International Journal of Information Management Data Insights 1 (2021) 100016

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Investigating diseases and chemicals in COVID-19 literature with text mining

Amir Karami[a,*], Brandon Bookstaver[b], Melissa Nolan[c], Parisa Bozorgi[c,d]

[a] School of Information Science, University of South Carolina, Davis College, 1501 Greene St, Columbia, SC 29208, United States
[b] College of Pharmacy, University of South Carolina, United States
[c] Arnold School of Public Health, University of South Carolina, United States
[d] South Carolina Department of Health and Environmental Control, United States

ARTICLE INFO

Keywords:
Chemical
Drug
Disease
Symptom
Text mining
SARS-CoV-2
COVID-19
Literature

ABSTRACT

Given the rapidly unfolding nature of the COVID-19 pandemic, there is an urgent need to streamline the literature synthesis of the growing scientific research to elucidate targeted solutions. Traditional systematic literature review studies have restrictions, including analyzing a limited number of papers, having various biases, being time-consuming and labor-intensive, focusing on a few topics, and lack of data-driven tools. This research has collected 9298 papers representing COVID-19 research published through May 5, 2020. We used frequency analysis to find highly frequent manifestations and therapeutic chemicals, representing the importance of the two biomedical concepts. This study also applied topic modeling that provided 25 categories showing associations between the two overarching categories. This study is beneficial to researchers for obtaining a macro-level picture of literature, to educators for knowing the scope of literature, and to policymakers and funding agencies for creating scientific strategic plans regarding COVID-19.

1. Introduction

An unprecedented outbreak of pneumonia of unknown etiology in Wuhan City, Hubei province in China, emerged in December 2019. In January 2020, the World Health Organization (WHO) declared the Chinese outbreak of COVID-19 to be a Public Health Emergency of International Concern, posing a high risk to countries with vulnerable health systems (WHO, 2020). In March 2020, the World Health Organization announced the COVID-19 pandemic. As of June 16, 2020, the number of positive cases was more than 8 million globally, with over 438,000 deaths (COVID-19 Map [Internet], 2020). These numbers continue to rise daily, and this pandemic has impacted nearly all aspects of life with schools closed, many businesses shuttered, travel curtailed, and major sports events canceled.

The WHO emergency committee has stated that the spread of COVID-19 may be interrupted by early detection, isolation, prompt treatment, and the implementation of a robust system to trace contacts. Other strategic objectives include a means of ascertaining clinical severity, the extent of transmission, and optimizing treatment options (Organization WH, 2020). The outbreak has posed significant threats to international health and the economy. In the absence of treatment for this virus, there is an urgent need to find alternative methods to find possible solutions from the current literature.

In March 2020, the White House Office of Science and Technology Policy issued a call to action to the Nation's artificial intelligence experts to utilize text and data mining techniques that can help the science community answer high-priority scientific questions related to COVID-19. This call to action addresses the fact that the vast majority of research scientists cannot afford to spend copious amounts of their time analyzing this growing literature dataset. Also, the traditional systematic literature review is a time-consuming and labor-intensive process. Therefore, computational methods can help experts rigorously find unbiased, statically meaningful patterns in COVID-19 literature. This study identifies and investigates clinical manifestations of disease and therapeutic chemical compounds in COVID-19 research papers and discloses the relationship between diseases and chemicals.

2. Literature review

To summarize research papers on a specific topic, researchers develop traditional surveys that involve finding, evaluating, and analyzing relevant publications (Kar & Navin, 2020). Several traditional literature review surveys were developed on the COVID-19 pandemic (Hatmi, 2021). The surveys have focused on some main issues, including symptoms (e.g., fever, Lovato & De Filippis, 2020), the coincidence of COVID-19 and other diseases (e.g., cardiovascular dis-

* Corresponding author.
E-mail address: karami@sc.edu (A. Karami).

https://doi.org/10.1016/j.jjimei.2021.100016
Received 28 February 2021; Received in revised form 10 May 2021; Accepted 11 May 2021

ease, Santoso et al., 2020), and treatments (e.g., using Remdesivir, Beigel et al., 2020; Fajgenbaum et al., 2020). While the traditional literature review studies and current tools offer valuable information. However, these studies suffer from some limitations:

- First, researchers select a sample of relevant papers, not all related papers. For example, one recent study reviewed less than 150 research papers related to COVID-19 (Zhang & Liu, 2020).
- Second, selecting a sample of relevant research by traditional approaches could be prone to various biases, such as focusing on specific journal articles or highly cited studies.
- Third, the traditional literature review process is a time-consuming and labor-intensive one.
- Fourth, current literature reviews are often topic-specific, limited to a few diseases or chemicals.
- Fifth, replicating results generated by previous review studies is a difficult task.
- Sixth, while research search engines (e.g., PubMed) provide access to research papers, they do not offer a data-driven tool or propose a limited analysis (e.g., LitCovid).

We developed a query from the PubMed website (https://pubmed.ncbi.nlm.nih.gov/advanced/) and found that more than 31 million documents were published between 1781 and 2020 on PubMed, that more than 127,000 of those documents contain "COVID-19." This huge amount of text data shows that there is a need to explore large collections of research databases. Big data studies are associated with a wide range of domains, from academic libraries (Al-Barashdi & Al-Karousi, 2018; Witten, Don, Dewsnip, & Tablan, 2004; Zhang & Gu, 2011) to health recommender systems (Valdez, Ziefle, Verbert, Felfernig, & Holzinger, 2016; Wiesner & Pfeifer, 2014). Big data is also a key component in biomedical research applications (van Altena, Moerland, Zwinderman, & Olabarriaga, 2016). Big data mining methods have been very effective and efficient in automatically identifying patterns from big biomedical data. For example, text mining methods have been utilized for new biomedical discoveries (Mirza et al., 2019). Text mining includes "*the methods of machine learning and statistics with the goal of recognizing patterns and disclosing the hidden information in text data*" (Karami, Gangopadhyay, Zhou, & Kharrazi, 2018). Text mining methods aim to obtain information from documents, discover novel patterns, and identify relationships between concepts (Gonzalez, Tahsin, Goodale, Greene, & Greene, 2016).

To address the limitations of traditional surveys, big data and text mining methods can assist in investigating large sets of written word data in an efficient and effective way (Grover & Kar, 2017). Text mining methods have been utilized for literature review surveys relevant to depressive disorder (Zhu et al., 2018), wearable technology (Shin et al., 2019), biomedical (Chen et al., 2017; van Altena et al., 2016), big data (Mohammadi & Karami, 2020 Aug 24), medical case reports (Karami, Ghasemi, Sen, Moraes, & Shah, 2019), services management and marketing (Verma, Sharma, Deb, & Maitra, 2021). However, text mining has not been used for analyzing infectious diseases literature. This research provides a fast data-driven framework to analyze COVID-19 related papers to overcome the limitations of traditional literature review studies.

3. Methods

In this section, the construction of our dataset is described, followed by frequency analysis and relationship analysis (Fig. 1). We also provide more details on the application of topic modeling for this research.

3.1. Data

We used the COVID-19 article collection of LitCovid through May 5, 2020 (Index of /pub/lu/LitCovid [Internet], 2020). This collection was annotated with six entity types, including gene, disease, chemical, mutation, species, and cell line. While we focused on chemical and disease concepts in this paper, future research could utilize the rest of the entities to provide a better understanding of the literature. The rest of our analysis is based on the identified chemical and disease concepts, not the whole text of the COVID-19 articles.

