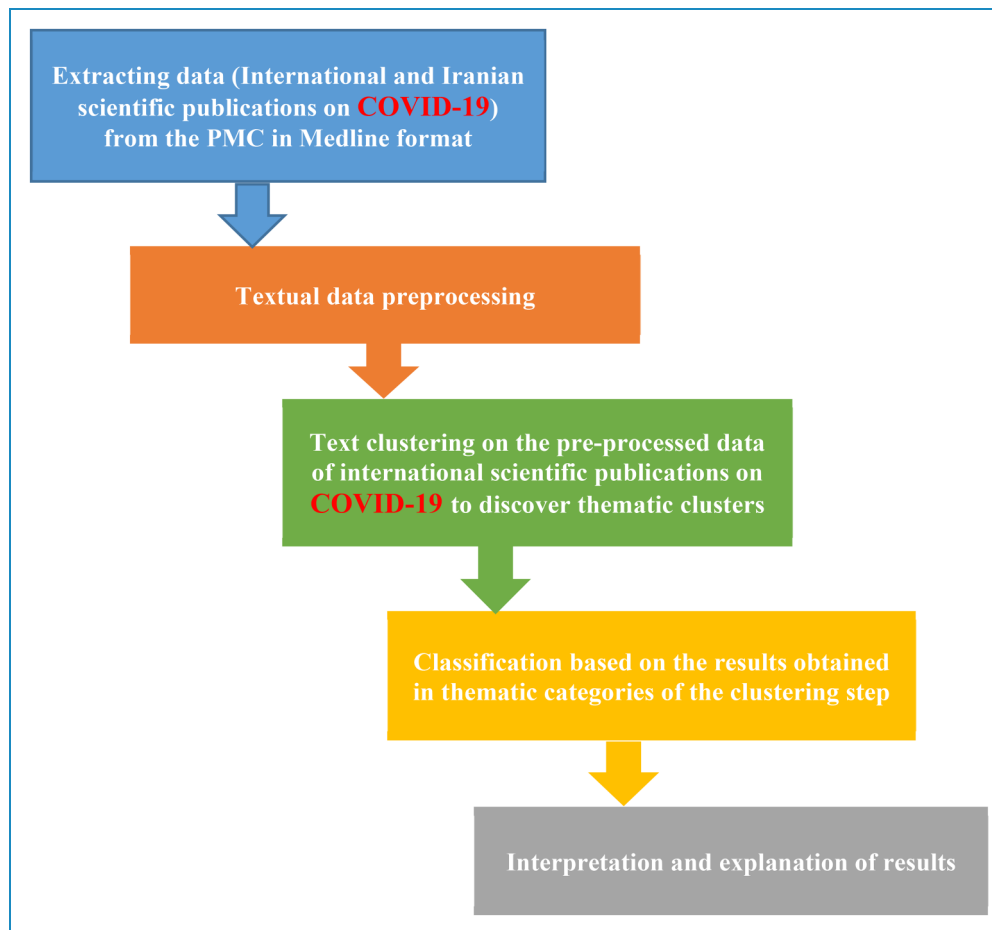
Figure 3. An overview of the data analysis framework used in this study.

"Vaccine Immunogenicity," with 26.25% and 15.83%, respectively, ranked second and third in Iranian COVID-19 publications. Moreover, the topic "Other" with 0.13%, indicates the lowest share of Iranian COVID-19 publications.

The monthly publication trends of topics extracted from international COVID-19 publications (Figure 6(a)) and Iranian COVID-19 publications (Figure 6(b)) are shown in Figure 6.

Figure 6(a) indicates that Topic 4 (Social and Technology in COVID-19) dedicated the highest number of international COVID-19 publications. The data analysis also showed the highest number of international publications on this topic was in April 2021. Moreover, Figure 6(b) shows that in February 2021, "Social and Technology in COVID-19" provided the highest rate in Iranian publications. Also, "Vaccine Immunogenicity" is another topic with the highest number of publications on COVID-19 in recent months. Furthermore, the highest number of international scientific publications related to this topic was published in March 2021, and the highest number of Iranian publications was on this topic in October 2020.

The topic "COVID-19 Complication" had the highest number of international publications on COVID-19 from the beginning of the pandemic until December 2020, followed by "Vaccine Immunogenicity."

"COVID-19 Complication" publications were most prevalent in July 2020 and "Vaccine Immunogenicity" publications in March 2021. Figure 6(a) shows that the growth trend of publications on all topics of COVID-19 has been decreasing since March 2021. Figure 6(b) indicates that Iranian publications on "COVID-19 Complication" have grown significantly; the highest number was in November 2020. Another topic growing in Iranian publications on COVID-19 recently is "Vaccine Immunogenicity," which had the highest number of publications in October 2020.

Degree of consistency of the subject classes of Iranian and international publications based on the text classification method

Figure 7 indicates the trends of international and Iranian publications on COVID-19 in seven topic categories.

