Mining of Emerging trends of Covid-19 thematic areas in National and International publications

Farshid Danesh [ID], Forough Rahimi* [ID]

Assistant Professor, Department of Information Management, Regional Information Center for Science and Technology (RICeST), Shiraz, Iran

ABSTRACT

Background and Aim: The results from the analysis of COVID-19 literature by employing text-mining techniques are of particular importance for researchers, policymakers, and planners of medical sciences at the national and international levels, avoiding parallel research and waste of time and budget. The paper explore emerging topics and the trend of scientific words at the national and international levels in the subject area of COVID-19.

Materials and Methods: This applied research was conducted by employing the text-mining and its related algorithms and classifying texts. The population consists of all COVID-19 articles indexed in PubMed Central® (PMC). The number of records retrieved was 160,862 items until June 10, 2021. Among these, 3143 national and 157,719 international COVID-19 articles. Python and its related libraries were applied. The most significant words were also identified and reported based on TF-IDF weighting. Emerging topics were identified according to the weighted average growth.

Results: "COVID", "infect", and "cell" were among the most important words used in international COVID-19 articles. In addition, the most important words in the national COVID-19 articles were "patient", "SARS-Cov", and "COVID".

Conclusion: Among the most important conclusions that can be inferred from the trend of word change used in the COVID-19 literature is that the most significant words in international literature differ significantly from those in national literature, as international research focuses on COVID-19 and the infections caused by it. In contrast, national research focuses on COVID-19 and patients. Another significant result is the annual word-changing national and international literature.

Keywords: Covid-19, Text Mining, TF-IDF, Classification, Clustering, Emerging Topics, Python

Use a device to scan and read the article online

Danesh F, Rahimi F. Mining of Emerging trends of Covid-19 thematic areas in National and International publications. Iran J Med Microbiol. 2023; 17(2);150-60.

Download citation: BibTeX | RIS | EndNote | Medlars | ProCite | Reference Manager | RefWorks
Send citation to: 😊 Mendeley Ⓩ Zotero 🔲 RefWorks

1. Introduction

By emerging the COVID-19 pandemic in December 2019 in the city of Wuhan, China, governments, as well as national and international institutions, strongly supported an extensive range of studies addressing this issue while adopting the appropriate policies, public investment, incentive policies, and favorable political orientations to ensure people's well-being and deal with this hazardous virus. As a result of the urgency and importance of dealing with COVID-19, researchers from different countries and in different subject areas of medical sciences have been prompted to conduct studies in various fields to solve the problem and provide solutions. Some studies were conducted to seek new treatment methods, and others were performed to discover novel vaccines or medicines that could help new patients and infected people in treating COVID-19 [1, 2].

In addition to the high quantity and velocity of articles on COVID-19, the mentioned scientific studies

included important and practical findings that led to the increased attention of researchers and the public (3). However, these studies are not enough to know the importance of published research and require analytical research and modeling (4). The remarkable effectiveness of the mentioned scientific studies reveals more than ever the necessity of analyzing the articles and scientific topics on COVID-19.

At the end of the third year since the beginning of this dangerous pandemic that has begun to decline, hundreds of thousands of studies on COVID-19 can be found indexed in the most reliable medical science database, PubMed Central® (PMC) (5). The large volume of credible literature on COVID-19 worldwide demonstrates the urgency of monitoring and analyzing scientific texts on COVID-19 for scholars at the micro level and policymakers and planners at the macro level more than ever, despite the downward trend of COVID-19 patients and deaths. Thus, the results derived from the analysis of the published documents on COVID-19 with text-mining techniques could be beneficial to researchers, policymakers, and planners of medical sciences at the national and international levels. Researchers in this field may find the study's findings very helpful in prioritizing future research projects, selecting research topics, and avoiding the time and money-wasteful parallel studies. Besides that, the policymakers in this area can use the results of the present study to set fruitful and efficient research priorities according to the word change process and emerging topics and have better strategic policies (3, 6-9). Health officials and policymakers can adapt the global and national trends of international articles with the topics, bases, and trends of national studies in addition to the mentioned necessities (10). They can also answer this serious question of whether the studies of Iranian scientists on COVID-19 are in the same direction and line with international literature. It is required to answer the following questions in order to solve the mentioned problem:

1. What are the most important keywords used in international and national scientific articles on COVID-19 in the PubMed Central® (PMC) database?

2. How is the trend of words changes used in international and national scientific articles on COVID-19 in the PubMed Central® (PMC) database?

3. What are the emerging topics in international and national scientific articles on COVID-19 in the PubMed Central® (PMC) database?

