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

<table>
<thead>
<tr>
<th>Category (Number)</th>
<th>Daily newspaper name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional journals (2)</td>
<td>Electronic Newspaper, Digital Times</td>
</tr>
<tr>
<td>Broadcasting company (5)</td>
<td>KBS, SBS, MBC, YTN, OBS</td>
</tr>
<tr>
<td>National daily newspaper (10)</td>
<td>Kyunghyang Shinnun, Chosun Ilbo, Donga Ilbo, Hankook Ilbo, Hankoreh, JoongAng Ilbo, Kookmin Ilbo, Munhwa Ilbo, Naeil Shinnun, Segye Ilbo, Seoul Shinnun</td>
</tr>
</tbody>
</table>

TABLE II. NUMBER OF ARTICLES ANALYSIS

<table>
<thead>
<tr>
<th></th>
<th>Before the COVID-19 pandemic</th>
<th>After the COVID-19 pandemic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of articles</td>
<td>5,232</td>
<td>23,884</td>
<td>29,126</td>
</tr>
<tr>
<td>Number of excluded articles</td>
<td>92</td>
<td>652</td>
<td>744</td>
</tr>
<tr>
<td>Number of analysis articles</td>
<td>5,140</td>
<td>23,232</td>
<td>28,372</td>
</tr>
</tbody>
</table>

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. 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>
<thead>
<tr>
<th>Table III. Top 15 Keywords Before and After the COVID-19 Pandemic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before the COVID-19 pandemic</strong></td>
</tr>
<tr>
<td>Word</td>
</tr>
<tr>
<td>Telemedicine</td>
</tr>
<tr>
<td>Korea</td>
</tr>
<tr>
<td>United States</td>
</tr>
<tr>
<td>Telemedicine</td>
</tr>
<tr>
<td>China</td>
</tr>
<tr>
<td>Job</td>
</tr>
<tr>
<td>Korea</td>
</tr>
<tr>
<td>Japan</td>
</tr>
<tr>
<td>Seoul</td>
</tr>
<tr>
<td>Medical Law</td>
</tr>
<tr>
<td>Big Data</td>
</tr>
<tr>
<td>A pilot project</td>
</tr>
<tr>
<td>Cheong Wa Dae</td>
</tr>
<tr>
<td>Smartphone</td>
</tr>
<tr>
<td>Competitiveness</td>
</tr>
</tbody>
</table>

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**

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

**TABLE V. TOPICS MODELING AFTER THE COVID-19 PANDEMIC**

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

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 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 legal validity 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


