

Comparative Analysis of Telemedicine in Media Coverage Pre- and Post-COVID-19 using Unsupervised Latent Dirichlet Topic Modeling

Haewon Byeon

Department of Digital Anti-Aging Healthcare (BK21),
Medical Big Data Research Center, Inje University, Gimhae 50834, South Korea

Abstract—Telemedicine, driven by technology, has become a game-changer in healthcare, with the COVID-19 pandemic amplifying its significance by necessitating remote healthcare solutions. This study explores the evolution of telemedicine through news big data analysis. Our research encompassed a vast dataset from 51 media outlets (total 28,372 articles), including national and regional dailies, economic newspapers, broadcasters, and professional journals. Using LDA analysis, we delved into pre- and post-pandemic telemedicine trends comprehensively. A crucial revelation was the prominence of "medical law" in telemedicine discussions, underscoring the need for legal reforms. Keywords like "artificial intelligence" and "big data" underscored technology's pivotal role. Post-pandemic, keywords like "COVID-19," "online healthcare," and "telemedicine" surged, reflecting the pandemic's impact on remote healthcare reliance. These keywords' increased frequency highlights the pandemic's transformative influence. This study stresses addressing healthcare's legal constraints and maximizing technology's potential. To seamlessly integrate telemedicine, policy support and institutional backing are imperative. In summary, telemedicine's rise, propelled by COVID-19, signifies a healthcare paradigm shift. This study sheds light on its trajectory, emphasizing legal reforms, tech innovation, and pandemic-induced changes. The post-pandemic era must prioritize informed policy decisions for telemedicine's effective and accessible implementation.

Keywords—Telemedicine; COVID-19; medical law; healthcare transformation; LDA topic modeling

I. INTRODUCTION

One important factor that is rapidly evolving in the modern healthcare environment is the increase in telemedicine services. This is achieved through the combination of advancements in information technology and innovative approaches in the medical field, providing patients with more effective and convenient healthcare services. Particularly, the importance of telemedicine has been further emphasized after the outbreak of COVID-19 in early 2020.

The COVID-19 pandemic has placed a significant burden on the global healthcare system. With the increase in patients and limitations in healthcare personnel, traditional methods of healthcare service delivery have faced constraints. As an alternative, telemedicine has gained prominence, offering patients the opportunity to safely receive medical consultations from their homes. This transformation has become a significant catalyst for revolutionizing the healthcare system, especially as

the COVID-19 pandemic has brought about revolutionary changes in the medical field worldwide. Patients can now consult with doctors through computers or smartphones within the comforts of their own homes. In response to these changes, healthcare institutions have strengthened their telemedicine systems and strived to introduce new technologies to provide the best possible healthcare services to patients.

Extensive research has been conducted on telemedicine in the last decade [1-4]. However, these studies predominantly focus on specific geographical regions or timeframes, resulting in a constrained understanding of telemedicine trends pre- and post-COVID-19 pandemic. Moreover, the reliance on survey-based methodologies in these studies introduces potential biases due to subjective responses and recall inaccuracies. To address these shortcomings and enhance the accuracy of findings, leveraging big data from news sources for unsupervised topic modeling emerges as a promising alternative. This method offers an objective and expansive analysis of telemedicine's evolution and its perception in media discourse across different temporal contexts.

Unsupervised topic modeling has been extensively applied in research utilizing text data, such as identifying industrial accident-related issues using website text data [5], discerning attitudes towards vaccines via Twitter [6], exploring topics on climate change through news articles [7], and investigating topics related to the Omicron variant [8]. With the increasing prevalence of such studies, the efficacy of unsupervised topic modeling applied to news articles has been substantiated [9, 10].

An examination of telemedicine trends pre- and post-COVID-19 constitutes a critical area of inquiry within contemporary healthcare research. Such comparative analyses are instrumental in elucidating the evolution and emerging patterns in telemedicine, thereby offering valuable insights for the enhancement of future healthcare systems and the optimization of patient services.