3.2. Frequency analysis

Frequency analysis plays an important role in text mining. This analysis is based on measuring the frequency of chemical and disease concepts. This analysis shows the number of papers containing a disease or a chemical. This is a binary process to find whether a disease or a chemical was used in a paper. If a disease or a chemical occurred once or multiple times in a paper, it would be 1. Otherwise, it would be 0. Frequency analysis has been used for different applications such as opinion mining (Heimerl, Lohmann, Lange, & Ertl, 2014) and content analysis (Karami, Lundy, Webb, & Dwivedi, 2020; Karami, White, Ford, Swan, & Spinel, 2020). Frequency analysis starts with splitting a document into a sequence of tokens by space between words (terms). To apply frequency analysis to our data, we utilized the unnest_tokens function in the R tidytext package (Silge & Robinson, 2016) and the count function in the plyr R package (Wickham, 2011) to extract tokens and find the total frequency.

3.3. Relationship analysis

Next, we turned our attention to find relationships between our two overarching categories. Co-occurrence analysis, clustering, and topic modeling can assist in identifying relationships between entities; however, we decided to use topic modeling because current literature illustrates better performance for topic modeling than co-occurrence analysis (Leydesdorff & Nerghes, 2017). In addition, topic modeling can disclose the relationship between documents and clusters (categories) (Karami, 2015). Topic models are a class of hidden variable models, structured distributions in which observed data interact with hidden random variables. With a hidden-variable model, the practitioner can posit a hidden structure in the observed data, and then learn that structure using posterior probabilistic inference.

Among different topic models, latent Dirichlet allocation (LDA) is a valid and popular model. LDA has been used for different applications (Kumar et al., 2021) such as analyzing online reviews (Karami & Pendergraft, 2018), social media data (Karami & Anderson, 2020; Karami & Elkouri, 2019; Karami & Shaw, 2019; Karami et al., 2021b, 2021a; Webb, Karami, & Kitzie, 2018), biomedical literature (Shin et al., 2019), and neurology case reports (Karami et al., 2019); this application to COVID-19 literature is novel. In our data, LDA assumes that there is an exchange between articles (documents) and biomedical concepts (words). LDA can find categories representing a cluster of related words (terms).

The outputs of LDA for n documents (papers), m words, and t categories, are two matrices. The first one is the probability of each of the words in each category or $P(W_i|C_k)$, and the second one is the probability of each of the categories in each document or $P(C_k|D_j)$ (Karami et al., 2020):

$$\begin{array}{cc} & \text{Categories} \\ \text{Words} & \begin{bmatrix} P(W_1|C_1) & \cdots & P(W_1|C_t) \\ \vdots & \ddots & \vdots \\ P(W_m|C_1) & \cdots & P(W_m|C_t) \end{bmatrix} \end{array}$$

$$P\left(W_i|C_k\right)$$

$$\begin{array}{cc} & \text{Documents} \\ \text{Categories} & \begin{bmatrix} P(C_1|D_1) & \cdots & P(C_t|D_n) \\ \vdots & \ddots & \vdots \\ P(C_t|D_1) & \cdots & P(C_t|D_n) \end{bmatrix} \end{array}$$

$$P\left(C_k|D_j\right)$$

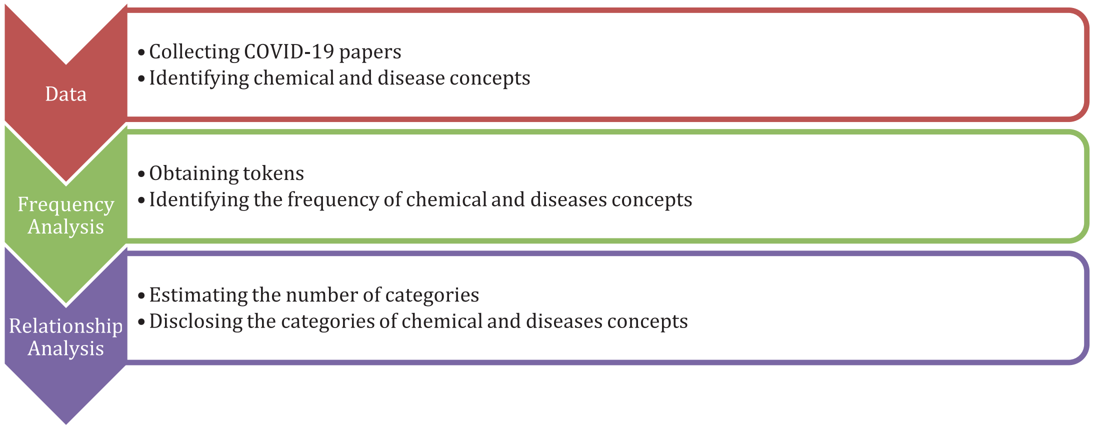

Fig. 1. Research framework.

Table 1
Statistics of diseases and chemicals.

	#Tokens	#Unique terms	Minimum frequency	Max frequency	Median
Chemicals	7612	2434	1	318	1
Diseases	33,838	3645	1	7100	1

The top words in each category based on the order of $P(W_i|C_k)$ represent the categories. To find the related documents (papers) for each topic, we sorted $P(C_k|D_j)$ from the highest value to the lowest one. The top documents have the highest probability of being related to a category. To disclose the relationship between diseases and chemicals, we used LDA. Before applying LDA, we needed to estimate the optimum number of clusters. We utilized five methods, of which four were developed in the R ldatuning package (Nikita & Nikita, 2019) and the fifth method was c_v developed in the Python gensim package (gensim: topic modelling for humans [Internet], 2020). While the maximum value of the three methods of the five estimation methods shows the optimum point, the minimum value of the two methods illustrates the optimum point. To combine the result of the five methods, we used the Technique of Order Preference Similarity to the Ideal Solution (TOPSIS), which is a multi-criteria decision analysis method (Yazdi, 2013) developed in the R topsis package (McCallum, 2002). The output of TOPSIS offered the optimum number of clusters at 25.

We measured the log-likelihood for five sets and found that the log-likelihood reached its approximate maximum before 4000 iterations. To assess the robustness of LDA, we compared the log-likelihood for five sets of 4000 iterations. The comparison indicated no significant difference (p-value > 0.05) between the iterations regarding their mean and standard deviation (Fig. 2). Then, we used the Mallet implementation of LDA (PubMed Search Results on COVID-19 [Internet], 2021) to cluster diseases and chemicals. We set the number of topics and iterations at 25 and 4000, respectively. Appendix A shows 25 clusters of disease and chemical terms. Each term was labeled as a chemical (e.g., remdesivir_chemical) or a disease (e.g., infection_disease). We provided one research example offered by LDA for each cluster containing some of the terms in the cluster. These examples can assist researchers in having a better understanding of the relationships.

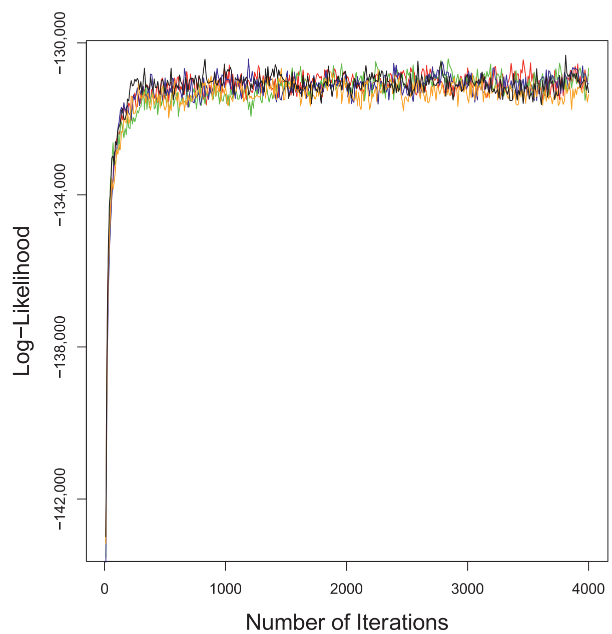

Fig. 2. Log-likelihood for five sets of 4000 iterations.

