Analyzing Research Trends in Digital Therapeutic: A Topic Modeling Approach

Eunsun Choi[1], Namje Park[2,*]

[1] Major in Computer Education, Faculty of Science Education, Graduate School, Jeju National University

[2]Department of Computer Education, Teachers College, Jeju National University
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 63294, Korea

choi910624@jejunu.ac.kr[1], namjepark@jejunu.ac.kr[2]

Corresponding author[*]: mobile Phone: +82-10-6397-1549,namjepark@jejunu.ac.kr

Article Info
Page Number: 464 – 471
Publication Issue:
Vol. 71 No. 3 (2022)

Abstract

Digital Therapy (DTx) is in the spotlight in the medical community as an efficient treatment method that allows low cost for patients and timely intervention for doctors. This paper aims to find the development direction of DTx by analyzing the trend of DTx research until April 2022.We first collected the paper mentioned DTx in the title or subject word of the paper, or abstract, in Springer through web crawling. We grouped synonyms into one word, and excluded foreign language papers that could not be analyzed for data analysis of the collected documents. The approach used for data analysis is a topic modeling method used to find the subject word in a set of documents, among text mining techniques. Through this, terms frequently used in the literature were found through frequency analysis. We also conducted centrality analysis on the top 10 words with high frequency identified through frequency analysis. In addition, we created groups of words by topic modeling to determine which words were clustered.We conducted three analyses on 179 literatures related to DTx through this study, and identified the research trends of DTx through the results. The first analysis is frequency analysis. Through this, it was found that a lot of research is being conducted on technology, datum, system, application, and App based on AI as well as health and patient. The second analysis is centrality analysis. The centrality between each word was not strongly revealed. The third analysis is topic modeling analysis. We created five topic groups, and device and tool were included in 2 groups. Along with technical research, we confirmed that research on the economic value of DTx is also being actively

conducted.This study analyzed DTx-related papers published in Springer and has limitations in viewing them as overall research trends for DTx. However, not only the keywords of papers conducted in most research trends analysis, but also the title, abstract, and keywords of the papers were used for analysis to derive more detailed research results. This study is on research trends. It can be applied to future policy research or development directions related to DTx.

Keywords— Digital Therapeutics(DTx), Research Trend Analysis, Web Crawling, Frequency Analysis, Centrality Analysis, Topic Modeling, Text Mining

Article History
Article Received: *12 January 2022*
Revised: *25 February 2022*
Accepted: *20 April 2022*
Publication: *09 June 2022*

1. Introduction

Digital transformation is taking place throughout the industry. Among them, the trend of accepting digital technology in the medical field is noteworthy. Medical technology covers diagnosis, rehabilitation, and treatment to improve human health and prevent diseases. In the medical field that deals with human life, accurate data analysis is required. In addition, unverified advanced technologies are challenging to apply to medical care. It has not been easy to meet integrated digital technology and medical technology for these reasons. However, with the outbreak of the digital revolution in the 1980s, digital technology developed explosively, and the speed and influence of development were brutal compared with anything previously [1]. Accordingly, digital technology is also being accepted in the medical field that can save human lives or improve the quality of life through disease treatment. These changes occur inside and around medicine, such as pharmacology, biochemistry, and biotechnology. In addition, changes outside of medicine, such as 3D computing and cloud computing, which seem unrelated to medicine at first glance, are currently bringing destructive changes to the medical world [2].

Digital Therapeutics (DTx) is considered a type of Software as Medical Device based software program. It refers to software medical devices that prevent and manage diseases or disorders and provide therapeutic interventions to patients for treatment. Globally, the DTx market was worth $2.1 billion in 2018. It is expected to grow at an average annual rate of about 20% to over $23 billion in 2030 [3]. There will be no explanation for the future of medicine without digital technology.

In this paper, we intend to analyze research trends by collecting research related to DTx, which has just entered the starting point of research. To this end, we searched for papers related to DTx and collected keywords on Springer, which can search for various academic works such as papers, books, reports, and institutional outputs. We extracted keywords by refining the collected data. Moreover, we identify the association and centrality between critical keywords and present topic trends for DTx through time series regression analysis.