The organization of this study is as follows: Section II delineates the methodology for data collection and the implementation of Latent Dirichlet Allocation (LDA) modeling. Section III details the outcomes of the network analysis conducted, Section IV discusses the implications of these findings. Finally, Section V concludes the paper.

II. MATERIALS AND METHODS

A. Data Collection

This study utilized news articles from a total of 51 media outlets, including 10 national dailies, 8 economic dailies, 26 regional dailies, 5 broadcasting companies, and 2 professional journals, analyzed through the news big data analysis service, BIGKinds, provided by the Korea Press Foundation (see Table I). BigKinds integrates big data analytics with a database composed of news articles, offering analytical support. Since 1990, approximately 70 million news contents have been transformed into big data within this platform. Furthermore, by converting unstructured text into structured data, it provides information, making it a service capable of analyzing societal phenomena through articles.

TABLE I. MEDIA SUBJECT TO ANALYSIS

Category (Number)	Daily newspaper name
Professional journals (2)	Electronic Newspaper, Digital Times
Broadcasting company (5)	KBS, SBS, MBC, YTN, OBS
National daily newspaper (10)	Kyunghyang Shinmun, Chosun Ilbo, Donga Ilbo, Hankook Ilbo, Hankyoreh, JoongAng Ilbo, Kookmin Ilbo, Munhwa Ilbo, Naeil Shinmun, Segye Ilbo, Seoul Shinmun
Economic daily Newspaper (8)	Maeil Economy, Money Today, Seoul Economy, Asia Economy, Aju Economy, Financial News, Korea Economy, Herald Economy
Local daily newspaper (26)	Busan Ilbo, Chungbuk Daily, Chungcheong Daily, Chungcheong Today, Daegu Ilbo, Daejeon Ilbo, Gangwon Ilbo, Gwangju Daily, Gwangju Ilbo, Gangwon Provincial Daily, Gyeonggi Ilbo, Gyeongin Ilbo, Gyeongnam Newspaper, Gyeongnam Provincial Daily, Gyeongsang Daily, Halla Daily, International Newspaper, Jemin Daily, Jeonbuk Ilbo, Jeonbuk Provincial Daily, Jeonnam Ilbo, Joongbu Daily, Joongbu Daily, Maeil Daily, Mudeung Ilbo, Ulsan Daily, Yeongnam Ilbo

TABLE II. NUMBER OF ARTICLES ANALYSIS

	Before the COVID-19 pandemic	After the COVID-19 pandemic	Total
Total number of articles	5,232	23,884	29,116
Number of excluded articles	92	652	744
Number of analysis articles	5,140	23,232	28,372

To investigate the trends in remote healthcare before and after the COVID-19 pandemic, keywords such as telemedicine, remote medical treatment, and non-face-to-face medical care were searched and analyzed. The collected articles were divided into two periods: pre-COVID-19 pandemic period, from January 1, 2016 to November 30, 2019, and post-COVID-19 pandemic period, from December 1, 2019 to the endemic declaration date on May 5, 2023. After excluding duplicate and irrelevant news articles, a total of 5,140 and

23,232 articles were analyzed for the two respective periods (see Table II).

B. Analysis Method

Firstly, frequency analysis was conducted to investigate the occurrence of words based on the collected text data of news articles, followed by the analysis of word weights using TF-IDF. TF-IDF is the most commonly used weighting algorithm and is widely used in keyword extraction and topic classification [11]. LDA topic modeling is an algorithm that is useful for extracting latent topics from big data consisting of text [12]. LDA (Latent Dirichlet Allocation) is one of the most common topic modeling methods, contributing to the extraction of coherent topics from data [13]. The fundamental concept of LDA is to represent the latent topics in a document composed of text data as a random mixture, wherein topics are characterized by the distribution of words [14]. A diagram of the procedure for this study is presented in Fig. 1. All analyses conducted in this study were performed using Python.