In order to analyze the published articles on COVID-19 in various documents, including the texts of users in social networks or the texts of articles indexed in databases, several studies have been carried out during the COVID-19 pandemic. Tran et al. concluded that the United States, China, and the European Union produce the most scientific outcomes in this field. The guidelines for emergency care and surgical procedures, viral pathogenesis, and global responses to the COVID-19 pandemic were the three most important topics covered in the studies under review (11). The main differences between their article and the current study are the time, the quantity and variety of study documents, the methodology, and the comprehensiveness of the results. Using the COVID-19 database (https://cord19.aws/), Cheng, Cao & Liao also carried out text-mining of the COVID-19 literature in three parts of SARS, MERS, and COVID-19. They discovered that the terms "infection," "patient," and "case" are used the most frequently in the documents under investigation (12). Doanvo, Qian, Ramjee, Piontkivska, Desai & Majumder studied the publishing features of COVID-19 in CORD-19 (COVID-19 Open Research Dataset) and analyzed them by applying topic modeling. The results of their study identified eight main topics for COVID-19 articles and indicated that researchers should conduct further investigations on diagnostics, therapeutics, vaccines, viral genomics, and pathogenesis (13). Han, Wang, Zhang, Wang utilized a text-mining method to examine the public opinions and sentiments of Chinese citizens regarding COVID-19 in social media. On COVID-19, seven significant topics and thirteen subtopics were found. The Weibo spatial distribution of COVID-19- relevant studies were mainly concentrated in Wuhan, Beijing-Tianjin-Hebei, the Yangtze River Delta, the Pearl River Delta, and the Chengdu-Chongqing urban agglomeration. The most expressed topics were related to "opinion and sentiments" (14).

ChireSaire and Pineda-Briseo used social network analysis and text-mining techniques to investigate the COVID-19 topic in Latin America. The findings demonstrated that the South American scientific community is more interested than those in Central and North America in using artificial intelligence to track the spread of the COVID-19 pandemic (15). "COVID-19," "public health," "world health," and "new coronavirus" were also thought to have a higher frequency and weight than other keywords, according to Abdeen, Hamed, and Wu. Their findings also demonstrated how misinformation spread within a reputable news outlet such as the "Daily Mail". The previous studies emphasize the importance of using text-mining techniques in analyzing and classifying a considerable amount of texts, and in particular, no study compared and analyzed the words of relevant COVID-19 documents in the world and Iran. Therefore, the urgency of writing and publishing this article was

felt even more due to the importance of the research problem and the lack of a similar study **(16, 17)**.

Danesh, Dastani, and Ghorbani aim to model global coronavirus articles based on topics in the last 50 years. The findings indicated that the most important keywords in global coronavirus articles are SARS, science, protein, MERS, veterinary, cell, human, RNA, medicine, and virology. Also, eight important topics were identified in the global coronavirus articles **(18)**.

A review of the research literature showed that various types of research had been conducted in medical sciences using the text-mining method. However, a limited number of researches on coronavirus, Sars, MERS, and Covid-19 have been conducted using text mining. One of the most important gaps and weaknesses of previous research compared to the current research is the limited number of the research community. In this research, more than 160 thousand documents have been analyzed and reviewed, which the value of the present study is.

2. Materials and Methods

This applied research was conducted by employing text-mining techniques with an analytical approach. The statistical population consisted of all the COVID-19 articles indexed in the PubMed Central® (PMC) database from November 2019 to July 2021. This database is among the most extensive and reliable databases of medical sciences in the world in terms of coverage and reliability **(5)**. PubMed Central® (PMC) was searched on June 10, 2021, to retrieve the data from international and Iranian articles on COVID-19. The data from international and Iranian articles were extracted distinctly. The data was retrieved in Medline file format and then converted to CSV using Science Space (http://medialab.github.io/sciencescape/medline_utils) to perform text-mining analysis. The number of records extracted from the data of international and Iranian articles on COVID-19 were 157,719 and 3143, respectively. For each record of COVID-19 articles, the title and abstract were applied to perform the text-mining process. Text-mining techniques were employed to analyze the data of this study, which included three steps as follows: "Text pre-processing," "text analysis," and "result interpretation" **(19)**.

Text mining is analyzing and exploring a mass of unstructured texts by software to identify concepts, patterns, topics, keywords, and other features of textual data. In other words, text mining aims to discover meaning (concept and purpose) and extract hidden information (for example, entities and relationships) in textual data. Text mining was done by the python programming language, an open source, with simple syntax, compact and multi-purpose. It is also easily developed and provides users with various libraries for working with texts **(20)**.
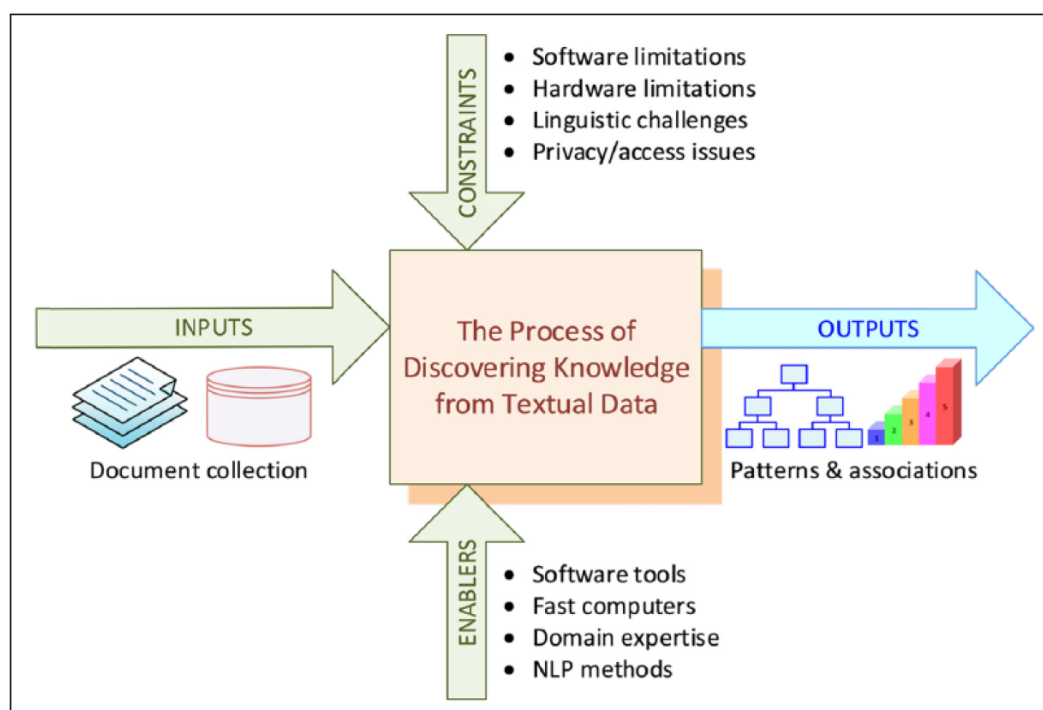

