Proceedings of the First Australian International Conference on Industrial Engineering and Operations Management, Sydney, Australia, December 20-21, 2022

Analysis of Research Trends in Artificial Intelligence and Healthcare Convergence using Text Mining Techniques

Kim See You, Shin Dong Ho
Student and Professor, MY PAUL SCHOOL
12-11, Dowontongmi-gil, Cheongcheon-myeon, Goesan-gun,
Chungcheongbuk-do, Republic of Korea
eavatar@hanmail.net

Abstract

This paper reviews the major research trends on the convergence research of artificial intelligence and healthcare technology. For research, we collected 15,260 English-language articles on AI and medical-related topics and performed text mining techniques on Scopus from 1963 to 55 years. As a result, seven core research topics, including 'AI for clinical decision support system (CDSS)', 'AI for medical imaging', 'Internet of Things (IoHT)', 'medical big data analysis', and 'medical medicine' were selected. defined Robotics", "Blockchain in Healthcare" and "Evidence-Based Medicine (EBM)". In this paper, text mining including Text Analysis, Frequency Analysis, Topic Modeling on LDA (Latent Dirichlet Allocation), Word Cloud, Ego Network Analysis, etc., the technique was performed.

Keywords
Artificial Intelligence, Healthcare, Text mining, Topic Modeling, Ego Network Analysis,

1. Introduction
Recently, convergence and complex research of state-of-the-art innovative technologies have been actively promoted in various academic fields around the world. Innovations developed through such convergence and complex research are bringing about disruptive innovation across all industries, and there is a huge wind of change throughout the industry, with the emergence of completely new business models. This change is also the core foundation of the 4th Industrial Revolution, which has recently become a hot topic. In other words, the 4th Industrial Revolution can also be seen as the emergence of an "age of convergence complex technologies" that overcome challenges in social and scientific technology as new levels of technology that did not exist were developed.

Since the confrontation between Lee Se-dol and AlphaGo in March 2016, the public's awareness of AI has increased, and policy interest in the fourth industrial revolution is also very high after the presidential election in May following the government's announcement of comprehensive measures for the intelligent information society in 2017.(Choi and Oh, 2017).

To lead the future health care industry, the government has identified six core projects as ① establishment of a healthcare big data showcase, ② establishment of an ecosystem for the healthcare industry, ③ establishment of a smart clinical trial center, ④ development of new drugs using artificial intelligence, ⑤ smart convergence, complex medical care Device development and system improvement, and ⑥ promotion of in vitro diagnostic device market entry are selected and discussed.(Healthcare Special Commitee, 2018).

The government is working on six key projects to lead the future healthcare industry. The medical community is actively promoting various convergence and complex projects such as the establishment of a next-generation hospital information system and HIS, which integrates and manages all information generated in hospitals such as clinical, medical services, and hospital administration using artificial intelligence technology and big data.(Heo, 2017).

Seoul Samsung Hospital, which is leading the field of precision medicine, introduced the next-generation HIS, Digitalized Analysis & Research Window for Integrated Knowledge in 2016, and Seoul Asan Hospital operated the next-generation HIS, Asian Medical Information System 3.0, in 2017.(Shin et al., 2018).

Korea University Medical Center established KUEM, a next-generation electronic medical record, in 2017 and developed Aibril Antibiotics Advisor; 3A in 2018. Taken together, it can be summarized that the emergence of innovative technologies based on artificial intelligence and big data in the healthcare field will have a very large economic and social ripple effect. (Lee, 2018).

As such, the government, medical (hospital), and industry (enterprise) are interested in artificial intelligence, healthcare convergence, and complex fields, but healthcare policies and R&D are still in the introduction stage, and research to support detailed trends and R&D strategy planning is insufficient. Therefore, this paper aims to systematically understand the convergence and complex R&D trends of artificial intelligence-based technology and healthcare-related technology, which are rapidly growing by inducing remarkable innovation in all industries, including healthcare.

To this end, text mining techniques and network analysis techniques were applied to the abstracts of papers collected by Scopus to identify key research topics that are drawing keen attention from researchers in this field worldwide and analyze them in depth. The research results of this paper can be used to select trendy core research topics for researchers in this field in the future and to establish the government's strategy for promoting R&D projects in the healthcare sector. (Christensen et al., 2015; Kim, 2017; Choi, 2017).

2. Body

2.1 Analysis Target

To understand research trends in artificial intelligence, healthcare convergence, and complex fields, this study collected 16,370 English-written papers published for about 55 years from 1963 to 2018 in Elsevier's Scopus. For text analysis, a total of 15,260 papers were selected as the final analysis targets, excluding 851 papers without abstracts, 196 papers without author information, and 63 papers that were not suitable for duplication and analysis.