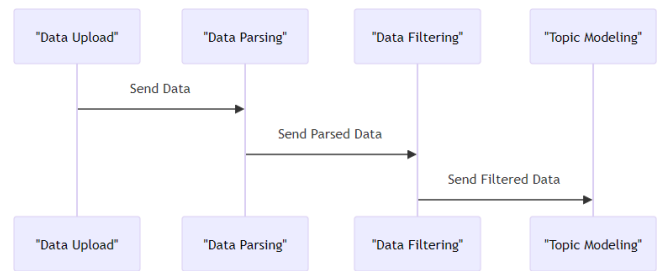


Fig. 1. Flowchart of the research procedure.

C. Network Analysis

Fig. 2 present the results of a keyword network analysis conducted on keywords related to telemedicine. The network analysis depicts the interconnectedness between words using noun phrases extracted through Structured SVM from the top 100 articles with high accuracy.

Fig. 2 represents the pre-pandemic network, with weights ranging from 6 to 58. The keyword analysis revealed that medical law exhibited the strongest association. This suggests that telemedicine is currently considered illegal under existing medical laws, and therefore, it is speculated that the revision of medical laws is crucial for the implementation of telemedicine. Additionally, keywords such as Ministry of Health and Welfare, United States, China, and Japan emerged as related keywords for similar reasons.

Fig. 3 illustrates the network configuration in the post-pandemic context, characterized by varying connection weights that range between 8 and 71. The keyword analysis revealed similarities to the pre-pandemic network, with the addition of keywords related to COVID-19, National Security Council, and Deputy Chief of Policy Committee. This difference can be attributed to the temporary rise in the importance of telemedicine due to the enabling of remote medical services during the COVID-19 pandemic.