Figure 1. Text-mining process in the form of a general conceptual model **(21)**

In order to begin the text-mining process, the title, abstract, and keywords from each retrieved article were merged. Pre-processing and data cleaning operations were then applied to the study data to improve the data quality and reliability of the extracted patterns and relationships. The pre-processing operation was carried out in the manner described as follows: (i) Removing non-essential characters such as extra spaces, text formatting tags, and non-alphabetic characters from the text; (ii) Fractioning text components into words; (iii) Converting uppercase letters to lowercase letters for text uniformity; (iv) Homogenizing synonyms to the word preferred; and (v) Homogenizing different forms of words using the lemmatization method, which includes replacing words or their basic or dictionary form instead of used forms of words (22); and (vi) removing stop words and words that have no value for retrieving or analyzing documents, such as conjunctions and prepositions (and, the, of, for) that have little informational content. According to topic modeling performed on the data (23) based on the weight assigned to each topic, the emergence of each topic was determined among national and international articles during the study period. Finally, the topics with higher growth were selected as emerging topics.

The weighted average of topics within each topic was calculated in order to identify the emerging topics based on the topics derived from topic modeling, and it was plotted over the study period from November 2019 to June 2021 (20 months) in distinct graphs based on seven topics at the international and national levels (Iran) for COVID-19 articles. The emerging topics were compared and identified based on the weighted average change of topics over time. It should be noted that in LDA topic modeling, a factor is assigned to each document in each topic based on how similar it is to each topic. The same topic is then provided in this document in each of the topics with the highest factor. As a result, each document within the topics has a weight factor. The average factors of documents from each topic and year were gathered for this study. Then, the annual trend of each topic's factors over the years is demonstrated.

3. Results

3.1. Most Important Words Used in International and National COVID-19 Articles

The ten important and most frequently used words, based on TF-IDF weighting, in global COVID-19 literature are presented in Table 1. The data from international literature indicate that "COVID" (W=132.45), "infect" (W=94.81), and "cell" (W=89.50) are the most significant words used in these articles. In addition, the data analysis of the most significant keywords applied in the national COVID-19 articles demonstrated that "patient" (W=31.77), "SARS-COV" (W=22.52), and "COVID" (W=19.36) are the most important words employed in these articles (Table 1).

Moreover, Figure 2 indicates the most important words used in the international and national COVID-19 articles as a word cloud.

3.2. Trend of Words Change in International and National COVID-19Articles

It was impossible to perform weighting based on TF-IDF and analyze the trend of words change due to the single-digit number of national articles in November and December 2019. Therefore, the most significant words of international articles were studied in the mentioned months. In the following, the national and international COVID-19 articles were analyzed for 2020 and the first six months of 2021, and the trend of words change was extracted monthly and reported on an annual basis after aggregation.

Table 1. Most significant words employed in international and national COVID-19 articles.

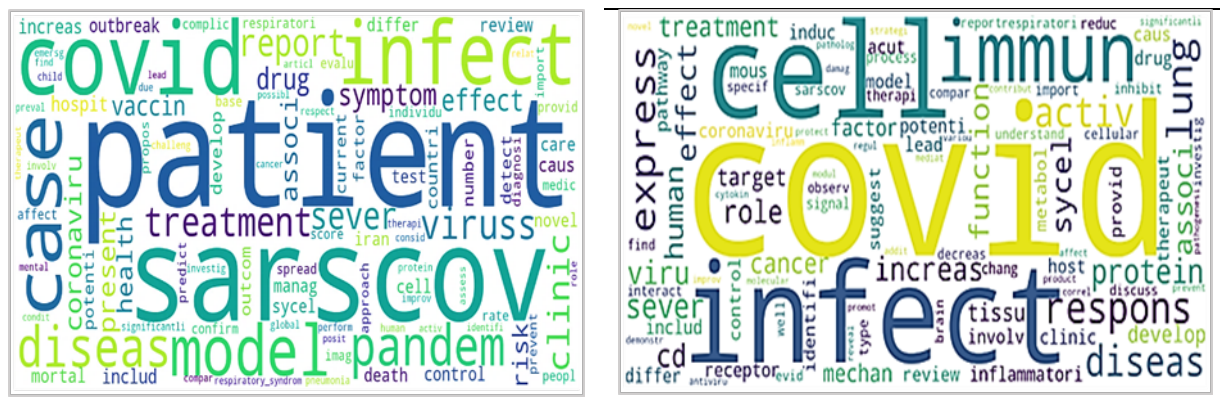| Rank	Keywords of international articles	TF-IDF	Keywords of national articles	TF-IDF
1	Covid	132.45	Patient	31.77
2	Infect	94.81	SARS-Cov	22.52
3	Cell	89.50	COVID	19.36
4	Immune	85.65	Infect	17.75
5	Response	84.22	Case	16.77
6	Express	82.21	Model	16.21
7	Lung	81.00	Pandemic	16.06
8	Disease	75.29	Disease	15.87
9	Active	74.98	Virus	15.83
10	Protein	73.04	Treatment	15.21