Table 1. Results of applying the topic modeling algorithm to international COVID-19 publications.

| Topic number	Topic name	Essential words in each topic
0	Diagnostic Tests	Performance, test, training, sample, image, propose, detecting, base, content, and technique
1	COVID-19 Proteins: Vaccine and Antibody Response	Virus, model, human, infection, show, gene, analysis, protein, mouse, and identify
2	Vaccine Immunogenicity	Group, participant, year, compare, outcome, rate, increase, analysis, factor, and age
3	Other	Clinical trial, drug, project, vaccine, efficacy, treatment, dose, safety, and build
4	Social and Technology in COVID-19	health, develop, provide, system, approach, social, Support, model, work, and care
5	COVID-19 Complication	Patient, treatment, clinical, disease, cancer, case, include, therapy, mental, and diagnosis
6	COVID-19 and the Immune System	Cell, effect, response, function, expression, role, increase, mechanism, immune, and level
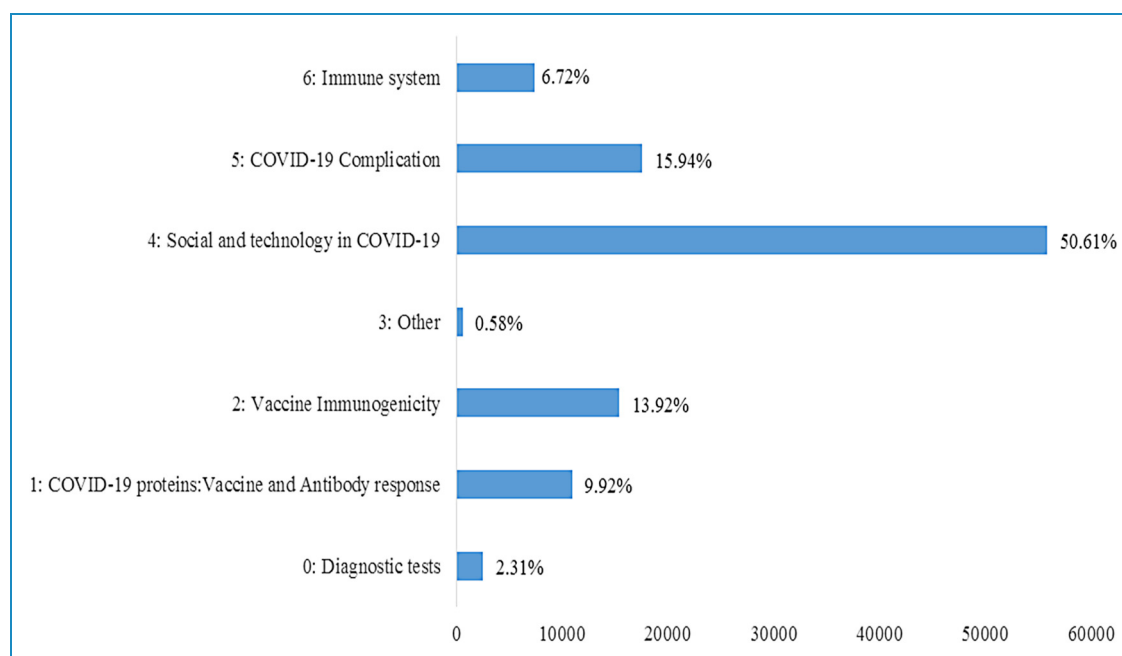
Figure 4. Share of international COVID-19 publications on the seven topics.