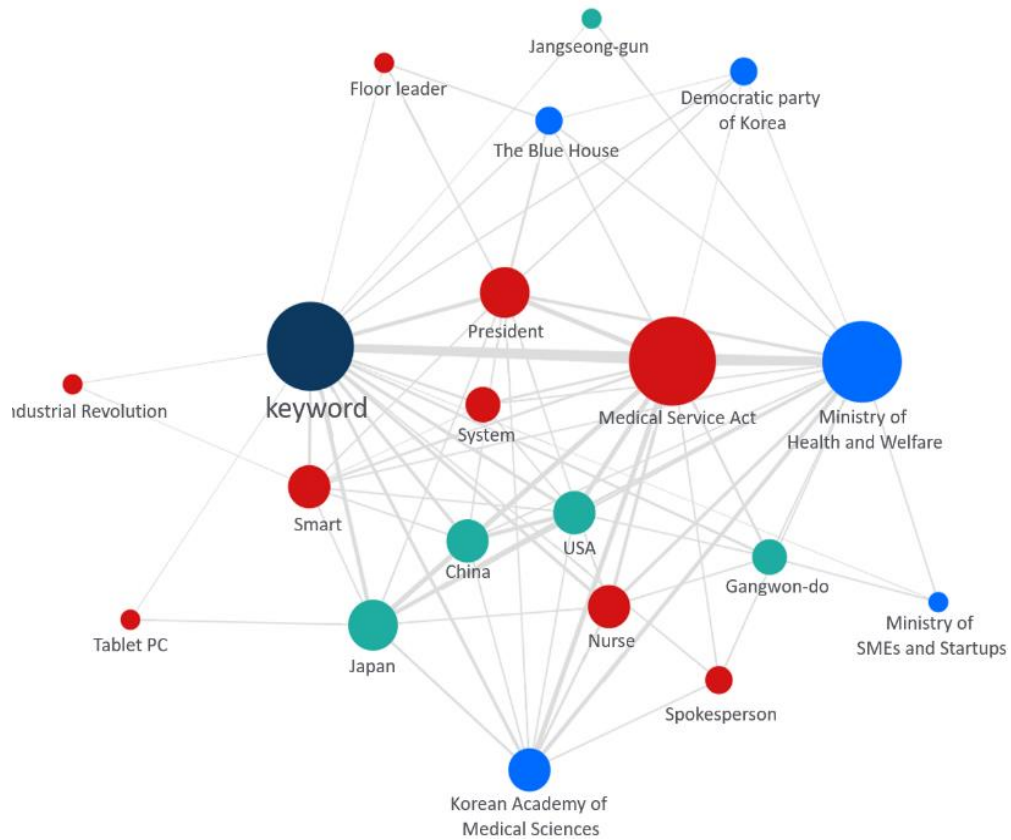


Fig. 2. Analysis of the relationship between keywords related to telemedicine: A comparative study before the COVID-19 pandemic.

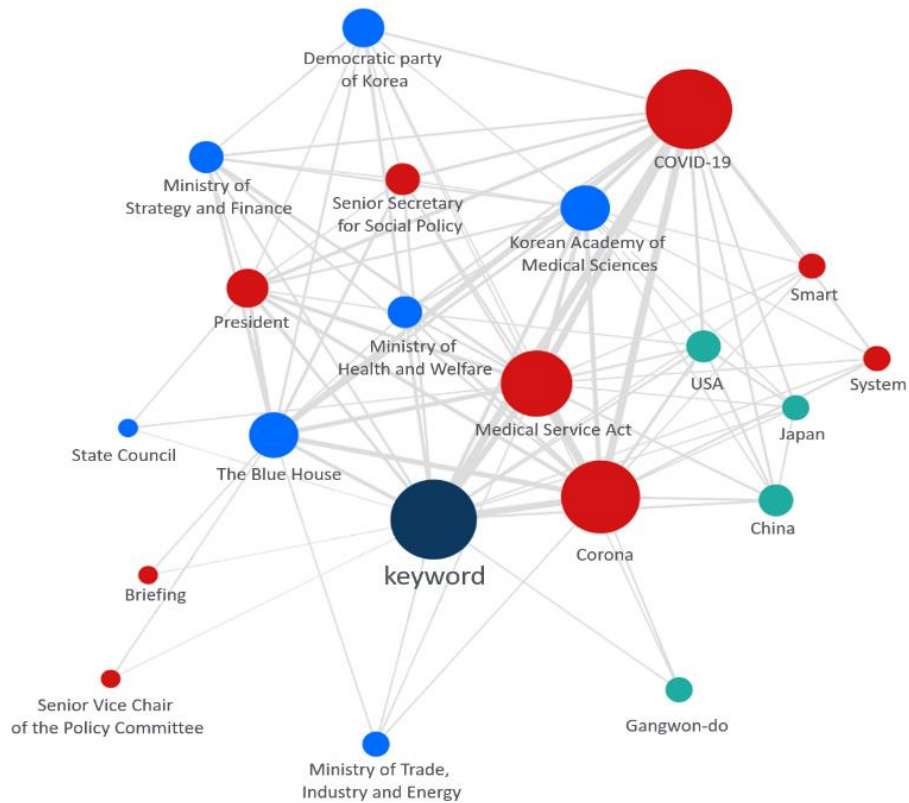


Fig. 3. Analysis of the relationship between keywords related to telemedicine: A comparative study after the COVID-19 pandemic.

III. RESULTS

A. Frequency Analysis Result

Table III presents the frequency analysis of relevant keywords before and after the COVID-19 pandemic. The results show that after the pandemic, additional keywords related to COVID-19, online, and telemedicine were identified.

B. LDA Topic Modeling

Fig. 4 illustrates the results of calculating coherence scores using Gensim, a Python package. Gensim is a tool for topic

modeling, analyzing the interrelationships between words to reinterpret the content of documents in a new and profound meaning. Additionally, by simplifying the representation of documents, it is utilized to enhance the efficiency of information processing [15]. The number of topics is determined based on the coherence scores, resulting in three topics before the occurrence of COVID-19 and seven topics after. The optimal number of topics is determined by the highest coherence score.