Figure 2. Word cloud of the most significant words used in international (left) and national (right) COVID-19 articles
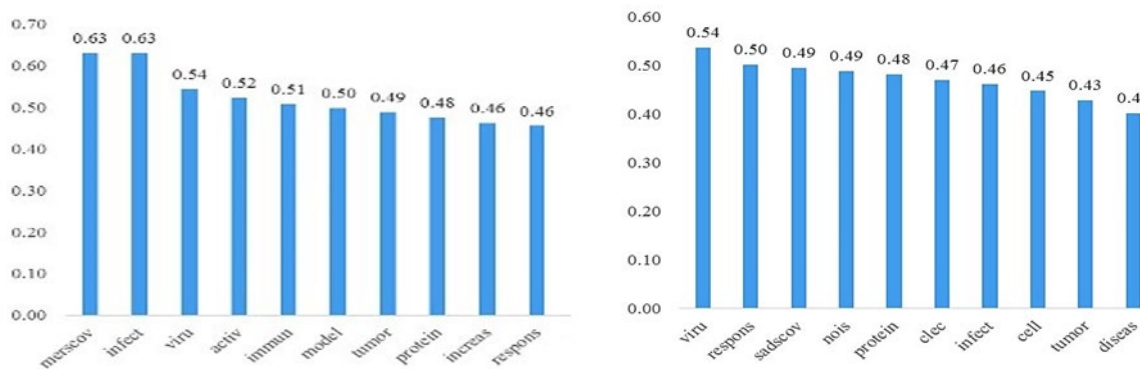

Figure 3. Most significant words used in international COVID-19 articles in November (left) and December (right) 2019

3.3. Trend of Words Change in International COVID-19 Articles in November and December 2019

According to Figure 3, "MERS-CoV" (W=0.63) and "infect" (W=0.63), "viru" (W=0.54), and "active" (W=0.52) were the most significant words employed in the international COVID-19 articles in November 2019 (left) and "viru" (W=0.54), "respons" (W=0.50), and "sarscov" (W=0.49) and "noise" (W=0.49) were the most significant words employed in the international COVID-19 articles in December 2019 (right).

Figure 4. Word cloud of international COVID-19 articles in November (left) and December (right) 2019.

3.4. Trend of Words Change in International and National COVID-19 Articles in 2020

The results of the data analysis were placed next to each other in Figure 5 and Figure 6 to study and compare the trend of words changes in the most important words employed in international and national COVID-19 articles in 2020. In other words, the diagram and word cloud related to international literature is placed on the right side and indicated in brown. While the blue diagram and word cloud on the left is related to national articles, indicating the most significant words employed in them.
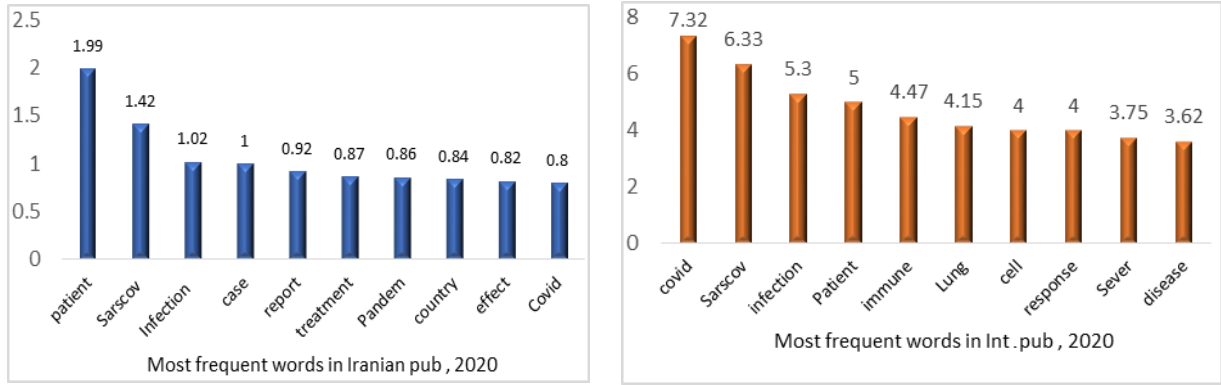
Figure 5. Most significant words employed in international (right) and national (left) COVID-19 articles in 2020