2. Related Background Research

DTx refers to a high-quality software system that helps patients prevent diseases and illnesses in advance, including health tips and behavior corrections, exercise programs, and medication intake alerts [4]. DTx needs to provide clinically proven results, which sets DTx apart from simple wellness and health apps or reminder schedule apps. According to the Digital Therapy Alliance (DTA), DTx refers to evidence-based, clinically evaluated software for preventing and treating a wide range of diseases that provides patients with direct medical commitment [5]. Various techniques such as Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) are used in this treatment. DTx functions as follows:

·Preventive care activities for people with chronic or severe illness

·Sources of health information that physicians use to make diagnosis and treatment decisions to patients

·Digital programs combined with independent or traditional treatments to treat addiction

·Tools for monitoring and continuous improvement of patient health

DTx is often confused and used in similar terms such as Digital Health and Digital Medicine [6]. These terms have overlapping parts. However, they involve different characteristics and intervention in treatment. Digital health is the broadest concept among the three terms, and refers to apps or health care systems used for collecting digital-based well-being-related data. It is not classified as a medical device, therefore there is no need for medical grounds or regulation. Digital medicine refers to evidence-based hardware or software for measuring health information, and it is characterized by various requirements for regulation, although medical evidence is necessary. Meanwhile, DTx is a digital-based treatment that supports clear medical verification when applied to humans. It is the strictest and narrowest of the three terms. Table 1 refers the difference of three terms.

Table.1.Difference between DTx, Digital Health, and Digital Medicine

	Digital Health		
		Digital Medicine	
			DTx
Relation	broad concept covering both digital medicine and DTx	sub-concept of digital health	sub-concept of digital medicine
Product	health or wellbeing related products	Products that measure human health	Products that deliver therapeutic intervention
Evidence	not always	required	required
Regulations	not required	vary	required

In 2017, reSET, a smartphone app that treats Pear Therapy's Sub-stance Use Disorder, was approved by the U.S. Food and Drug Administration for the first time [7]. As such, DTx

is a relatively recent new technology, but it is expected to provide various advantages such as reducing treatment costs for patients in the future and enhancing the efficiency of treatment delivery due to timely intervention for doctors. Hence, there is a high possibility of development in the future. Therefore, it is important to understand the research trends and development directions of DTx.

3. Materials and methods

3.1.*Data Collection and Preprocessing*

We wanted to collect DTx-related documents among all Springer's documents through web scraping. Therefore, among all the studies, we searched for studies containing the word 'Digital Therapeutics' in the thesis title, keyword, and abstract. Among the results, we excluded literature containing the word 'digitalis', which is similar in form to digital but has different meanings. Through this process, we were able to collect 179 documents. Since the literature volume is not large, we tried to use keywords and the title of the paper, keywords, and abstracts for analysis. Fig. 1 refers to the number of papers collected by year. According to Fig. 1, the number of research literature has increased significantly since 2019.Data for 2022 is the number of papers until April 2022.

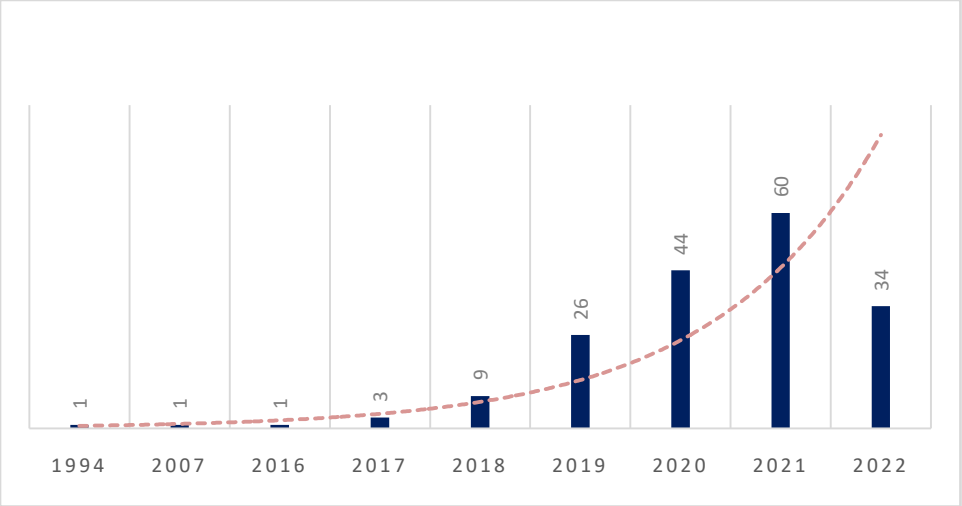

Fig.1.Number of PublicationsUsed in the Analysis per Year

Data preprocessing is essential for data analysis [8]. In this document, the data purification process that processed synonyms, thesaurus, and exclusions is as follows.