TABLE III. TOP 15 KEYWORDS BEFORE AND AFTER THE COVID-19 PANDEMIC

Before the COVID-19 pandemic			After the COVID-19 pandemic		
Word	Frequency	TF-IDF	Word	Frequency	TF-IDF
Telemedicine	1550	1857.13745	COVID-19	11943	7945.69675
Korea	1230	1757.94827	The coronavirus	4257	7222.98417
United States	950	1602.92958	Online	3643	6748.47047
Telemedicine	781	1470.58678	Confirmed case	2981	6119.80347
China	750	1442.55203	Face-to-face	2893	6025.80369
Job	748	1440.69989	United States	2757	5875.23554
Korea	564	1245.3022	Seoul	2391	5435.70132
Japan	482	1139.8296	Korea	2377	5417.82679
Seoul	448	1092.12789	Remote Healthcare	1850	4680.13852
Medical Law	432	1068.79851	Infectious disease	1849	4678.6079
Big Data	429	1064.35893	AI	1659	4377.62493
A pilot project	399	1018.78418	Medical staff	1632	4333.1425
Cheong Wa Dae	383	993.565504	Start-up	1579	4244.51904
Smartphone	365	964.393884	Coronavirus	1526	4154.11628
Competitiveness	347	934.33424	Job	1480	4074.16339

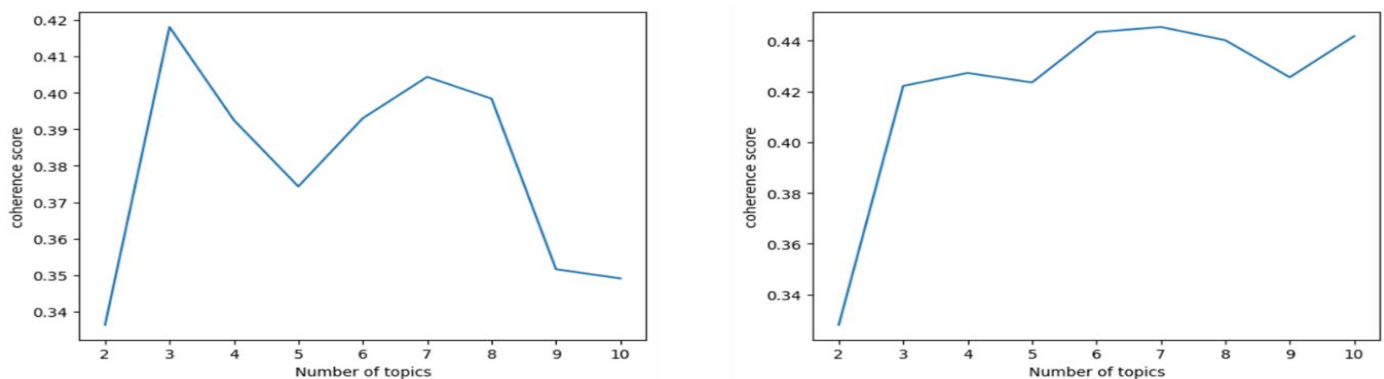


Fig. 4. Coherence scores before and after the COVID-19 pandemic.

The results of topic modeling before and after the COVID-19 pandemic are presented in Table IV and Table V. Since the topic weights were low in the 6th word of each topic, the top 5 words for each topic are provided. Before the COVID-19 pandemic, topics related to remote healthcare and technology (Topic 1), global remote healthcare (Topic 2), and remote healthcare and medical law (Topic 3) were identified, focusing on the definition and technological aspects of remote healthcare, global implementation of remote healthcare, and

policy and technological considerations in the adoption of remote healthcare.

The results of topic modeling after the COVID-19 pandemic are presented in Table V. It was observed that topics related to COVID-19 emerged, given the temporary allowance of remote healthcare during the pandemic. Additionally, topics related to the definition of remote healthcare and its social

implications were identified, such as Topic 6 and Topic 7, in order to facilitate the full implementation of remote healthcare.