Figure 7 indicates the monthly trend of international and Iranian publications on COVID-19. The results indicate that the trend of Iranian COVID-19 publications on the topic "COVID-19 Proteins: Vaccine and Antibody Response" in January 2020 has increased significantly (higher than 70%). However, in February and March 2020, there was a significant decrease in Iranian publications on this subject. From April 2020 to June 2021, there was a steady trend of 10%–15% in Iranian publications. An investigation of international publications on the topic "COVID-19 Proteins: Vaccine and Antibody Response" also demonstrates that there has been a
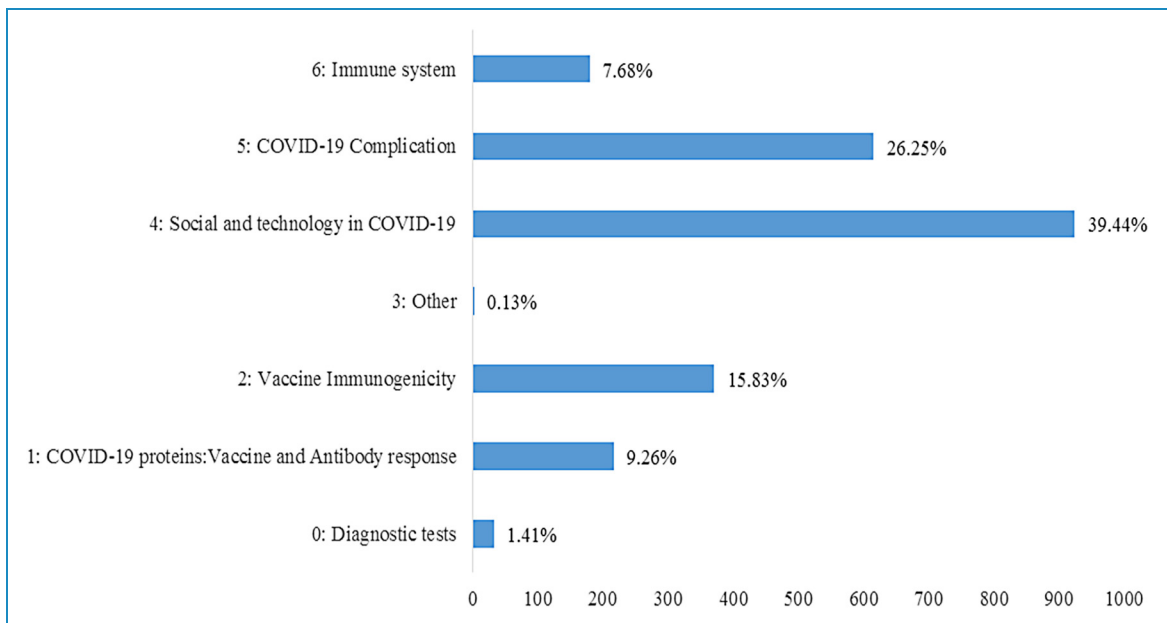
Figure 5. Topic classification of Iranian COVID-19 publications based on the topics extracted from the topic modeling of international publications on COVID-19.
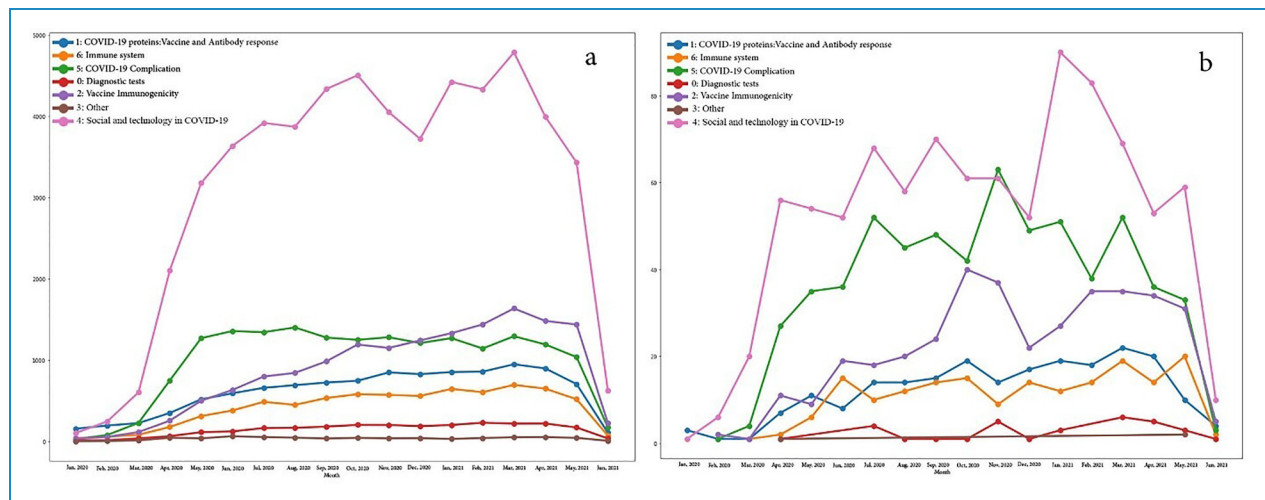

Figure 6. Monthly publication trend of COVID-19 topics of (A) international publications and (B) Iranian publications.

declining trend from January to April 2020. From May 2020 to June 2021, a steady trend of 10% is observed. Therefore, from mid-2020 to June 2021, Iranian publications followed the global publishing trend in "COVID-19 Proteins: Vaccine and Antibody Response." A common trend in this category is observed between international and Iranian publications. In other words, Iranian publications have a common publishing and research process with international publications in "COVID-19 Proteins: Vaccine and Antibody Response." Figure 7 indicates that the Iranian publication on COVID-19 is in line with international publications only in one of the seven topic categories. There was no consistency between Iranian and international publications in the other six topic categories.

Discussion

The COVID-19 global pandemic has posed a significant crisis for public health and the lives of all people worldwide. Its rapid spread and danger have endangered public health and all aspects of human life, including economic and social development. The solution to this crisis is to fully understand the problem and its associated

consequences in different areas and identify possible solutions to deal with it. Therefore, it is necessary to improve scientific knowledge about COVID-19 because it leads to answers to questions in real life. However, in-depth knowledge in this area is needed to identify solutions and answers to the numerous existing problems and challenges.[51] Accordingly, the present article has indicated a clear picture of the topics of Iranian and international publications on COVID-19. The topic modeling results of international COVID-19 publications showed that these publications had been published on seven main topics. In order of their publication share from highest to lowest, these topics are "Social and Technology in COVID-19"; "COVID-19 Complication"; "Vaccine Immunogenicity"; "COVID-19 Proteins: Vaccine and Antibody response"; "Immune System"; "Diagnostic Tests"; and "Other."

"Social and Technology in COVID-19" is the topic that has attracted the most attention of researchers; 2 months after the outbreak of the COVID-19 pandemic in late 2019, it has had the highest publication among other topics. The study results indicate that this issue's highest publication rate was in April 2021.