The most important words, based on TF-IDF weighting, used in international and national COVID-19 articles in 2020 are presented in Figure 5. "COVID" (W=7.32), "SARS-CoV" (W=6.33), and "infection" (W=5.3) were the most significant words used in international COVID-19articles in 2020, while the words "Patient" (W=1.99), "SARS-CoV" (W=1.42), and "infection" (=1.02) were the most significant words employed in the national COVID-19 articles in 2020.
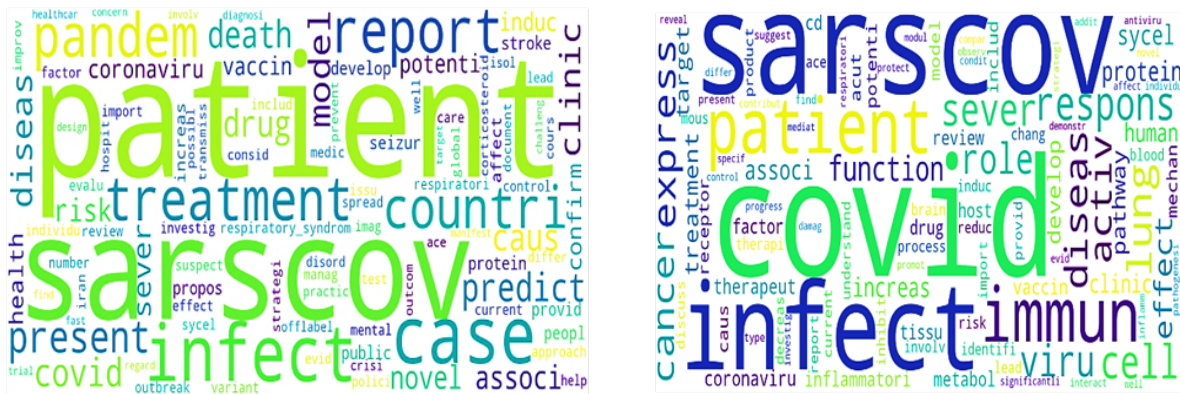

Figure 6. Words used in international (right) and national (left) COVID-19 articles in 2020.

The trend of word change of the most significant words used in the international and national COVID-19 articles in 2021 is presented in Figure 7 in graphs and a word cloud.
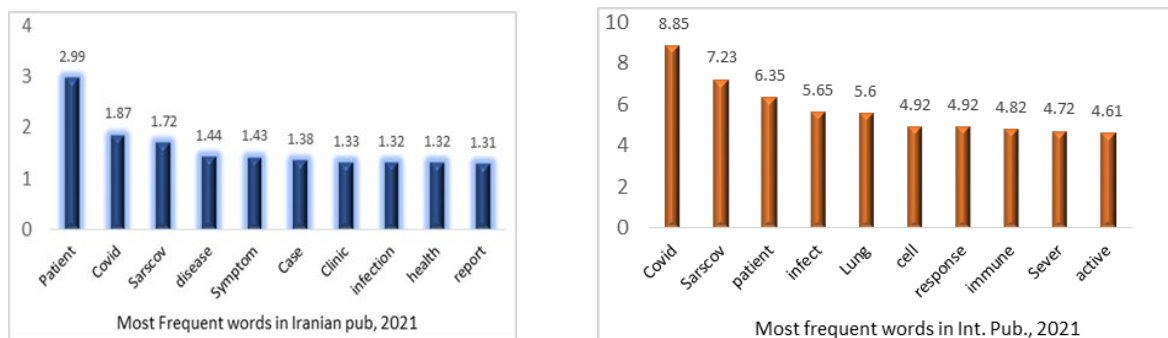

Figure 7. Most significant words used in international (right) and national (left) COVID-19 articles in 2021.

The most significant words, based on TF-IDF weighting, used in international and national COVID-19 articles in 2021 are presented in Figure 7. "Covid" (W=8.85), "Sarscov" (W=7.23), and "Patient" (W=6.35) were the most significant words used in the international COVID-19articles in the first six months of 2021. In addition, "Patient" (W=2.99), "Covid" (W=1.87), and "Sarscov" (W=1.72) were the most significant words employed in the national COVID-19 literature in the first six months of 2021.
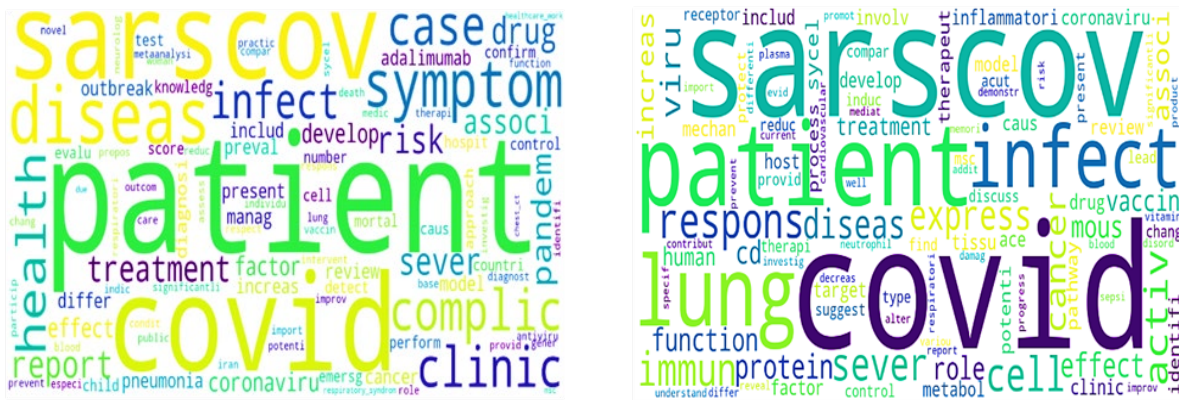
Figure. 8. Word cloud of international (right) and national (left) COVID-19 articles in 2021.