1. Since we included the thesis title and abstract in the analysis in addition to keywords consisting of nouns, meaningless pronouns, verbs, adverbs, adjectives, prepositions, conjunctions, and interjections were all excluded from the analysis.
2. Numbers that are difficult to understand with simple words such as '1960s' and 'Arabia 2' were excluded from the analysis.
3. Foreign languages other than English were converted to English and included or excluded from the analysis.
4. Synonyms were treated as one word. For example, COVID, COVID19, and

COVID19Pandemic were all treated as COVID19.

5. Duplicate words identified due to abbreviations were treated as the same word. For example, Artificial Intelligence and AI were processed by AI.

3.2.*Data Analysis*

We purified the data collected through web crawling and then conducted data analysis. The method used for data analysis is topic modeling. Topic modeling is one of the text mining techniques used to find critical topics in a set of documents and discover the implications of the text [9]. We used this technique to perform keyword analysis to extract the main words and perform frequency analysis.Through this process, the association of the analyzed keywords was analyzed. In addition, centrality analysis was performed to grasp how key the primary control nodes play a crucial role in the social network. Finally, through topic modeling, we created a group of topics and visualized them through a network map.The tool used for analysis is NetMiner 4.0.

4. Results

4.1.*Frequency Analysis*

We extracted 2,673 words from thesis titles, abstracts, and keywords of the collected literature. Frequency analysis was performed on the extracted words, and the words were sorted in the order of the most frequent occurrences. The top 30 words are summarized, 'health' occupied the most frequency with 239 times, 'patient' 176 times, 'technology' 124 times, 'datum' 117 times, and immediately followed by 'treatment' 116 times was counted. A characteristic feature is that 'app' appeared 54 times, 'device' appeared 52 times, and 'AI' appeared in the literature with a high frequency of 44 times. From the frequently appearing words, it can be seen that various studies are being conducted to treat the health of people and patients through technology and data, and the number of treatments through apps or digital devices applying AI technology is increasing. Table 2 shows the frequency analysis results up to the top 30 of the extracted words.

Table.2.Top 30 Most Frequent Words in Research Papers

Rank	Topic	Frquen-cy	Rank	Topic	Frequen-cy	Rank	Topic	Frequen-cy
1	health	239	11	healthcare	62	21	outcome	49
2	patient	176	12	application	62	22	development	49
3	technology	124	13	quality	59	23	research	47
4	datum	117	14	intervention	56	24	AI	44
5	treatment	116	15	tool	54	25	evidence	42
6	care	110	16	management	54	26	disorder	41
7	study	102	17	App	54	27	product	40
8	disease	79	18	digital	53	28	medicine	40
9	system	77	19	device	52	29	solution	34
10	use	74	20	approach	51	30	risk	27

4.2.*Centrality Analysis*

We conducted a centrality analysis to determine how crucial the topics in the study on DTx are playing within the study among the studies published in Springer. We analyzed nodes' degree centrality, between centrality, and closeness centrality. The top 10 topics for degree centrality were health, patient, treatment, datum, technology, care, study, disorder, use, and system. It can be seen that these words have high connectivity with other nodes. Betweenness centrality appeared in the order of health, patient, AI, datum, treatment, care, study, technology, digital, and death. We found that the top 10 words of betweenness centrality frequently appeared in combinations of shortest paths with nodes other than themselves. The top topics of closeness centrality were health, patient, datum, care, treatment, technology, use, noisease, application, and study. It can be seen that the paths to other nodes are arranged in a short order. Table 3 shows the results of centrality analysis for topics.