Social challenges are among the most critical issues resulting from COVID-19.[52] Social distance and lockdown have been successful methods for managing pandemics, along with the use of technology.[53] In this regard, the Centers for Disease Control and Prevention (CDC) considers the use of information technology in the management and control of COVID-19 necessary.[54] Given the development and progress of the existing information and communication technology infrastructures, governments and health organizations at the national and international levels can use intelligent approaches to overcome the pandemic; therefore, during the COVID-19 crisis, there is an upward trend in applying new technologies in the health field for dealing with and controlling this crisis.[55] The above explanation is one of the main reasons for researchers' popularity of "Social and Technology in COVID-19." In this regard, Gupta et al. have proposed an extensive range of topics in society, including "Socioeconomic Impact" and "Communication," for COVID-19 publications and have shown an increase in publications related to this topic. Their study also showed "Health care, Telemedicine" as another topic with many publications.[18] In another study, Rodríguez-Rodríguez et al. highlighted the applying technology in society during the COVID-19 pandemic as one of the main topics in publications related to using AI in COVID-19.[9] The second important issue was the "COVID-19 Complication."

According to the findings and the publication trend in this issue, as it is known, the area of complications has been more frequent than other areas and has had a steady trend over time. The reason may be that the new coronavirus has systemic involvement and appeared in various variants. Therefore, at any time, there are reports of new complications

of this disease, among which long-term complications are worrying. Regarding "Social and Technology in COVID-19," it can be added that since the beginning of the COVID-19 pandemic, various complications have been reported worldwide, some with high and some with low prevalence. For example, complications such as adult respiratory distress syndrome (ARDS), cardiovascular, coagulation, and thrombotic problems are among the most critical complications in the investigations. These three essential factors have been reported as an essential reason for the death of patients suffering from COVID-19. However, these are not the only complications. The articles also mention rare complications, including blood, gastrointestinal, external, microbial agent infection, and cardiac arrest. The psychological complications of COVID-19, among others, are categorized in this topic. In addition to physical symptoms, COVID-19 can cause mental problems for patients. Besides the physical symptoms, COVID-19 can cause psychological problems for patients.

On the other hand, changes in people's lifestyles following the outbreak of COVID-19, such as travel restrictions, lockdown and isolation, social distancing, and economic status, may similarly lead to mental disorders. In addition to the mentioned cases, some psychological problems are determined in specific groups, such as pregnant women, medical workers, and children during the COVID-19 pandemic.[56] Anderson identified different topic clusters of COVID-19 complications, such as brain issues and stroke/blood vessels, heart/blood vessels, respiratory, breathing, and lung issues, and public and mental health concerns in COVID-19 publications.[3] Moreover, Gupta et al. identified topic clusters associated with "COVID-19 Complication," including gastroenterology, cardiovascular, pulmonary, and mental health, in COVID-19 publications. They showed that mental health is a primary and growing topic in COVID-19 publications.[15] Älgå, Eriksson, and Nordberg believe that in the first 6 months of the COVID-19 pandemic, some issues, such as healthcare responses, clinical manifestations, and psychological effects, have been the most important topics published in the articles related to COVID-19.[57]

Another topic that has received the most attention from researchers at COVID-19 global publications is "Vaccine Immunogenicity." The power of immunogenicity against the virus is an important feature of a vaccine. In discussing the power of the vaccine in creating collective safety and immunity of individuals, the higher the immunization of the vaccine, the greater its efficiency and effectiveness. In the discussion of immunization of vaccines, both arms of the specific immune system, namely, humoral immunity (with the role of neutralizing antibodies) and cellular immunity (with the role of the cytotoxic T lymphocytes (CD8$^+$)), are involved and active and any vaccine that can address both areas, especially in the field of immunity.
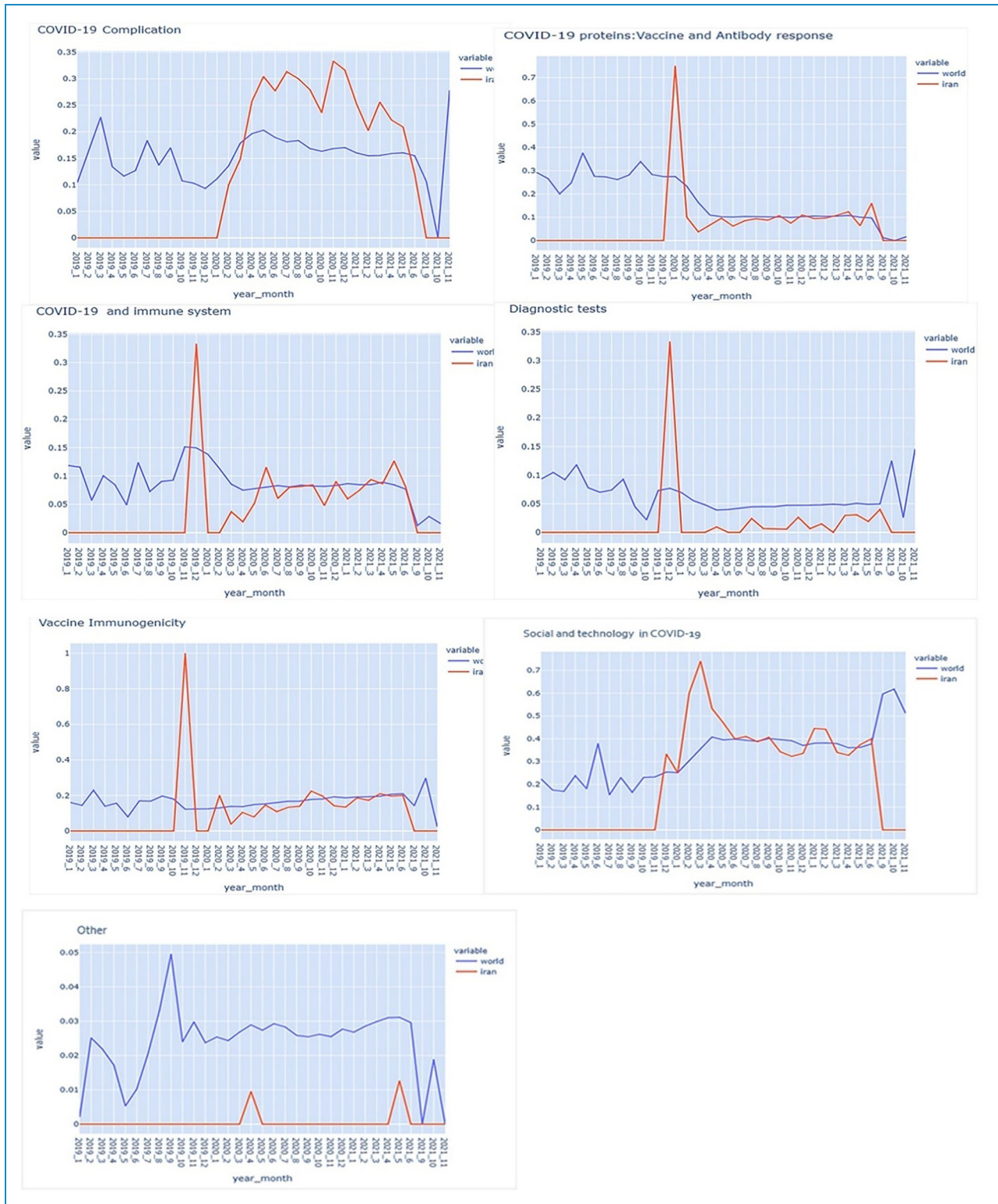
Figure 7. Monthly trends of international and Iranian COVID-19 publications on seven topics.