4. Results

The annotated collection with 9298 records was obtained on May 5, 2020. We found 3645 disease-related and 2434 chemical-related terminology in the investigation (Table 1).

Frequency analysis of disease-related terminology assisted in finding highly frequent symptoms and clinical aspects of diseases. We did not apply any pre-processing methods (e.g., stemming) to allow opportunities for all relationships to develop in an unbiased manner. Fig. 3 displays the top-50 list of disease-related terminology unearthed by our frequency analysis. Several overlapping clinical terms and iteratives were involved, including “SARS-CoV-2; COVID; coronavirus; COVID-19; coronavirus 19; novel coronavirus.” We removed the top four terms, including Covid-19, Infection, Coronavirus-2019, and Infected from Fig. 3. A few trends were noted by frequency analysis of disease-related variables. First, multiple search terms are related to COVID-19 disease (n=16), highlighting the confusion around naming this clinical disease and the inconsistency of search terms and/or literature references used. A few of these terms include the virus’s common name ‘SARS-CoV-2,’ generalized ‘coronavirus,’ a shortened disease name of ‘covid,’ the provisional name ‘2019 n-CoV,’ and others. Second, a similar trend of multiple annotations of pulmonary disease terminology was noted. In general, vague

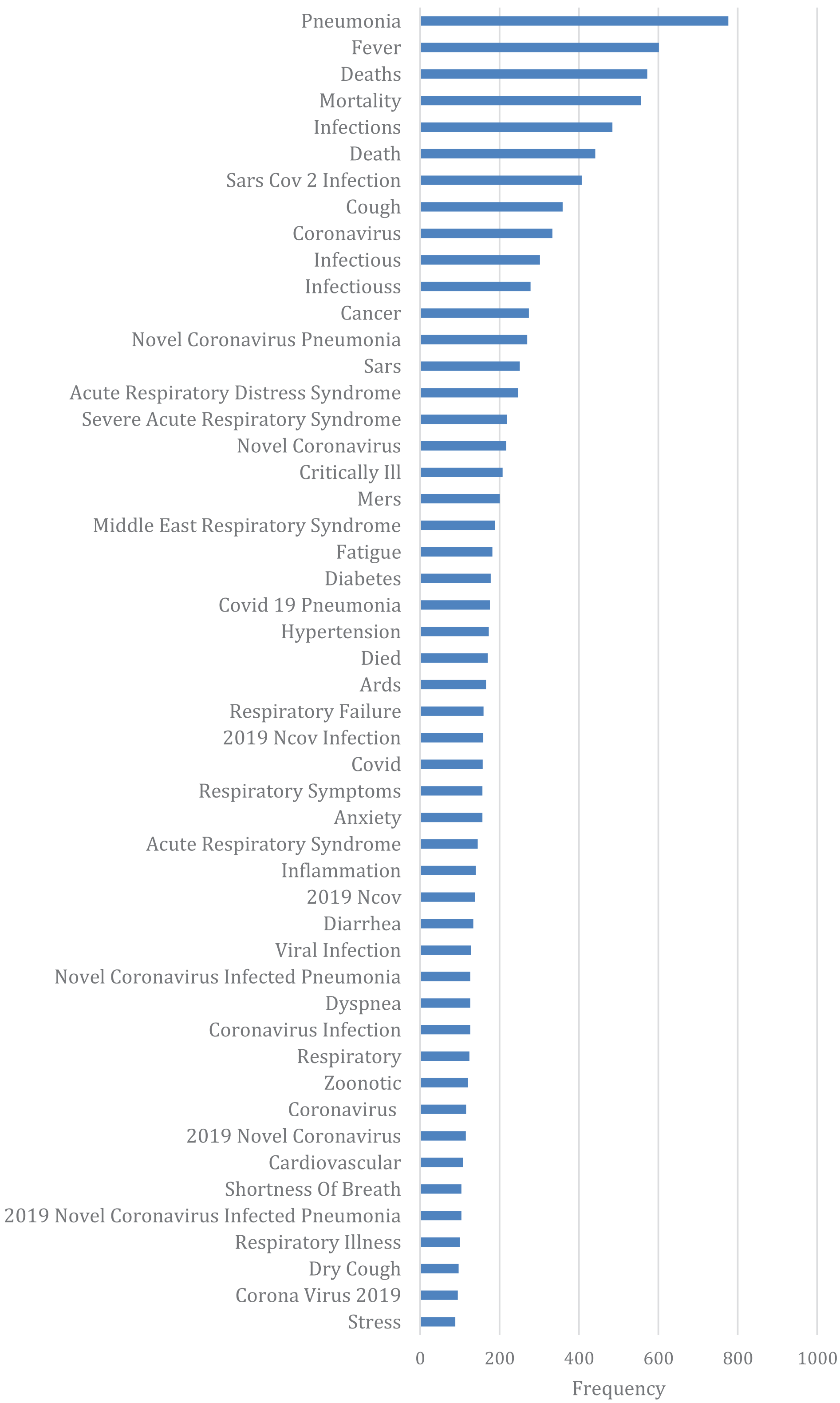

Fig. 3. Top-50 high frequent disease-related terminology identified via frequency analysis.

terms were used more frequently (infection, pneumonia, fever, death) compared to less commonly used detailed clinical terms (cardiovascular, inflammation, dyspnea). Infectious disease transmission route terminology was infrequently used (zoonotic, respiratory). Comorbidities and risk factors for severe diseases were moderately referenced (diabetes, hypertension). Interestingly, geographical terms (e.g., United States, China, and Italy) or population specific terms (e.g., elderly, nursing home, and cruiseship) were not identified in our frequency analysis. Using the same methodology, we executed a frequency analysis of therapeutic chemicals related to SARS-CoV-2 infection. As noted in Fig. 4, chemical compounds were notably less frequently utilized in the scientific literature than clinical manifestations of diseases.