Table.3.Centrality Analysis Results

Rank	Topic	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	health	0.094	0.145	0.404
2	patient	0.084	0.088	0.389
3	technology	0.062	0.041	0.372
4	datum	0.058	0.059	0.379
5	treatment	0.053	0.052	0.372
6	care	0.052	0.049	0.377
7	study	0.046	0.049	0.360
8	disease	0.046	0.040	0.367
9	system	0.045	0.030	0.360
10	healthcare	0.040	0.025	0.357

4.3.*Topic Modelling Analysis*

Topic modeling was performed on the collected data for detailed topic search for DTx. We performed iterative topic modeling to find the optimal number of topics for which the key words belonging to each topic can form a common semantic group, and finally selected a set of five topics.

The first topic group includes participants, controls, devices, effects, and trials. It shows that clinical trials are being conducted on various groups in the DTx study, and in particular, it can be seen that the test of a therapeutic agent using a device is performed. The second topic group includes AI, blood, machine, language, etc. We found that AI is used in research with a high frequency related to blood pressure. App, solution, tool, opacity, device, etc. are grouped in the third topic group. The 'device' connects the common node of topic-3 and topic-1. In addition, we show that digital-based apps and devices are being used as tools for digital therapy. A fourth topic group includes disorders, person, training, tools, rehabilitation, and ADHD. We found the 'tool' as a common node for topic-3 and topic-4. DTx is also used in the treatment and recovery of various cognitive disorders, and is particularly attracting attention as a treatment for ADHD. The fifth topic group is connected to quality, product,

strategy, and need. Like topic-2, there were no nodes connected to other topics. In this group, the node size of the quality and product is significant, especially the thick link of quality. It refers the economic side of DTx. Many researchers have marketability in mind as a DTx development strategy, and it can be understood as a need for large-scale investment. Fig. 2 is a 2-mode spring topic map formed by topic modeling analysis.

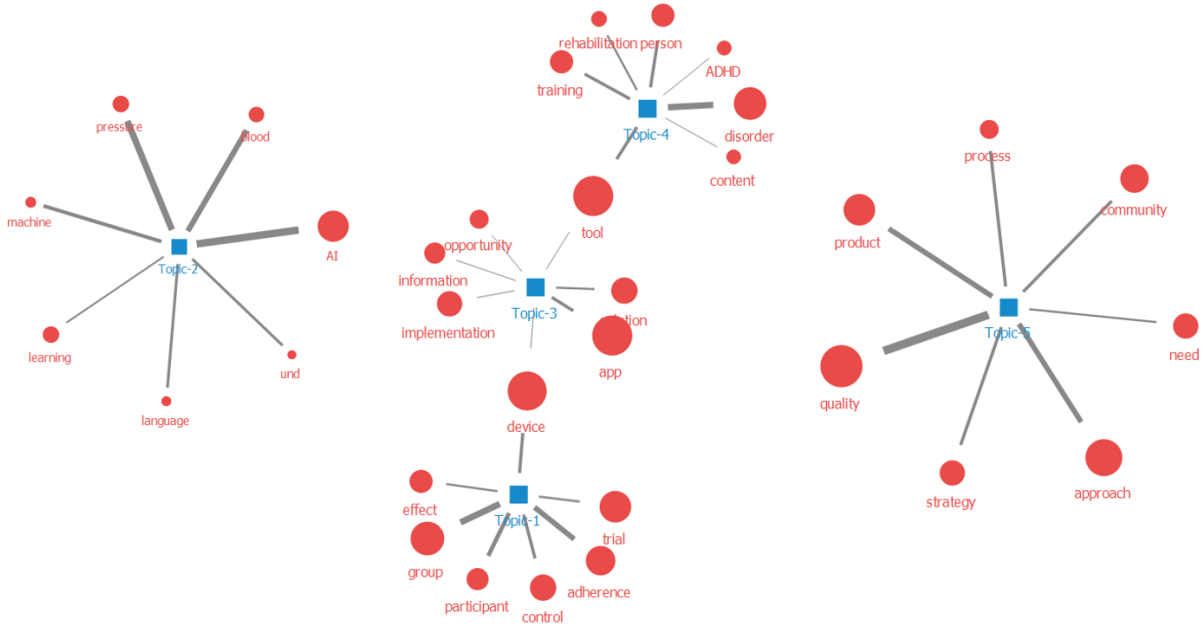

Figure 2.Topic Network Map

According to the number of literature corresponding to the topic group, Topic-3 is the largest with 58, followed by Topic-5 with 45. In addition, 27 documents for Topic-2, 25 documents for Topic-1, and 24 documents for Topic-4 are related. These results show that DTx research trends to date are the most active in research on various digital-based devices, and many researchers are also paying attention to the economic needs. Table 4 shows the number of literature related to the core topics of the study accompanying the topic group.