It stimulates the cell more and has high efficiency and effectiveness. Vaccine immunization is influenced by important host and human factors and the vaccine, and factors such as vaccine dose, injection site, and vaccine dose schedule can be effective in vaccine immunization in addition to viral target.[58]

Another important topic is the "Vaccine and Antibody Response," the vaccine injection aimed at stimulating the immune system, stimulating a specific system or the same acquired immunity. Most studies in this field examine the immune system of individuals after vaccination.[59] Haleem et al. concluded that extensive research is needed to develop vaccines to prevent COVID-19 infection. To combat this disease comprehensively, there is an urgent need for the early production and development of various drugs, vaccines, and treatments.[60]

"Immune System" and "Diagnostic Tests" are other topics in which international COVID-19 publications are categorized in this article. However, organs of the immune system play an important role in the fight against infectious agents (especially viruses) in both innate and specific immune systems, including cellular and humoral immunity. However, this close connection between the coronavirus and the immune system in different people causes various symptoms. Therefore, studies on the immune system can include therapeutic mechanisms of the immune system or strengthen it versus the defense against viral agents and indicate the more acute clinical condition of the disease in individuals, which may not improve and can lead to disease progression. It becomes acute as well.[61] In the discussion of diagnostic tests, the standard diagnosis of a virus is by molecular-genetic methods, e.g. reverse transcription-polymerase chain reaction (RT-PCR) or the determination of the viral genome in nasopharynx samples. However, other diagnostic methods include approaches based on serological methods or the determination of antivirus antibodies. In addition, a computerized tomography scan (CT scan) of the lungs and observation of the ground glass view are also of diagnostic value.[62]

Danesh and Dastani also indicated that the 50-year publications on coronavirus had been categorized into eight topics: "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."[13] In another study, the main topics of COVID-19 publications include pathogenicity, epidemiology, transmission, diagnosis, treatment, prevention, and complications.[63] Another research also focused on articles on COVID-19 and publications focusing on virology, immunology, epidemiology, pharmacology, public health, critical care, and emergency medicine.[64] Due to the extensive range of categories and topics published on COVID-19, it can be said that the results of previous research are in line with the present article. In general, virology, vaccines, antivirals, and health research are at the core of the scientific response to the COVID-19 pandemic; there are other studies in other fields.[65]

A topic classification method in text analysis was used in the present study to assess the consistency between Iranian and international publications regarding COVID-19. There was a greater proportion of Iranian publications related to three topics: "vaccine and virus purposes," "COVID-19 complications," and "vaccine safety." In accordance with the study results, Iranian researchers place a greater emphasis on "complications." Additionally, the trend in Iranian publications on "COVID-19 Proteins: Vaccine and Antibody Response" is consistent with that in international publications. In contrast, no consistency was observed in other topics among publications.