TABLE IV. TOPIC MODELING BEFORE THE COVID-19 PANDEMIC

Group	Topic name	Top 5 words
1	Remote healthcare and Science and Technology	Seoul, Network, Artificial Intelligence, Big Data, Deputy Prime Minister
2	Global remote healthcare	Korea, United States, China, jobs
3	Remote healthcare and medical law	Remote healthcare, medical law, pilot projects, telemedicine, Ministry of Health and Welfare

TABLE V. TOPICS MODELING AFTER THE COVID-19 PANDEMIC

Group	Topic name	Top 5 words
1	Temporary introduction	Remote healthcare, Telemedicine, Temporary, Medical, Medical law
2	Position on remote healthcare	Medical circles, medical associations, Korean Medical Association, collusion, Cheong Wa Dae
3	COVID-19 and remote healthcare	Confirmed cases, Health centers, Medical institutions, Respiratory tract, Patients with diseases
4	COVID-19 and remote healthcare in major countries	Korea, United States, online, China, COVID-19
5	COVID-19 and home treatment	COVID-19, Seoul, Online, Coronavirus, Confirmed Cases
6	Remote healthcare and Science and Technology	AI, Artificial intelligence, Untact, Big data, Cloud
7	Social Effects of remote healthcare	New Deal, Jobs, Korean Version, COVID-19, Infectious Diseases

IV. DISCUSSION

In this study, we analyzed keywords related to telemedicine and visualized the relationships between them in a network format. Through this, we were able to explore the trends and implications for the adoption of telemedicine before and after the COVID-19 pandemic. Particularly, we found that the keyword "medical law" had the highest centrality in both networks. This indicates that telemedicine is currently constrained by medical laws and suggests the need for policy discussions to improve this situation [15,16]. Additionally, keywords such as "Ministry of Health and Welfare," "United States," "China," and "Japan" appeared frequently in both networks for similar reasons, indicating significant discussions regarding telemedicine in each country [17-20]. In summary, this study confirms the importance of the legality of medical laws for the expansion and development of telemedicine.

In this study, we also observed that various forms of medical technology and artificial intelligence are closely linked with the activation of telemedicine. Keywords such as "Seoul," "network," "artificial intelligence," "big data," and "Deputy Prime Minister" were connected in the network. These connections indicate a strong association between telemedicine and scientific technology and emphasize the importance of technology in the future of the medical field [21, 22].

Frequency analysis revealed that after the COVID-19 pandemic, not only keywords related to COVID-19 but also keywords related to online, non-face-to-face, and remote consultations increased. This indicates an increased need for telemedicine due to the COVID-19 pandemic. Moreover, these keywords carry higher weights compared to before, highlighting the impact of COVID-19 on the medical and health sectors [23, 24].

Furthermore, through LDA topic modeling, we identified three topics before the COVID-19 pandemic and seven topics after. This revealed various subjects such as telemedicine and scientific technology, global telemedicine, and telemedicine and medical laws. This diversity demonstrates the rapid activation of telemedicine and the ongoing discussions and research from various aspects [25, 26].

In synthesizing these findings, it becomes evident that the COVID-19 pandemic has amplified awareness and underscored the necessity of telemedicine. Nevertheless, legal constraints within the medical field persist, necessitating future discourse and policy development to rectify these issues. Moreover, comprehensive research exploring the social impacts and equity of telemedicine from multifaceted perspectives remains a critical need.

V. CONCLUSION

This study has established the role of the COVID-19 pandemic as an accelerating force in the transformation of the telemedicine sector, highlighting the importance of adaptive medical regulations and the application of advanced scientific technology, with our big data analysis emphasizing the growing necessity of telemedicine and the essential nature of structured policies for its effective deployment. Future research should build on this groundwork by integrating an expanded range of big data sources, such as social media, healthcare forums, and patient surveys, while implementing longitudinal studies to thoroughly characterize the long-term sustainability and patterns of telemedicine utilization.