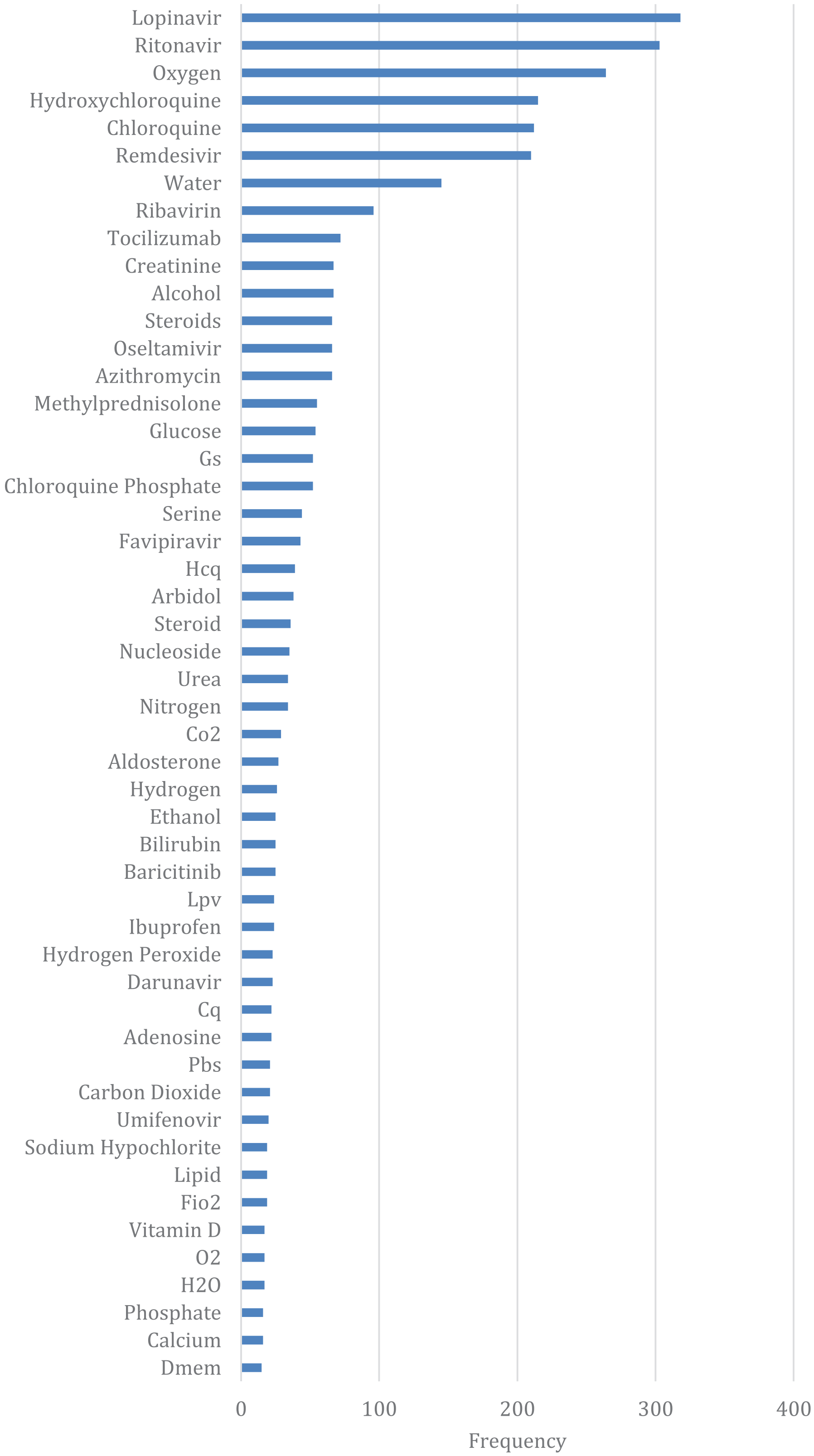

Fig. 4. Top-50 high frequent chemical-related terminology identified via frequency analysis.

Table 2
Summary of pertinent LDA relationship analysis.

Category	Disease or Symptom	Chemical Element or Drug	Category	Disease or Symptom	Chemical Element or Drug
C1	Infection; SARS; COVID19; Pneumonia; Inflammation	Serine; Oxygen	C14	Cough; Fever; Infection; Pneumonia; Deaths	Oxygen
C2	Coronavirus; Inflammation; Acute Respirator; Distress Syndrome; Coagulopathy; Disseminated Intravascular Coagulation; Sepsis	Tocilizumab; Vitamin D; Heparin	C15	Fever; Cough; Fatigue; Diarrhea; Pneumonia; Dry Cough; Myalgia; Shortness of Breath; Vomiting; Dyspnea	N/A
C3	Coronavirus; Pneumonia; Infection; Immunodeficiency	Lopinavir/Ritonavir; Steroid; LPV/R; EEA	C16	Acute Respiratory Distress Syndrome; ARDs; Pneumonia; Respiratory Failure; Septic Shock; Death	Oxygen
C4	Infection; Coronavirus; Coughing; Cross Infection	Water; Alcohol	C17	Infection; Pneumonia; Ncov Infection; Fever; Deaths	N/A
C5	COVID Coronavirus; Coronary Syndrome; Infarction	F FDG; Carbon Dioxide; Copper; LPV; TCM; Stainless Steel	C18	Infected; Pneumonia	Water; Penicillin; Streptomycin; Dmem; Carbon; PBS; SDs; Sialic Acid
C6	Mortality	Remdesivir; Lopinavir; Ribavirin; Ritonavir; Chloroquine; Favipiravir; Oseltamivir	C19	Coronavirus; Respiratory Syndrome; Pneumonia; Pleural Effusion; Hypoxia; Chest Tightness	Steroids
C7	Middle East Respiratory Syndrome; MERS; SARS; Severe Acute Respiratory Syndrome; Pneumonia; Infections; Death	N/A	C20	Inflammation; Viral Infection; Coronavirus Infection; Acute Respiratory Syndrome; Virus Infection; Toxicity; Fibrosis	N/A
C8	Malaria; Rheumatoid Arthritis; Rheumatics; SLE; Systemic Lupus Erythematosus; Autoimmunes	Hydroxychloroquine; Chloroquine; Chloroquine; Phosphate; Baricitinib	C21	Anxiety; Pain; Depression; Stress; Fatigue; Infectious; Psychiatric; Delirium	Melatonin
C9	Coronavirus; Infection; Anxious; SAR; CO; Infections	TCB	C22	SARS Coronavirus Infection; Cytotoxicity; Acute Respiratory Infections	Hydrogen; Serine; Oligonucleotides; Glycine; Asparagine
C10	Coronavirus; Deaths; Infection; Infection; Emphysema; Multi Organ Failure; Toxicity	Oxygen; Colchicine; Situazione	C23	Hypertension; Diabetes; Cardiovascular; Death; Respiratory Failure; Myocarditis; Infection	Aldosterone
C11	Diabetes; Hypertension; Diabetes Mellitus; Mortality; Novel Coronavirus; Obesity; Diabetic; Respiratory Infections	Glucose; Ibuprofen	C24	Coronavirus; QT Prolongation; QCT Prolongation	Hydroxychloroquine; Chloroquine; Azithromycin; HCQ; Tocilizumab; CQ; QCT
C12	Cancer; Tumor; Bleeding; Lung Cancer; Malignancies; Hemorrhage	Formalin; Dexamethasone	C25	Viral Infection; Infection; Virus Infection; Toxicity; HIV	Lipid; ATP
C13	Coronavirus; Acute Kidney Injury; Liver; Death	Nitrogen; Urea; Creatinine; Bilirubin			

Compounds already licensed for use in other clinical conditions were the most commonly correlated terms: Lopinavir/Ritonavir (HIV), Hydroxychloroquine (Malaria), Remdesivir (Hepatitis C and Ebola), and Tocilizumab (Rheumatoid Arthritis). The involvement of a few terms was hard to distinguish (e.g., oxygen, carbon dioxide, nitrogen) between their relevance to pharmaceutical clinical trial measurement outcomes versus being a spill-over artifact of respiratory-related clinical manifestations of diseases. Additionally, a limitation of the frequency analysis was the inability to distinguish the interactions between multiple drug compounds, such as Lopinavir–Ritonavir. The full list of chemicals and diseases is available at https://github.com/amir-karami/COVID-19-Chemicals-Diseases.