Among Iranian publications on COVID-19, Dastani and Danesh found that the highest quantity was on "Prevention," which includes the following topics: "Pandemic status," "Management," "Policy," "Control," "Behaviors," and "Other diseases." It was found that "Treatment" was the most common topic, with subtopics such as "Clinical features of mortality," "Clinical features of the disease," "Drug," and Outcome.[17] Due to Dastani and Ghorbani's analysis, Iran's research projects related to COVID-19 can be categorized into 12 subtopics.[66] The following subtopics were identified: "Treatment," "Care needs of medical staff," "Disease severity factors," "Mental health and preventive behavior," "Diagnostic and laboratory," "Immunological studies," "Vitamins and minerals," "cardiovascular disease," "Vaccine studies," "Job and life stress," "Experiences of nurses," "Patients and families," and "Prevalence and symptoms." Based on the results of this study, the previous studies are inconsistent.[66,17]

Conclusion

The results of this study indicated a rich concept of the published issues of COVID-19 and the trend of publishing Iranian and international topics. The results of this study showed that the number of scientific publications in "Diagnostic Tests"; "COVID Proteins: Vaccine and Antibody Response"; "Vaccine Immunogenicity"; "Other"; "Social and Technology in COVID-19"; "COVID-19 Complication"; and "COVID-19 and Immune System" were aligned at the international and national levels. Hence, the highest number of scientific publications on COVID-19 internationally and in Iran was in "Social and Technology in COVID-19."

Also, the trend of scientific publications on COVID-19 at the national and international levels on the topic "Social and Technology in COVID-19" is similar. It has the highest number of publications in the study period. However, there are differences in international and national publication trends on other topics.

The main limitations of this article are as follows: The time frame of this study is from December 2019 to July 2021. Therefore, it has not been investigated from August 2021 to June 2023. Many articles related to the subject area of COVID-19 have been published and indexed in PMC, which

are not part of the statistical population of the present article and need to be studied in an independent research from August 2021 to July 2023. The second limitation is geographical. In this article, international publications are compared with Iranian publications. This limitation can be studied in future research. The value and originality of this article is to provide an enlightened model for the study of international publications with a specific country that can be used in future studies. The two limitations of the time frame and the geographical limitation were clearly explained, and suggestions for future studies were presented.

The methodology used in this paper, which led to identifying national subject classes based on international ones, would also be used by researchers in the future. The manual classification of topics in scientific publications is an impossible task. Therefore, using supervised and unsupervised learning methods in text mining, the best results can be obtained in the shortest possible time so that the least amount of interference has a human.

Correspondingly, the important results of the present study were among the monthly trends of national and international publications based on the identified topic categories and the discovered degree of consistency of the national and international research. Therefore, researchers and policymakers could understand COVID-19 knowledge based on the results of this study. Thus, in future studies, it is suggested that researchers study and analyze other areas of COVID-19 knowledge, such as prevention methods, treatment methods, and vaccines, separately and by applying text classification and clustering techniques. It is also suggested that in a longitudinal study from July 2021 to the present, text mining of International COVID-19 publications be performed, and the results will be compared with the findings of the present paper. Furthermore, this investigation examined the degree of consistency of subject classes of Iranian and international publications on COVID-19. It is suggested that in another study, the publications of other countries be compared and analyzed with the international publications on COVID-19, the degree of consistency of their research be examined, and the results be compared with the present study.

Contributorship: FD has contributed to this article in the following sections: conceptualization, data curation, methodology, writing – review and editing, and approved the version to be published. MD has contributed to this article in the following sections: formal analysis, investigation, resources, software, visualization, and writing, and drafted the article or revised it critically for important intellectual content.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: Only articles indexed in PMC on COVID-19 were extracted and analyzed in this research. Also, no human or animal was used to conduct the present research. Therefore, based on the nature of text-mining research, "Ethics approval" was not required for this type of research.

Guarantor: M.D.

Informed consent: Not applicable, because this article does not contain any studies with human or animal subjects.

Trial registration: Not applicable, because this article does not contain any clinical trials.

ORCID iD: Meisam Dastani https://orcid.org/0000-0002-5631-539X

Note

1. http://medialab.github.io/sciencescape/medline_utils

References

1. Davenport T. *Big data at work: dispelling the myths, uncovering the opportunities*. Brighton, MA: Harvard Business Review Press, 2014.
2. Simon C, Davidsen K, Hansen C, et al. Bioreader: a text mining tool for performing classification of biomedical literature. *BMC Bioinform* 2019; 19: 165–170.
3. Anderson BS. Using text mining to glean insights from COVID-19 literature. *J Inf Sci* 2021; 49: 01655515211001661. https://doi.org/10.3390/healthcare8010046
4. Wilson D, Sheikh A, Görgens M, et al. Technology and universal health coverage: examining the role of digital health. *J Glob Health* 2021; 11: 1–12.
5. Krallinger M, Leitner F and Valencia A. Analysis of biological processes and diseases using text mining approaches. *Bioinform Methods Clin Res* 2010; 593: 341–382.
6. Kumari S, Vani V, Malik S, et al. Analysis of text mining tools in disease prediction. In Hybrid Intelligent Systems: 20th International Conference on Hybrid Intelligent Systems (HIS 2020), December 14–16, 2020, 2021 (pp. 546–564). Springer International Publishing. https://doi.org/10.1007/978-3-030-73050-5_55
7. Meskarpour Amiri M, Nasiri T and Mehdizadeh P. Subjects clustering analysis and science mapping on COVID-19 researches in scopus database. *J Military Med* 2020; 22: 663–669.
8. Song P and Karako T. COVID-19: real-time dissemination of scientific information to fight a public health emergency of international concern. *Biosci Trends* 2020; 14: 1–2.