Further investigations are warranted to probe the public's acceptance of telemedicine and its incorporation into health systems, necessitating the application of sophisticated opinion analysis methods—including sentiment analysis—and the execution of comparative international studies. Such exploratory research is necessary to delineate strategies that are sensitive to cross-cultural differences and capable of adjusting to the evolving healthcare demands informed by global technological trends and shifting societal expectations in the aftermath of the pandemic.

ACKNOWLEDGMENT

This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF- RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

REFERENCES

- [1] "K. S. Zachrison, K. M. Boggs, E. M. Hayden, J. A. Espinola, and C. A. Camargo Jr, "Understanding barriers to telemedicine implementation in rural emergency departments," *Annals of Emergency Medicine*, vol. 75, no. 3, pp. 392-399, 2020. <https://doi.org/10.1016/j.annemergmed.2019.04.026>.
- [2] L. S. Wilson and A. J. Maeder, "Recent directions in telemedicine: review of trends in research and practice," *Healthcare Informatics Research*, vol. 21, no. 4, pp. 213-222, 2015. <https://doi.org/10.4258/hir.2015.21.4.213>.
- [3] W. H. Bos, A. van Tubergen, and H. E. Vonkeman, "Telemedicine for patients with rheumatic and musculoskeletal diseases during the COVID-19 pandemic; a positive experience in the Netherlands," *Rheumatology International*, vol. 41, no. 3, pp. 565-573, 2021. <https://doi.org/10.1007/s00296-020-04746-5>.
- [4] A. Elawady, A. Khalil, O. Assaf, S. Toure, and C. Cassidy, "Telemedicine during COVID-19: a survey of Health Care Professionals' perceptions," *Monaldi Archives for Chest Disease*, vol. 90, no. 4, 2020. <https://doi.org/10.4081/monaldi.2020.1445>.
- [5] K. B. Min, S. H. Song, and J. Y. Min, "Topic modeling of social networking service data on occupational accidents in Korea: latent dirichlet allocation analysis," *Journal of Medical Internet Research*, vol. 22, no. 8, e19222, 2020. <https://doi.org/10.2196/19222>.
- [6] G. Lindelöf, T. Aledavood, and B. Keller, "Dynamics of the Negative Discourse Toward COVID-19 Vaccines: Topic Modeling Study and an Annotated Data Set of Twitter Posts," *Journal of Medical Internet Research*, vol. 25, e41319, 2023. <https://doi.org/10.2196/41319>.
- [7] F. Rabitz, A. Telešienė, and E. Zolubienė, "Topic modelling the news media representation of climate change," *Environmental Sociology*, vol. 7, no. 3, pp. 214-224, 2021. <https://doi.org/10.1080/23251042.2021.1913991>.
- [8] E. Mayor and A. Miani, "A topic models analysis of the news coverage of the Omicron variant in the United Kingdom press," *BMC Public Health*, vol. 23, no. 1, 2023. <https://doi.org/10.1186/s12889-023-13824-1>.
- [9] Z. Li, W. Shang, and M. Yan, "News text classification model based on topic model," in 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, Japan, 2016, pp. 1-5. <https://doi.org/10.1109/ICIS.2016.7550881>.
- [10] X. Hu, "News hotspots detection and tracking based on LDA topic model," in 2016 International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 2016, pp. 248-252. <https://doi.org/10.1109/PIC.2016.7949534>.
- [11] H. Liu, X. Chen, and X. Liu, "A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis," *IEEE Access*, vol. 10, pp. 32280-32289, 2022. <https://doi.org/10.1109/ACCESS.2022.3142238>.
- [12] G. Eom and H. Byeon, "Development of keyword trend prediction models for obesity before and after the COVID-19 pandemic using RNN and LSTM: analyzing the news big data of South Korea," *Frontiers in Public Health*, vol. 10, 894266, 2022. <https://doi.org/10.3389/fpubh.2022.894266>.