3.5. Emerging Topics in International and National COVID-19Articles

The trend of international and national articles in the seven topics derived from the implementation of LDA was further studied to answer this question and determine the emerging topics. Emerging topics in the international and national articles are presented in Table 2.

Table 2. Emerging topics in the international and national literature of the subject area.

Row	International Subject	Emerging Subject	Row	National (Iran) Subject	Emerging Subject
1	Diagnostic Tests	No	1	COVID-19 and molecular structure	No
2	COVID Proteins: Vaccine and Antibody Response	No	2	Social and Mental Status	Yes
3	Vaccine Immunogenicity	Yes	3	Treatment	No
4	Other	No	4	Other	No
5	Social and technology in COVID-19	Yes	5	Application of artificial intelligence	No
6	Covid Complication	No	6	Clinical features	No
7	Covid & immune system	No	7	Vaccine	No

The weighted average changes of the international COVID-19 articles in each of the seven subject areas were demonstrated in Table 2, indicating the descending and ascending trends of topics. Among them, two "Vaccine Immunogenicity" and "Social and technology in COVID-19" topics were the emerging topics of international literature on COVID-19.
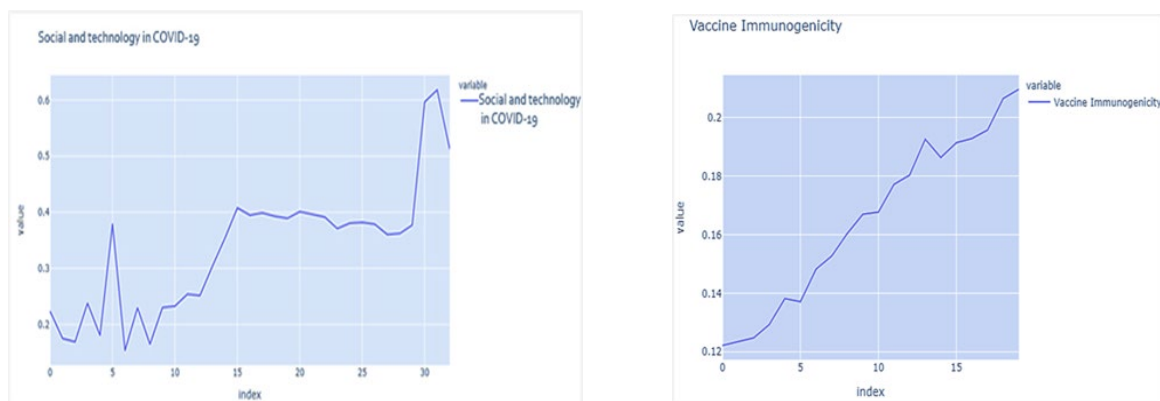

Figure 9. Change trend of the average weight (%) of the emerging topics of "Vaccine Immunogenicity" and "Social and technology in COVID-19" in international articles

At the national level, the results suggested that only "Social and mental status in COVID-19" is among the emerging topics of Iranian articles in this field.

Figure. 10. Change trend of the monthly average weight (%) of the national articles on COVID-19 related to "Social and Mental Status"

4. Discussion

Based on TF-IDF weighting and the results of this study, the most significant words used in national and international COVID-19 literature were suggested until June 2021. The most significant and frequently used words in the international COVID-19 literature were "response," "immune," "cell," "infect," and "Covid," whereas the most considerable and commonly used words in the Iranian COVID-19 literature were "case," "infect," "Covid," "SARS-CoV," and "Patient." Thus, "infect," and "Covid" were commonly used in both international and Iranian articles, according to the results commonality. Moreover, to solve the problem of COVID-19, the wide range of this dangerous pandemic's effects and its impact on various fields and sub-fields of medical sciences, such as immunology and microbiology, were indicated by the thematic diversity of frequently used words.

The significant and common words in the literature on the COVID-19 field introduced by Dornick, Kumar, Seidenberger, Seidle which were "patient," "infect," "disease," "case," and "pandem" **(24)**, by Haghani et al. were "human*," "Coronavirus infection*," and "viral pneumonia" **(25)**, by Doanvo, Qian, Ramjee, Piontkivska, Desai & Majumder were "case," "coronavirus," "pandemic," "patient," and "Covid" **(11)**, and by Cheng, Cao & Liao were "study", "infection," "number," "case," and "patient" **(10)**. Meanwhile, Gupta et al., while identifying the most significant keywords of COVID-19 literature, expected "mask" and "personal protection equipment" to be the most frequent keywords, but the results did not confirm this issue **(26)**. The words and keywords employed in the COVID-19 articles indicated the complexity and breadth of this scientific area. They suggested the inclusion of various fields in the COVID-19 literature **(27)**.