9. Rodríguez-Rodríguez I, Rodríguez JV, Shirvanizadeh N, et al. Applications of artificial intelligence, machine learning, big data and the internet of things to the COVID-19 pandemic: a scientometric review using text mining. *Int J Environ Res Public Health* 2021; 18: 8578–8604.

10. Silva TF, Demelo AA, Dos Santos DLA, et al. Systematic review of bibliometric studies on SARSCoV-2. *Revista Ciências em Saúde* 2020; 10: 116–125.

11. Tran BX, Ha GH, Nguyen LH, et al. Studies of novel coronavirus disease 19 (COVID-19) pandemic: a global analysis of literature. *Int J Environ Res Public Health* 2020; 17: 4095–4111.

12. Delen D and Crossland MD. Seeding the survey and analysis of research literature with text mining. *Expert Syst Appl* 2008; 34: 1707–1720.

13. Feng L, Chiam YK and Lo SK. Text-mining techniques and tools for systematic literature reviews: a systematic literature review. In 2017 24th Asia-Pacific Software Engineering Conference (APSEC), 2017 (pp.41–50). IEEE. https://doi.org/10.1109/APSEC.2017.10

14. Liu Q, Zheng Z, Zheng J, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. *J Med Internet Res* 2020; 22: e19118.

15. Ghasiya P and Okamura K. Investigating COVID-19 news across four nations: a topic modeling and sentiment analysis approach. *IEEE Access* 2021; 9: 36645–36656.

16. Danesh F, Dastani M and Ghorbani M. Retrospective and prospective approaches of coronavirus publications in the last half-century: a latent Dirichlet allocation analysis. *Libr Hi Tech* 2021; 39: 855–872.

17. Dastani M and Danesh F. Iranian COVID-19 publications in LitCovid: text mining and topic modeling. *Sci Program* 2021; 2021: 1–9.

18. Gupta A, Aeron S, Agrawal A, et al. Trends in COVID-19 publications: streamlining research using NLP and LDA. *Front Digit Health* 2021; 3: 1–15.

19. Koh K, Lee S, Park S, et al. Media reports on COVID-19 vaccinations: a study of topic modeling in South Korea. *Vaccines (Basel)* 2022; 10: 2166.

20. Silalahi SF, Arini HM and Mulyani YP. Topic modelling and sentiment analysis during COVID-19 pandemic response: a systematic review. In AIP Conference Proceedings, 2023 Feb 14 (Vol. 2654, No. 1, p. 040018). AIP Publishing LLC. https://doi.org/10.1063/5.0117344

21. Amores JJ, Blanco-Herrero D and Arcila-Calderón C. The conversation around COVID-19 on twitter—sentiment analysis and topic modelling to analyze tweets published in English during the first wave of the pandemic. *Journalism Media* 2023; 4: 467–484.

22. Dehghani M and Ebrahimi F. ParsBERT topic modeling of Persian scientific articles about COVID-19. *Inform Med Unlocked*. 2023; 36: 101144.

23. Chin H, Lima G, Shin M, et al. User-chat bot conversations during the COVID-19 pandemic: study based on topic modeling and sentiment analysis. *J Med Internet Res* 2023; 25: e40922.

24. Lee CH and Yang HC. Construction of supervised and unsupervised learning systems for multilingual text categorization. *Expert Syst Appl* 2009; 36: 2400–2410.

25. Falagas ME, Pitsouni EI, Malietzis GA, et al. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 2008; 22: 338–342.

26. Malaterre C, Lareau F, Pulizzotto D, et al. Eight journals over eight decades: a computational topic-modeling approach to contemporary philosophy of science. *Synthese* 2021; 199: 2883–2923.

27. Blei DM. Probabilistic topic models. *Commun ACM* 2012; 55: 77–84.

28. Guo Y, Barnes SJ and Jia Q. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Manag* 2017; 59: 467–483.

29. Cao Q, Cheng X and Liao S. A comparison study of topic modeling based literature analysis by using full texts and abstracts of scientific articles: a case of COVID-19 research. *Libr Hi Tech* 2023; 41(2): 543–569. https://doi-org.librarylogin-cupey.uagm.edu/10.1108/LHT-03-2022-0144

30. Watanabe G, Conching A, Nishioka S, et al. Themes in neuronavigation research: a machine learning topic analysis. *World Neurosurg: X* 2023; 18: 100182.

31. Liu L, Tang L, Dong W, et al. An overview of topic modeling and its current applications in bioinformatics. *Springer Plus* 2016; 5: 1608.

32. Muchene L and Safari W. Two-stage topic modelling of scientific publications: a case study of university of Nairobi, Kenya. *PloS one* 2021; 16: e0243208.

33. Park H and Park MS. Capturing the trend of mHealth research using text mining. *mHealth* 2019; 5: 1–12.

34. Gough D and Richardson M. *Advanced research methods for applied psychology*. Oxfordshire, UK: Routledge, 2018, pp. 75–87. https://doi.org/10.4324/9781315517971

35. Carmichael JJ and Eaton SE. Security risks, fake degrees, and other fraud: a topic modelling approach. In: Eaton SE, Carmichael JJ and Pethrick H (eds) *Fake degrees and fraudulent credentials in higher education*. Cham: Springer International Publishing, 2023, pp. 227–250.