Table 2 provides a summary of our relationship analysis findings comparing disease-related and therapeutic chemical compound-related terminologies in Appendix A. The summarization of Appendix A is based on removing duplicate concepts and splitting each category into two parts: diseases or symptoms and chemical or drug. For example, category 1 in Appendix A contains the following terms: infection_disease, sars_cov infected_disease, infections_disease, pneumonia_disease, serine_chemical, inflammation_disease, oxygen_chemical, sars_cov_infection_disease, critically_ill_disease, and covid_disease. This category contains five terms related to disease or symptoms, including infection, SARS, COVID19, pneumonia, and inflammation, and two terms related to chemical or drug, including serine and oxygen.

Each row of Table 2 shows a category, including the relationships between diseases and chemicals. For example, the first row shows C1 illustrating the relationships between five diseases (symptoms), including infection disease, SARS, COVID-19, pneumonia, and inflammation, between two chemicals, including serine and oxygen, and between the five diseases and two chemicals. While most categories contain statistically relevant combinations of terminology from both categories, four (C7, C15, C17, and C20) combinations did not result in chemical-associated terminology. This is likely due to the higher usage of disease-related terminology in current scientific literature, suggesting that current publications predominate on clinical manifestations of disease with a lower focus on therapeutic options.

5. Discussion

In just a few months, the COVID-19 literature has exponentially grown to over thousands of scientific publications (Harvard Health Publishing. Treatments for COVID-19 [Internet], 2021). The vast majority of research scientists cannot afford to spend their time analyzing this dataset. Also, the traditional literature review survey is a time-consuming and labor-intensive process. Therefore, computational methods can help experts find patterns in this dataset. This paper utilized text mining methods to identify associations between diseases and chemicals in thousands of COVID-19 papers. Our approach provided a bird's eye view to understand the importance of diseases and chemicals for researchers using frequency analysis and disclose their possible relationship using topic modeling. The frequency of the two bioconcepts showed the most discussed chemicals and diseases. The list of diseases contained diseases, conditions, sequalae, and symptoms. The list of chemicals showed chemical elements and known drugs that have been utilized for non-COVID-19 diseases.

By April 2021, (MED1) remdesivir, (MED2) dexamethasone, (MED3) baricitinib, (MED4) anticoagulation drugs (e.g., heparin, MED5) fever, aches, and pain reducers (e.g., ibuprofen, MED6) Vitamin D, (MED7) tocilizumab, (MED8) monoclonal antibodies, (MED9) convalescent plasma, (MED10) bamlanivimab plus etesevimab (AIIa), and (MED11) casirivimab plus imdevimab (AIIa), were found to be helpful medications for hospitalized people (CDC Coronavirus Disease, 2019; NIH, 2021). While our data collection was developed in May 2020, our research identified seven (MEDs 1–7) out of the 11 medications.

On February 2021, the Center for Disease Control and Prevention (CDC) announced the following 11 COVID-19 symptoms: (S1) fever, (S2) cough, (S3) shortness of breath or difficulty breathing, (S4) fatigue, (S5) pains in muscle or body, (S6) headache, (S7) nausea or vomiting, (S8) diarrhea, (S9) new loss of taste or smell, (S10) sore throat, and (S11) congestion or runny nose (CDC, 2021). Out of the 11 symptoms, this research identified eight symptoms (S1–S8).

On March 2021, CDC also announced medical risk factors that can lead to severe illness from COVID-19, including (RF1) cancer, (RF2) kidney disease, (RF3) lung diseases, (RF4) diabetes, (RF5) heart conditions, (RF6) HIV, (RF7) liver disease, (RF8) overweight and obesity, (RF9) down syndrome, (RF10) Immunocompromised state, (RF11) pregnancy, (RF12) sickle cell disease, (RF13) dementia, (RF14) smoking, (RF15) blood stem cell transplant, (RF16) stroke, and (RF17) substance use disorders (Strollo & Pozzilli, 2020). Out of these 17 risk factors, eight factors (RF1-RF8) appeared in our findings.

5.1. Contributions to literature

The detected diseases and chemicals can assist researchers in finding relevant studies. The relationship between diseases and chemicals in a category (1) can be used as hypotheses for further investigation, (2) has the ability to help with long term effects from COVID-19, (3) assist with vaccine discovery and monitoring of vaccine-related effects, and (4) can show real word management patterns.

This study has some advantages over classic literature review research. First, while current research has analyzed a limited number of research papers, this study provided an analysis of thousands of papers. Second, traditional literature review studies utilized qualitative methods to develop a codebook based on a limited data sample of related papers. This approach might not capture all main patterns. However, this paper got the benefit of using machine learning methods without a codebook. Third, this study provided an efficient process to avoid the time-consuming and labor-intensive processes of traditional qualitative methods.

5.2. Implications for practice

With the rapid growth of biomedical research publications, there is a continuous need to systematically and efficiently analyze the publications. This study is beneficial to researchers for obtaining a macro-level picture of literature, to educators for knowing the scope of literature, to journals for exploring most discussed diseases and chemicals, and to policymakers and funding agencies for creating strategic science plans regarding COVID-19. This paper can also offer a new approach for academic libraries to facilitate identifying research trends and patterns of a large number of papers in not only COVID-19 but also other domains.

5.3. Limitations

While this study provides a new perspective on COVID-19 literature, it has some limitations. First, the data was collected from a single source. Second, this research focused on papers published in English. Third, we collected papers published by May 2020. Fourth, this study is limited to two biomedical concepts. Fifth, the categories of chemical and disease concepts do not show relationships or prescribed treatments. So, they should be seen as suggestions for medical practitioners. Sixth, we focused on the frequency and categories of diseases and chemicals. However, we did not consider the results of relevant studies. For example, we do not know whether the identified medicines were useful for COVID-19 treatment. Future work could address the limitations by collecting both English and non-English papers published during several years from multiple sources and analyzing other biomedical concepts such as genes.

6. Conclusion

The COVID-19 pandemic has impacted nearly every aspect of life and has posed significant threats to international health and the economy. In the absence of treatment for this virus, there is an urgent need to find possible solutions from the current literature. While traditional literature review studies provide valuable insights, these studies have limitations, including analyzing a limited number of papers, having various biases, being time-consuming and labor-intensive, focusing on a few topics, and lack of data-driven tools. This study fills the mentioned limitations and gaps in the literature and practice by analyzing diseases and chemicals with text mining methods in a corpus containing COVID-19 research papers and by finding associations between the two biomedical concepts. Appreciating this context is vital due to the lack of a systematic large-scale literature review survey and the importance of fast literature review during the current COVID-19 pandemic for developing treatments.

This study describes a cohesive research plan that will significantly help COVID-19 researchers streamline publication search for timely informed decision making. Better knowledge of COVID-19 literature should impact our understanding of biomedical concepts and provide better preparation for future waves of COVID-19 and other infectious diseases.

Declaration of Competing Interest

The authors state that they have no conflict of interest

Acknowledgments

This work is partially supported by a COVID-19 Research Initiative grant (157000-20-54159) from the Office of the Vice President for Research and the Big Data Health Science Center (BDHSC) at the University of South Carolina. All opinions, findings, conclusions, and recommendations in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

Appendix A

25 categories, including ten diseases or chemicals and one research example containing some of the words in a category.