The most important COVID-19 words have changed over time in national and international literature, and this trend has coincided with significant scientific events. According to results pertaining to the twenty-month trend of international literature, "COVID" came first in the fourteen months between the first of 2020 and June 2021. This is true even though "patient" was used more frequently in Iranian articles during eighteen months instead of twenty months. In contrast to international studies, which have focused on COVID-19 and the infections it causes, Iranian researchers have concentrated their studies first on hospitalized patients before moving on to the virus itself. Chire Saire and Pineda-Briseo also looked into the trend of words change in Latin American regions, and their findings are consistent with the findings of this study **(15)**. Gupta et al. studied the trend of word change of COVID-19 weekly **(26)**.

In the international literature on COVID-19, "Vaccine Immunogenicity" is an emerging topic that has significantly expanded since the pandemic's start in the first half of 2021, according to the current study's findings. An essential component of a vaccine is its ability to induce an immune response against the virus and even reduce public anxiety **(14)**. The stronger the vaccine's immunogenicity, the more efficient and effective it will be when discussing its ability to build individual and collective immunity. This developing field of study has helped scientists better control the pandemic and is regarded as the cornerstone of research on the COVID-19 vaccine.

The impact of applying new medical technologies in clinical trials and the mass production of vaccines is the objective manifestation of the importance and attention to this issue, according to "society and technology in COVID-19," another emerging topic at the international level. The first COVID-19 vaccine's production stages were successfully finished within a few weeks thanks to new medical software and hardware, and the vaccine was then put on the market **(28)**. Modern medical tools and technologies have significantly assisted in the treatment of patients with COVID-19, benefiting not only the research and production sectors but also the treatment, healthcare, and hospital sectors.

The weighted average increase to more than 20% in the first half of 2021 at the level of Iranian literature resulted in the topic of "Social and Mental Status" emerging. It may be possible to list the psychological effects and harms at the macro and micro levels in society and families, including the rise in conflicts and arguments between parents and children as well as the couple themselves during the quarantine, among the reasons for the significance of this issue and the publication of numerous articles in this area. The degree of acceptance and prevention—or non-

acceptance and denial—in the social context and the culture of various societies in dealing with this disease are among the topics covered in the articles **(29)**. The results and analyses presented at this time may no longer be novel due to the rapid percentage of scientific production on COVID-19 in the national and mainly international venues, given that the study data were collected in June 2021.

Limitations & Suggestions

As evidenced by the more than 200,000 references that have been added to the PMC database in the past year, the field of COVID-19 has a high publication speed and a high rate of scientific production. This makes it more required to repeat this research in different periods.

As a result, it is advised to conduct a similar independent study between 2019 and the end of 2022 and compare the findings. Future research is recommended to include a thorough analysis using methods such as sentiment analysis and opinion mining based on data from WOSCC and Scopus.

5. Conclusion

In general, it can be concluded that considering the widespread contagion of COVID-19 in the world and extensive studies in the world to find treatment methods or to discover new vaccines and medicines to deal with this virus, awareness, and knowledge of the results of such studies in the strategic subject area, e.g., the COVID-19 pandemic, can be useful and practical for researchers and policymakers in the field of health and treatment to make better decisions and provide a

knowledge map and studies in the field of this virus and can guide them in providing effective solutions and optimal decision-making. Studies genuinely demonstrate that developed nations provide a greater number of global studies. Since those nations have better and more modern infrastructures for the production of vaccines, drugs, and molecular medicines, they concentrate their research according to the time of disease emergence with an eye toward "problem-solving," which is "controlling the COVID-19 pandemic through the development of collective immunity by vaccines" **(30)**. Most of their studies are categorized under the main topic of "vaccination." Iranian scientists, however, have focused their research on COVID-19 patients and the complications caused by the disease.

Conflict of Interest

The authors declare that they have no conflict of interest.

Reference

1. Cui X, Wang P, Wei Z. Emergency use of COVID-19 vaccines recommended by the World Health Organization (WHO) as of June 2021. Drug Discov Ther. 2021;15(4):222-4. [DOI:10.5582/ddt.2021.01064] [PMID]

2. Saeidnia H, Mohammadzadeh Z, Saeidnia M, Mahmoodzadeh A, Ghorbani N, Hasanzadeh M. Identifying Requirements of a Self-care System on Smartphones for Preventing Coronavirus Disease 2019 (COVID-19). Iran J Med Microbiol. 2020; 14(3):241-51. [DOI:10.30699/ijmm.14.3.241]

3. Danesh F, Ghavidel S. Coronavirus: Scientometrics of 50 Years of Global Scientific Productions. Iran J Med Microbiol. 2020;14(1):1-16. [DOI:10.30699/ijmm.14.1.1]

4. Shehata A, El Dakar M, Salem N, editors. Top COVID-19 100 vaccine papers: An Altmetric study. 2021 22nd International Arab Conference on Information Technology (ACIT); 2021. [PMCID] [PMID] [DOI:10.1109/ACIT53391.2021.9677241]

5. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J. 2008;22(2):338-42. [DOI:10.1096/fj.07-9492LSF] [PMID]

6. Cheng X, Shuai C, Liu J, Wang J, Liu Y, Li W, et al. Topic modelling of ecology, environment and poverty nexus: An integrated framework. Agric Ecosyst Environ. 2018;267:1-14. [DOI:10.1016/j.agee.2018.07.022]

7. Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed Tools Appl. 2019;78(11):15169-211. [DOI:10.1007/s11042-018-6894-4]

8. O'Callaghan D, Greene D, Carthy J, Cunningham P. An analysis of the coherence of descriptors in topic modeling. Expert Syst Appl. 2015;42(13):5645-57. [DOI:10.1016/j.eswa.2015.02.055]

9. Steyvers M, Griffiths T. Probabilistic topic models. Handbook of latent semantic analysis. 15 ed: Psychology Press; 2007. p. 439-60.