Table.4.Core Topics and the Number of Associated Papers

Group	Node	No. of Documents
Topic-1	group, adherence, participant, control, device, effect, trial	25
Topic-2	AI, pressure, blood, machine, language, und, learning	27
Topic-3	App, solution, implementation, information, tool, opportunity, device	58
Topic-4	disorder, person, training, tool, rehabilitation, content, ADHD	24
Topic-5	quality, product, approach, strategy, process, community, need	45

5.Conclusion

Digital-based medical interventions will gradually increase in the future. This paper aims to understand the research trends of DTx to date. Research papers published in Springer related to DTx began to soar in 2018, and are continuously increasing. We first analyzed which topics were mentioned in DTx-related studies, and various topics such as health, patient, technology, and datum were detected. As a result of analyzing the centrality of the aggregated topics, it was found that the topics of high frequency that appeared in degree centrality, betweenness centrality, and closeness centrality were different. According to the topic modeling analysis, key words were clustered into five groups. The literature, in which digital tools and opportunities are grouped together, has been identified the most, indicating that many researchers think digital technology brings new opportunities for patient treatment. However, since this paper used only Springer's paper for research, it cannot be seen that it deals with the overall research trend of DTx. However, it is significant that we analyzed key words in the literature related to DTx through a topic modeling approach.

6. Acknowledgment

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5C2A04083374). And, this work was supported by the Korea Foundation for the Advancement of Science and Creativity(KOFAC) grant funded by the Korea government(MOE).

6. References

1. R. Nick Bryan. The Digital rEvolution: The Millenial Change in Medical Imaging. Radiology. 2003 Nov;229(2): 299-304. DOI: https://doi.org/10.1148/radiol.2292030485.
2. Sandra Bucci, Matthias Schwannauer, Natalie Berry. The Digital Revolution and its Impact on Mental Health Care. Psychology and Psychotherapy. 2019 Mar;92(2): 277-297. DOI: https://doi.org/10.1111/papt.12222.
3. Sanjivan Gill, Vikita Thakur, Onkar Sumant. 2022 Digital Therapeutics Market. Allied Market Research; 2021 Nov. Available from: https://www.alliedmarketresearch.com/digital-therapeutics-market.
4. Amit Dang, Dimple Arora, Pawan Rane. Role of Digital Therapeutics and the Changing Future of Healthcare. Journal of Family Medicine and Primary Care. 2020 May;9(5): 2207-2213. DOI: https://doi.org/10.4103/jfmpc.jfmpc_105_20
5. Digital Therapeutics Alliance. Digital Therapeutics Definition and Core Principles. Digital Therapeutics Alliance; 2019 Nov. Available from: https://dtxalliance.org/understanding-dtx/
6. Giuseppe Recchia, Daniela Maria Capuano, Neeraj Mistri, Roberto Verna. Digital Therapeutics-What they are, what they will be. ACTA Scientific Medical Sciences. 2020 March;4(3): 1-9. DOI: https://dor.org/10.31080/ASMS.2020.04.0575.
7. Raj Khirasaria, Vikramjit Singh, Angelika Batta. Exploring Digital Therapeutics: The Next Paradigm of Modern Health-care Industry. Perspectives in Clinical Research. 2020 Apr-Jun;11(2): 54-58. DOI: https://dor.org/10.4103/picr.PICR_89_19.
8. Salvador Garcia, Julian Luengo, Francisco Herrera. Data Preprocessing in Data Mining. 2015. Springer. Cham: Switzerland. DOI: https://doi.org/10.1007/978-3-319-10247-4.
9. Ike Vayansky, Sathish A.P. Kumar. A Review of Topic Modeling Methods. Information Systems. 2020 Dec;94: 101582. DOI: https://doi.org/10.1016/j.is.2020.101582.