C1	**Example**	**C14**	**Example**
infection_disease	(Lei et al., 2020)	infected_disease	(Barrett, Moore, Yaffe, & Moore, 2020)
sars_cov infected_disease		cough_disease	
infections_disease		fever_disease	
pneumonia_disease		infection_disease	
serine_chemical		oxygen_chemical	
inflammation_disease		infectious_disease	
oxygen_chemical		pneumonia_disease	
sars_cov_infection_disease		death_disease	
critically_ill_disease		died_disease	
covid_disease		mortality_disease	
C2	**Example**	**C15**	**Example**
mortality_disease		fever_disease	(Nakamura et al., 2020)
tocilizumab_chemical	(Dzieciatkowski et al., 2020)	cough_disease	
coronavirus_disease		fatigue_disease	
inflammation_disease		diarrhea_disease	
acute_respiratory_distress_syndrome_disease		pneumonia_disease	
vitamin_d_chemical		dry_cough_disease	
coagulopathy_disease		myalgia_disease	
disseminated_intravascular_coagulation_disease		shortness_of_breath_disease	
heparin_chemical		vomiting_disease	
sepsis_disease		dyspnea_disease	
C3	**Example**	**C16**	**Example**
coronavirus_disease	(Orleans, is Vice, & Manchikanti, 2020)	acute_respiratory_distress_syndrome_disease	(De Vitis, Passiatore, Perna, Proietti, & Taccardo, 2020)
novel_coronavirus_disease		ards_disease	
pneumonia_disease		critically_ill_disease	
infection_disease		pneumonia_disease	
lopinavir/ritonavir_chemical		oxygen_chemical	
steroid_chemical		mortality_disease	
lpv/r_chemical		respiratory_failure_disease	
eea_chemical		died_disease	
immunodeficiency_disease		death_disease	
inability_disease		septic_shock_disease	
C4	**Example**	**C17**	**Example**
infection_disease	(Chen et al., 2020)	infection_disease	(Yao, Qian, Zhu, Wang, & Wang, 2020)
water_chemical		novel_coronavirus_infected_pneumonia_disease	
alcohol_chemical		pneumonia_disease	
coronavirus_disease		novel_coronavirus_pneumonia_disease	
coronavirus_disease		infected_disease	
coughing_disease		infections_disease	
infected_disease		ncov_infection_disease	
covid_disease		fever_disease	
cross_infection_disease		novel_coronavirus_infection_disease	
hydrogen_peroxide_chemical		deaths_disease	
C5	**Example**	**C18**	**Example**
coronavirus_disease	(Kim et al., 2020)	water_chemical	
corona_virus_disease		infected_disease	(Costanzo, De Giglio, & Roviello, 2020)
f_fdg_chemical		pneumonia_disease	
carbon_dioxide_chemical å		penicillin_chemical	
copper_chemical		streptomycin_chemical	
lpv_chemical		dmem_chemical	
tcm_chemical		carbon_chemical	
coronary_syndrome_disease		pbs_chemical	
infarction_disease		sds_chemical	
stainless_steel_chemical		sialic_acid_chemical	
C6	**Example**	**C19**	**Example**
remdesivir_chemical	(Lombardy, 2020)	coronavirus_disease	(Caselli & Aricò, 2020)
lopinavir/ritonavir_chemical		coronavirus_disease	
lopinavir_chemical		respiratory_syndrome_disease	
ribavirin_chemical		steroids_chemical	
ritonavir_chemical		new_coronavirus_pneumonia_disease	
chloroquine_chemical		pleural_effusion_disease	
favipiravir_chemical		novel_coronavirus_disease	
lopinavir_ritonavir_chemical		hypoxia_disease	
mortality_disease		ncov_disease	
oseltamivir_chemical		chest_tightness_disease	
C7	**Example**	**C20**	**Example**
middle_east_respiratory_syndrome_disease	(Shneider, Kudriavtsev, & Vakhrusheva, 2020)	inflammation_disease	(Monti & Montecucco, 2020)
mers_disease		viral_infections_disease	
infected_disease		viral_infection_disease	
sars_disease		coronavirus_infection_disease	
severe_acute_respiratory_syndrome_disease		acute_respiratory_syndrome_disease	
deaths_disease		cov_infection_disease	
pneumonia_disease		virus_infection_disease	
infection_disease		toxicity_disease	
infections_disease		coronavirus_infections_disease	
death_disease		fibrosis_disease	

C8	**Example**	**C21**	**Example**
hydroxychloroquine_chemical	(Liu et al., 2020)	anxiety_disease	(Moccia et al., 2020)
chloroquine_chemical		pain_disease	
malaria_disease		depression_disease	
rheumatoid_arthritis_disease		stress_disease	
rheumatic_diseases_disease		fatigue_disease	
chloroquine_phosphate_chemical		infectious_disease	
sle_disease		melatonin_chemical	
systemic_lupus_erythematosus_disease		psychiatric_disease	
autoimmune_diseases_disease		delirium_disease	
baricitinib_chemical		infectious_diseases	
C9	**Example**	**C22**	**Example**
coronavirus_disease	(Tian et al., 2020)	infected_disease	(Members, Wang, Zeng, Wu, & Sun, 2020)
infection_disease		sars_cov_infection_disease	
infected_disease		hydrogen_chemical	
covid_disease		sars_coronavirus_infection_disease	
infections_disease		serine_chemical	
anxious_disease		cytotoxicity_disease	
critical_illness_disease		acute_respiratory_infections_disease	
sars_cov_infections_disease		oligonucleotides_chemical	
covid_infection_disease		glycine_chemical	
tcb_chemical		asparagine_chemical	
C10	**Example**	**C23**	**Example**
coronavirus_disease	(Kang et al., 2020)	hypertension_disease	(Carboni et al., 2020)
deaths_disease		diabetes_disease	
infection_disease		cardiovascular_disease	
coronavirus_disease		death_disease	
oxygen_chemical		heart_failure_disease	
emphysema_disease		respiratory_failure_disease	
multi_organ_failure_disease		myocarditis_disease	
colchicine_chemical		infection_disease	
toxicity_disease		aldosterone_chemical	
situazione_chemical		cardiovascular_diseases_disease	
C11	**Example**	**C24**	**Example**
diabetes_disease	(Shah, Das, Jain, Misra, & Negi, 2020)	hydroxychloroquine_chemical	(Bartlett et al., 2020)
hypertension_disease		chloroquine_chemical	
diabetes_mellitus_disease		azithromycin_chemical	
glucose_chemical		coronavirus_disease	
mortality_disease		hcq_chemical	
novel_coronavirus_disease		tocilizumab_chemical	
ibuprofen_chemical		cq_chemical	
obesity_disease		qt_prolongation_disease	
diabetic_disease		qtc_prolongation_disease	
respiratory_infections_disease		qtc_chemical	
C12	**Example**	**C25**	**Example**
cancer_disease	(Elfiky & Ribavirin, 2020)	viral_infection_disease	(Cheng et al., 2020)
tumor_disease		infection_disease	
cancers_disease		infections_disease	
bleeding_disease		infected_disease	
lung_cancer_disease		viral_infections_disease	
malignancies_disease		virus_infection_disease	
tumors_disease		lipid_chemical	
hemorrhage_disease		toxicity_disease	
formalin_chemical		atp_chemical	
dexamethasone_chemical		hiv_disease	
C13	**Example**		
coronavirus_disease	(Cheng et al., 2020)		
nitrogen_chemical			
urea_chemical			
creatinine_chemical			
bilirubin_chemical			
death_disease			
liver_injury_disease			
acute_kidney_injury_disease			
liver_damage_disease			
liver_disease			

References

Al-Barashdi, H., & Al-Karousi, R. (2018). Big data in academic libraries: Literature review and future research directions. *Journal of Information Studies & Technology (JIS&T), 201*(2), 13 Hamad bin Khalifa University Press (HBKU Press).

Barrett, C. D., Moore, H. B., Yaffe, M. B., & Moore, E. E. (2020). ISTH interim guidance on recognition and management of coagulopathy in COVID-19: A comment. *Journal of Thrombosis and Haemostasis, 18*(8), 2060–2063.