36. Blei DM, Ng AY and Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res* 2003; 3: 993–1022.

37. Cao J, Xia T, Li J, et al. A density-based method for adaptive LDA model selection. *Neurocomputing* 2009; 72: 1775–1781.

38. Gupta RK, Agarwalla R, Naik BH, et al. Prediction of research trends using LDA-based topic modeling. *Global Trans Proc* 2022; 3: 298–304.

39. Wang M and Mengoni P. How pandemic spread in news: text analysis using topic model. In 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2020 Dec 14 (pp. 764–770). IEEE. https://doi.org/10.1109/WIIAT50758.2020.00118

40. Vapnik VN. An overview of statistical learning theory. *IEEE Trans Neural Netw* 1999; 10: 988–999.

41. Deng X, Li Y, Weng J, et al. Feature selection for text classification: a review. *Multimed Tools Appl* 2019; 78: 3797–3816.

42. Goh YM and Ubeynarayana CU. Construction accident narrative classification: an evaluation of text mining techniques. *Accid Anal Prev* 2017; 108: 122–130.

43. Gürbüz T and Uluyol Ç. Research article classification with text mining method. *Concurrency Comput: Pract Exp* 2023; 35: e7437.

44. Leopold E, May M and Paaß G. Data mining and text mining for science & technology research. In: *Handbook of quantitative science and technology research*. Dordrecht: Springer, 2004, pp. 187–213. https://doi.org/10.1007/1-4020-2755-9_9

45. Srivastava SK, Singh SK and Suri JS. Effect of incremental feature enrichment on healthcare text classification system:

a machine learning paradigm. *Comput Methods Programs Biomed* 2019; 172: 35–51.

46. Žizka J, Darena F and Svobada A (2019) *Text mining with machine learning: principles and techniques*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9780429 469275

47. Shokouhian M, Asemi A, Shabani A, et al. Presenting a thematic model of health scientific productions using text-mining methods. *Iran J Inform Proc Manag* 2022; 35: 553–574.

48. de Godoi Brandão J and Calixto WP. N-Gram and TF-IDF for feature extraction on opinion mining of tweets with SVM classifier. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), 2019 Sep 21 (pp.1–5). IEEE. https://doi.org/10.1109/IDAP.2019.8875900

49. Varoquaux G, Buitinck L, Louppe G, et al. Scikit-learn: machine learning without learning the machinery. *GetMobile: Mobile Comput Commun* 2015; 19: 29–33.

50. Rehurek R and Sojka P. Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks, 2010.

51. Aristovnik A, Ravšelj D and Umek L. A bibliometric analysis of COVID-19 across science and social science research landscape. *Sustainability* 2020; 12: 9132–9162.

52. Oyebode O, Ndulue C, Adib A, et al. Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: text mining and thematic analysis approach. *JMIR Med Inform* 2021; 9: e22734.

53. Allam Z and Jones DS. On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. In Healthcare, 2020 Feb 27 (Vol. 8, No. 1, p.46). MDPI. https://doi.org/10.3390/ healthcare8010046

54. WHO C, World Health Organization. Responding to community spread of COVID-19. *Interim Guidance* 2020; 3: 1–6.

55. Handforth C. Bringing the power of global innovation to tackle COVID-19. United Nations Development Program (UNDP). 2020 Apr.

56. Hosseini M and Daryabor A. Psychological effects resulting from COVID-19 and rehabilitation intervention: a review of literature. *Sci J Rehabil Med* 2021; 10: 618–629.

57. Älgå A, Eriksson O and Nordberg M. Analysis of scientific publications during the early phase of the COVID-19 pandemic: topic modeling study. *J Med Internet Res* 2020; 22: e21559.

58. He Q, Mao Q, Zhang J, et al. COVID-19 vaccines: current understanding on immunogenicity, safety, and further considerations. *Front Immunol* 2021; 4: 669339.

59. Wei J, Stoesser N, Matthews PC, et al. Antibody responses to SARS-CoV-2 vaccines in 45965 adults from the general population of the United Kingdom. *Nat Microbiol* 2021; 6: 1140–1149.

60. Haleem A, Javaid M, Vaishya R, et al. Areas of academic research with the impact of COVID-19. *Am J Emerg Med* 2020; 38: 1524–1526.

61. Paces J, Strizova Z, Daniel SM, et al. COVID-19 and the immune system. *Physiol Res* 2020; 69: 379.

62. Benzigar MR, Bhattacharjee R, Baharfar M, et al. Current methods for diagnosis of human coronaviruses: pros and cons. *Anal Bioanal Chem* 2021; 413: 2311–2330.

63. Dehghanbanadaki H, Seif F, Vahidi Y, et al. Bibliometric analysis of global scientific research on coronavirus (COVID-19). *Med J Islam Repub Iran* 2020;34: 51–59.

64. Pal JK. Visualizing the knowledge outburst in global research on COVID-19. *Scientometrics* 2021; 126: 4173–4193.

65. Liu N, Chee ML, Niu C, et al. Coronavirus disease 2019 (COVID-19): an evidence map of medical literature. *BMC Med Res Methodol* 2020; 20: 1–11.

66. Dastani M and Ghorbani M. Bibliometric and text mining analysis on COVID-19 research projects in Iran. *Depiction Health* 2021; 12: 333–344.