- [13] P. K. Bogović, A. Meštrović, S. Beliga, and S. Martinčić-Ipšić, "Topic Modelling of Croatian News During COVID-19 Pandemic," in 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2021, pp. 1044-1051. <https://doi.org/10.23919/MIPRO52101.2021.9596908>.
- [14] Pratinida, T. M., & Setyohadi, D. B. (2021). "Automatization News Grouping Using Latent Dirichlet Allocation for Improving Efficiency," *International Journal of Innovative Computing, Information and Control*, vol. 17, no. 5, pp. 1643-1651.
- [15] I. Akef, J. S. M. Arango, and X. Xu, "Mallet vs GenSim: Topic modeling for 20 news groups report," *University of Arkansas Little Rock Law Journal*, vol. 2, 39205, 2016.
- [16] Y. S. Choi, "A study on the introduction of telemedicine-Coronavirus disease 2019 and the need for the introduction of telemedicine," *International Law Review*, vol. 12, no. 1, pp. 113-137, 2020.
- [17] A. de Oliveira Andrade et al., "On the use of telemedicine in the context of COVID-19: legal aspects and a systematic review of technology," *Research on Biomedical Engineering*, vol. 38, pp. 209-227, 2022. <https://doi.org/10.1007/s42600-021-00141-4>.
- [18] I. Gareev et al., "The opportunities and challenges of telemedicine during COVID-19 pandemic," *Frontiers in Bioscience (Elite Edition)*, vol. 13, no. 2, pp. 291-298, 2021.
- [19] A. Miyawaki, T. Tabuchi, M. K. Ong, and Y. Tsugawa, "Age and social disparities in the use of telemedicine during the COVID-19 pandemic in Japan: cross-sectional study," *Journal of Medical Internet Research*, vol. 23, no. 7, e27982, 2021. <https://doi.org/10.2196/27982>.
- [20] J. A. Betancourt, M. A. Rosenberg, A. Zevallos, J. R. Brown, and M. Mileski, "The impact of COVID-19 on telemedicine utilization across multiple service lines in the United States," *Healthcare*, vol. 8, no. 4, 380, 2020. <https://doi.org/10.3390/healthcare8040380>.
- [21] Y. T. Shen, L. Chen, W. W. Yue, and H. X. Xu, "Digital technology-based telemedicine for the COVID-19 pandemic," *Frontiers in Medicine*, vol. 8, 646506, 2021. <https://doi.org/10.3389/fmed.2021.646506>.
- [22] T. A. Suleiman and A. Adinoyi, "Telemedicine and Smart Healthcare—The Role of Artificial Intelligence, 5G, Cloud Services, and Other Enabling Technologies," *International Journal of Communications, Network and System Sciences*, vol. 16, no. 3, pp. 31-51, 2023.
- [23] D. M. Mann, J. Chen, R. Chunara, P. A. Testa, and O. Nov, "COVID-19 transforms health care through telemedicine: evidence from the field," *Journal of the American Medical Informatics Association*, vol. 27, no. 7, pp. 1132-1135, 2020. <https://doi.org/10.1093/jamia/ocaa072>.
- [24] A. Ramaswamy et al., "Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study," *Journal of Medical Internet Research*, vol. 22, no. 9, e20786, 2020. <https://doi.org/10.2196/20786>.
- [25] G. S. Mosnaim, H. Stempel, D. Van Sickle, and D. A. Stempel, "The adoption and implementation of digital health care in the post-COVID-19 era," *Journal of Allergy and Clinical Immunology: In Practice*, vol. 8, no. 8, pp. 2484-2486, 2020. <https://doi.org/10.1016/j.jaip.2020.07.013>.
- [26] P. Agarwal et al., "Adoption, feasibility and safety of a family medicine-led remote monitoring program for patients with COVID-19: a descriptive study," *Canadian Medical Association Journal Open Access*, vol. 9, no. 2, E324-E330, 2021. <https://doi.org/10.9778/cmajo.20200250>.