Bartlett, D. L., Howe, J. R., Chang, G., Crago, A., Hogg, M., Karakousis, G., et al. (2020). Management of cancer surgery cases during the COVID-19 pandemic: Considerations. *Annals of Surgical Oncology*, 1–4.

Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19—Preliminary report. *The New England Journal of Medicine, 383*(19), 1813–1826.

Carboni, E, Carta, AR, & Carboni, E. (2020). Can pioglitazone be potentially useful therapeutically in treating patients with covid-19? *Medical Hypotheses*, Article 109776.

Caselli, D., & Aricò, M. (2020). *2019-nCoV: Polite with children!*: 12. Pediatric Reports PAGEPress.

CDC. COVID-19 and your health [Internet]. Centers for disease control and prevention. (2021). [cited 2021 Apr 28]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html

CDC. Coronavirus Disease (2019). (COVID-19) – Symptoms [Internet]. Centers for disease control and prevention. 2021 [cited 2021 Apr 28]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

Chen, Q., Ai, N., Liao, J., Shao, X., Liu, Y., & Fan, X. (2017). Revealing topics and their evolution in biomedical literature using Bio-DTM: A case study of ginseng. *Chinese Medicine, 12*(1), 27.

Chen, W., Lan, Y., Yuan, X., Deng, X., Li, Y., Cai, X., et al. (2020). Detectable 2019-nCoV viral RNA in blood is a strong indicator for the further clinical severity. *Emerging Microbes & Infections, 9*(1), 469–473.

Cheng, Y., Luo, R., Wang, K., Zhang, M., Wang, Z., Dong, L., et al. (2020). Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney International, 97*(5), 829–838.

Costanzo, M., De Giglio, M. A. R., & Roviello, G. N. (2020). *SARS-CoV-2: Recent reports on antiviral therapies based on Lopinavir/Ritonavir, Darunavir/Umifenovir, hydroxychloroquine, Remdesivir, Favipiravir and other drugs for the treatment of the new coronavirus*. Current Medicinal Chemistry Bentham Science Publishers.

COVID-19 Map [Internet]. Johns Hopkins coronavirus resource center. [cited (2020). May 28]. Available from: https://coronavirus.jhu.edu/map.html

De Vitis, R., Passiatore, M., Perna, A., Proietti, L., & Taccardo, G. (2020). COVID-19 contagion and contamination through hands of trauma patients: What risks and what precautions? *Journal of Hospital Infection, 105*(2), 354–355.

Dzieciatkowski, T., Szarpak, L., Filipiak, K. J., Jaguszewski, M., Ladny, J. R., & Smereka, J. (2020). COVID-19 challenge for modern medicine. *Cardiology Journal, 27*(2), 175–183.

Elfiky, A. A., & Ribavirin, R. (2020). Sofosbuvir, Galidesivir, and Tenofovir against SARS–CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sciences, 253*(2020), 117592.

Fajgenbaum, D. C., Khor, J. S., Gorzewski, A., Tamakloe, M-A., Powers, V., Kakkis, J. J., et al. (2020). Treatments administered to the first 9152 reported cases of COVID-19: A systematic review. *Infectious Diseases and Therapy, 9*, 435–449.

gensim: Topic modelling for humans [Internet]. (2020).

Gonzalez, G. H., Tahsin, T., Goodale, B. C., Greene, A. C., & Greene, C. S. (2016). Recent advances and emerging applications in text and data mining for biomedical discovery. *Briefings in Bioinformatics, 17*(1), 33–42.

Grover, P., & Kar, A. K. (2017). Big data analytics: A review on theoretical contributions and tools used in literature. *Global Journal of Flexible Systems Management, 18*(3), 203–229.

Harvard Health Publishing. Treatments for COVID-19 [Internet]. Harvard health. (2021). [cited 2021 Apr 28]. Available from: https://www.health.harvard.edu/diseases-and-conditions/treatments-for-covid-19

Hatmi, Z. N. (2021). A systematic review of systematic reviews on the COVID-19 pandemic. *SN Comprehensive Clinical Medicine*, 1–18.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. In *Proceedings of the 47th Hawaii international conference on system sciences* (pp. 1833–1842). IEEE.

Index of /pub/lu/LitCovid [Internet]. [cited (2020). May 28]. Available from: https://ftp.ncbi.nlm.nih.gov/pub/lu/LitCovid /

Kang, Y., Chen, T., Mui, D., Ferrari, V., Jagasia, D., Scherrer-Crosbie, M., et al. (2020). *Cardiovascular manifestations and treatment considerations in covid-19*. Heart BMJ Publishing Group Ltd and British Cardiovascular Society.

Kar, A. K., & Navin, L. (2020). Diffusion of blockchain in insurance industry: An analysis through the review of academic and trade literature. *Telematics and Informatics, 58*(2021), 101532.

Karami, A., & Anderson, M. (2020). Social media and COVID-19: Characterizing anti-quarantine comments on Twitter. *Proceedings of the Association for Information Science and Technology, 57*(1), e349.

Karami, A., Dahl, A. A., Shaw, G., Valappil, S. P., Turner-McGrievy, G., Kharrazi, H., et al. (2021a). Analysis of social media discussions on (#)diet by blue, red, and swing states in the U.S. *Healthcare Multidisciplinary Digital Publishing Institute, 9*(5), 518. 10.3390/healthcare9050518.

Karami, A., & Elkouri, A. (2019). Political popularity analysis in social media. In *Proceedings of the international conference on information* (pp. 456–465). Springer.

Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2018). Fuzzy approach topic discovery in health and medical corpora. *International Journal of Fuzzy Systems, 20*(4), 1334–1345.

Karami, A., Ghasemi, M., Sen, S., Moraes, M. F., & Shah, V. (2019). Exploring diseases and syndromes in neurology case reports from 1955 to 2017 with text mining. *Computers in Biology and Medicine, 109*, 322–332.

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020a). Twitter and research: A systematic literature review through text mining. *IEEE Access, 8*, 67698–67717.

Karami, A., Lundy, M., Webb, F., Turner-McGrievy, G., McKeever, B. W., & McKeever, R. (2021b). Identifying and analyzing health-related themes in disinformation shared by conservative and liberal Russian trolls on Twitter. *International Journal of Environmental Research and Public Health, 18*(4), 2159.

Karami, A., & Pendergraft, N. M. (2018). Computational analysis of insurance complaints: GEICO case study. *Proceeding of the International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.*

Karami, A., & Shaw, G. (2019). An exploratory study of (#) exercise in the Twittersphere. In *Proceedings of the iConference 2019 Proceedings iSchools*. 10.21900/iconf.2019.103327.

Karami, A., White, C. N., Ford, K., Swan, S., & Spinel, M. Y. (2020b). Unwanted advances in higher education: Uncovering sexual harassment experiences in academia with text mining. *Information Processing & Management, 57*(2), Article 102167.

Karami, A. (2015). *Fuzzy topic modeling for medical corpora*. Baltimore County: University of Maryland.

Kim, J-M., Chung, Y-S., Jo, H. J., Lee, N-J., Kim, M. S., Woo, S. H., et al. (2020). Identification of coronavirus isolated from a patient in Korea with COVID-19. *Osong Public Health and Research Perspectives, 11*(1), 3.

Kumar, S, Kar, AK, & Ilavarasan, PV. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(1), Article 100008. 10.1016/j.jjimei.2021.100008.