10. Meskarpour Amiri M, Nasiri T, Mehdizadeh P. Subjects clustering analysis and science mapping on COVID-19 researches in scopus database. J Mil Med. 2020;22(6):663-9.

11. Tran BX, Ha GH, Nguyen LH, Vu GT, Hoang MT, Le HT, et al. Studies of novel coronavirus disease 19 (COVID-19) pandemic: a global analysis of literature. Int J Environ Res Public Health. 2020; 17(11):4095. [DOI:10.3390/ijerph17114095] [PMID] [PMCID]

12. Cheng X, Cao Q, Liao SS. An overview of literature on COVID-19, MERS and SARS: Using text mining and latent Dirichlet allocation. J Inf Sci. 2022; 48(3):304-20. [DOI:10.1177/0165551520954674] [PMCID]

13. Doanvo A, Qian X, Ramjee D, Piontkivska H, Desai A, Majumder M. Machine Learning Maps Research Needs in COVID-19 Literature. Patterns. 2020;1(9):100123. [PMID] [PMCID] [DOI:10.1016/j.patter.2020.100123]

14. Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. Int J Environ Res Public Health. 2020;17(8):2788. [DOI:10.3390/ijerph17082788] [PMID] [PMCID]

15. Chire Saire J, Pineda-Briseño A. Text Mining for Covid-19 Analysis in Latin America. Artif Intell. 2021;2021:257-94. [DOI:10.1007/978-3-030-69744-0_16]

16. Abdeen MAR, Hamed AA, Wu X. Fighting the COVID-19 Infodemic in News Articles and False Publications: The NeoNet Text Classifier, a Supervised Machine Learning Algorithm. Appl Sci. 2021;11(16):7265. [DOI:10.3390/app11167265]

17. Jafari Baghiabadi S, Razieh F. Studying of Research Related to COVID-19 Vaccine in Iran and the World: A Thematic Analysis and Scientific Collaborations. Iran J Med Microbiol. 2021;15(4): 414-57. [DOI:10.30699/ijmm.15.4.414]

18. Danesh F, Dastani M, Ghorbani M. Retrospective and prospective approaches of coronavirus publications in the last half-century: a Latent Dirichlet allocation analysis. Libr Hi Tech. 2021;39(3):855-72. [DOI:10.1108/LHT-09-2020-0216]

19. Žižka J, Dařena F, Svoboda A. Text mining with machine learning: principles and techniques: Crc Press; 2019. [DOI:10.1201/9780429469275]

20. Rehurek R, Sojka P. Software framework for topic modelling with large corpora. Proceedings of the LREC workshop on new challenges for NLP frameworks: Citeseer; 2010.

21. Kim YM, Delen D. Medical informatics research trend analysis: A text mining approach. Health Informatics J. 2018;24(4):432-52. [DOI:10.1177/1460458216678443] [PMID]

22. Malaterre C, Lareau F, Pulizzotto D, St-Onge J. Eight journals over eight decades: a computational topic-modeling approach to contemporary philosophy of science. Synthese. 2021;199(1):2883-923. [DOI:10.1007/s11229-020-02915-6]

23. Danesh F, Dastani M. Application of Artificial Intelligence for Discovering the Subject Classes of National Publications Based on International Publications: COVID-19 Publications. Digit Health. under review; 2023.

24. Dornick C, Kumar A. Seidenberger S, Seidle E, Mukherjee P. Analysis of Patterns and Trends in COVID-19 Research. Procedia Comput Sci; 2021 [DOI:10.1016/j.procs.2021.05.032]

25. Haghani M, Bliemer MCJ, Goerlandt F, Li J. The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review. Saf Sci. 2020;129:104806. [PMCID] [DOI:10.1016/j.ssci.2020.104806] [PMID]

26. Gupta A, Aeron S, Agrawal A, Gupta H. Trends in COVID-19 publications: streamlining research using NLP and LDA. Front digit health. 2021;3:686720. [PMID] [PMCID] [DOI:10.3389/fdgth.2021.686720]

27. Hossain MM. Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping. F1000Research. 2020(9):374. [DOI:10.12688/f1000research.23690.1]

28. Mullard A. COVID-19 vaccine development pipeline gears up. Lancet. 2020;395(10239):1751-2. [DOI:10.1016/S0140-6736(20)31252-6] [PMID]

29. Farahati M. Psychological impacts of the spread of coronavirus in society. Soc impact assess. 2020(2):207-5.

30. Haleem A, Javaid M, Vaishya R, Deshmukh SG. Areas of academic research with the impact of COVID-19. Am J Emerg Med. 2020;38(7):1524-6. [DOI:10.1016/j.ajem.2020.04.022] [PMID] [PMCID]