Lei, S., Jiang, F., Su, W., Chen, C., Chen, J., Mei, W., et al. (2020). Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection. *EClinicalMedicine, 21*(2020), 100331.

Leydesdorff, L., & Nerghes, A. (2017). Co-word maps and topic modeling: A comparison using small and medium-sized corpora (N< 1,000). *Journal of the Association for Information Science and Technology, 68*(4), 1024–1035.

Liu, N., Zhang, F., Wei, C., Jia, Y., Shang, Z., Sun, L., et al. (2020). Prevalence and predictors of PTSS during COVID-19 outbreak in China hardest-hit areas: Gender differences matter. *Psychiatry Research, 287*(2020), 112921.

Lombardy, S. I. S. I. (2020). Vademecum for the treatment of people with COVID-2.0, 13 March 2020. *Le Infezioni in Medicine, 28*(2), 143.

Lovato, A., & De Filippis, C. (2020). Clinical presentation of COVID-19: A systematic review focusing on upper airway symptoms. *Ear, Nose & Throat Journal, 99*(9), 569–576.

McCallum, A. K. (2002).Mallet: A machine learning for language toolkit 2002.

Members, W. C., Wang, H., Zeng, T., Wu, X., & Sun, H. (2020). Holistic care for patients with severe coronavirus disease 2019: An expert consensus. *International Journal of Nursing Sciences, 7*(2), 128–134.

Mirza, B., Wang, W., Wang, J., Choi, H., Chung, N. C., & Ping, P. (2019). Machine learning and integrative analysis of biomedical big data. *Genes, 10*(2), 87.

Moccia, L., Janiri, D., Pepe, M., Dattoli, L., Molinaro, M., De Martin, V., et al. (2020). Affective temperament, attachment style, and the psychological impact of the COVID-19 outbreak: An early report on the Italian general population. *Brain, Behavior, and Immunity, 87*(2020), 75–79.

Mohammadi, E., & Karami, A. (2020). Exploring research trends in big data across disciplines: A text mining analysis. *Journal of Information Science*. 10.1177/0165551520932855.

Monti, S., & Montecucco, C. (2020). Can hydroxychloroquine protect patients with rheumatic diseases from COVID-19? Response to: 'Does hydroxychloroquine prevent the transmission of COVID-19?' by Heldwein and Calado and 'SLE, hydroxychloroquine and no SLE patients with COVID-19: A comment' by Joob and Wiwanitkit. *Annals of the Rheumatic Diseases, 79*(6), e62.

Nakamura, K., Hikone, M., Shimizu, H., Kuwahara, Y., Tanabe, M., Kobayashi, M., et al. (2020). A sporadic COVID-19 pneumonia treated with extracorporeal membrane oxygenation in Tokyo, Japan: A case report. *Journal of Infection and Chemotherapy, 26*(7), 756–761.

NIH. Therapeutic management [Internet]. COVID-19 treatment guidelines. (2021). [cited 2021 Apr 28]. Available from: https://www.covid19treatmentguidelines.nih.gov/therapeutic-management/

Nikita, M., & Nikita, M. M. (2019).Package 'ldatuning.;

Organization WH. Coronavirus disease 2019 (COVID-19): Situation report, 67. World Health Organization; 2020;

Orleans, L. A., Vice, H., & Manchikanti, L. (2020). Expanded umbilical cord mesenchymal stem cells (UC-MSCs) as a therapeutic strategy in managing critically ill COVID-19 patients: The case for compassionate use. *Pain Physicia, 23*, E71–E83.

PubMed Search Results on COVID-19 [Internet]. Available from: https://pubmed.ncbi.nlm.nih.gov/?term=covid-19&filter=years.2020-2020 (2021).

Santoso, A., Pranata, R., Wibowo, A., Al-Farabi, M. J., Huang, I., & Antariksa, B. (2020). Cardiac injury is associated with mortality and critically ill pneumonia in COVID-19: A meta-analysis. *The American Journal of Emergency Medicine*. 10.1016/j.ajem.2020.04.052.

Shah, S., Das, S., Jain, A., Misra, D. P., & Negi, V. S. (2020). A systematic review of the prophylactic role of chloroquine and hydroxychloroquine in Coronavirus Disease-19 (COVID-19). *International Journal of Rheumatic Diseases, 25*(3), 613–619.

Shin, G., Jarrahi, M. H., Fei, Y., Karami, A., Gafinowitz, N., Byun, A., et al. (2019).Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. *Journal of Biomedical Informatics*;103153.

Shneider, A., Kudriavtsev, A., & Vakhrusheva, A. (2020). Can melatonin reduce the severity of COVID-19 pandemic? *International Reviews of Immunology*, 1–10.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software, 1*(3), 37.

Strollo, R., & Pozzilli, P. (2020). DPP4 inhibition: Preventing SARS-CoV-2 infection and/or progression of COVID-19? *Diabetes/Metabolism Research and Reviews, 36*(8), e3330.

Tian, X., Li, C., Huang, A., Xia, S., Lu, S., Shi, Z., et al. (2020). Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerging Microbes & Infections, 9*(1), 382–385.

Valdez, A. C., Ziefle, M., Verbert, K., Felfernig, A., & Holzinger, A. (2016). Recommender systems for health informatics: State-of-the-art and future perspectives. *Machine Learning for Health Informatics*, 391–414.

van Altena, A. J., Moerland, P. D., Zwinderman, A. H., & Olabarriaga, S. D. (2016). Understanding big data themes from scientific biomedical literature through topic modeling. *Journal of Big Data, 3*(1), 23.

Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021). Artificial intelligence in marketing: Systematic review and future research direction. *International Journal of Information Management Data Insights, 1*(1), 100002.

Webb, F., Karami, A., & Kitzie, V. (2018). Characterizing diseases and disorders in gay users' tweets. The Southern Association for Information Systems (SAIS) 2018 Proceedings. 31. https://aisel.aisnet.org/sais2018/31.

WHO Timeline – COVID-19 [Internet]. [cited (2020). May 28]. Available from: https://www.who.int/news-room/detail/27-04-2020-who-timeline—covid-19

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software, 40*(i01).

Wiesner, M., & Pfeifer, D. (2014). Health recommender systems: Concepts, requirements, technical basics and challenges. *International Journal of Environmental Research and Public Health, 11*(3), 2580–2607.

Witten, I. H., Don, K. J., Dewsnip, M., & Tablan, V. (2004). Text mining in a digital library. *International Journal on Digital Libraries, 4*(1), 56.

Yao, T-T., Qian, J-D., Zhu, W-Y., Wang, Y., & Wang, G-Q. (2020). A systematic review of Lopinavir therapy for SARS coronavirus and MERS coronavirus—A possible reference for coronavirus disease-19 treatment option. *Journal of Medical Virology, 92*(6), 556–563.

Yazdi, M. M. (2013).Topsis: Topsis method for multiple-criteria decision making (MCDM). R package version;1.

Zhang, L., & Liu, Y. (2020). Potential interventions for novel coronavirus in China: A systematic review. *Journal of Medical Virology, 92*(5), 479–490. 10.1002/jmv.25707.

Zhang, Y., & Gu, H. (2011). Text mining with application to academic libraries. *International Workshop on Computer Science for Environmental Engineering and EcoInformatics*, 200–205.

Zhu, Y., Kim, M-H., Banerjee, S., Deferio, J., Alexopoulos, G. S., & Pathak, J. (2018). Understanding the research landscape of major depressive disorder via literature mining: An entity-level analysis of PubMed data from 1948 to 2017. *JAMIA Open, 1*(1), 115–121. 10.1093/jamiaopen/ooy001.