2.2 Analysis Method

This research conducted text analysis, major word frequency analysis by era, LDA-based topic modeling, word cloud, and ego network analysis on the convergence of artificial intelligence and healthcare, and English thesis abstracts in complex fields collected by Scopus. The collected abstract of the paper was standardized for text analysis, and unnecessary information such as special characters was deleted through the preprocessing process. Abbreviations, singular, and plural are added to the synonym dictionary and created based on 15,260 thesis keywords thesaurus, Defined Words, and terminology dictionaries were applied. For text analysis, the OpenNLP morpheme analyzer provided by Netminer 4.4.2 was used (Figure 1, Table 1)

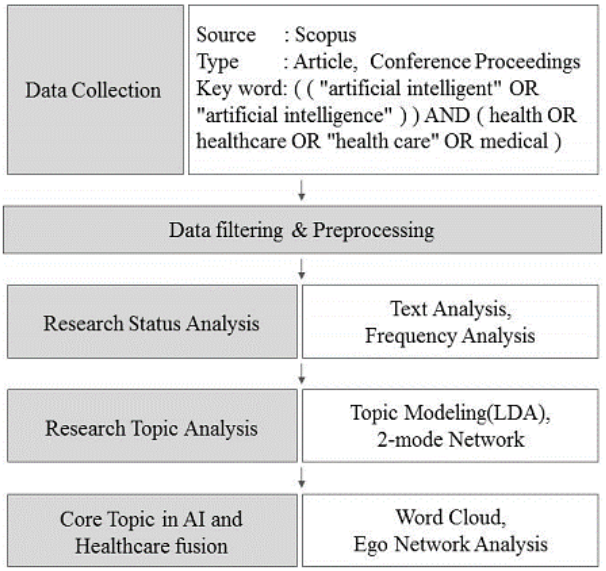

Figure 1. Analysis Flowchart of Study

Table 1. Customer Dictionary

Thesaurus	synonym	medical doctor → doctor, surgeon, physician, medical professional etc.
	abbreviated form	Hospital Information System(HIS) → hospital information system, his
	singular	informations → information
Defined word		Hospital Information System(HIS)
Stop Word		method, study, ieee etc.

LDA-based topic modeling is a generative probabilistic model utilization technique for analyzing text data, which is unstructured data, to find meaningful major topics (hereinafter referred to as 'research topics').(Blei, 2012). Topic modeling can infer meaningful research topics through pattern analysis of words related to the context from vast and unstructured text data. Because of these characteristics, topic modeling is widely used as a research methodology to extract major research topics by analyzing the text of the thesis abstract and to grasp research trends over time.(Nam, 2016).

In topic modeling, the values that researchers can designate in advance include the number of research topics, *a* value, *b* value, and Gibbs sampling value. Steyvers and Griffiths recommended finding the best value by fixing the value of *a* to 50/k (number of research topics) and setting the value of b differently, such as 0.1, and 0.01. In this study, to perform LDA-based topic modeling, the number of research topics) *a* value 7.0 (50/7) and *b* value 0.1 were set, and Gibbs sampling was repeated 1000 times.(Choi, 2017).

Network Analysis is defined as a link between a particular type of node and a node as a way of visualizing complex relationships so that they can be more intuitively identified. Network analysis may analyze the entire network, but it may be analyzed by setting individual nodes in the network as ego. Therefore, analyzing neighboring nodes that form a direct connection relationship by setting a specific node to ego according to the purpose of the analysis is called ego network analysis. Ego networks can be usefully used to observe the partial characteristics of networks connected to the Egonodes.(Choi, 2017; Lee, 2012).

Important indicators for understanding the structural characteristics of the network include Degree Central and Betweenness Central. Connection centrality is an index that measures how many connections nodes in the network have, and it can be seen that the higher the connection centrality, the more actively related research is being conducted. Mediation centrality is an indicator that measures how much one node performs the role of an intermediary between other nodes in the network. (Lee, 2012).

It can be seen that nodes with higher mediation centrality have the influence to control the flow of research and play the role of intermediaries and gatekeepers. The network analysis method is presented as one of the methods to supplement the existing qualitative analysis, which is difficult to predict the future or reflect trends. In this study, 'To more effectively analyze research trends in artificial intelligence, healthcare convergence, and complex fields, we analyzed the word cloud for the top 100 words with a high probability of allocation and analyzed the Ego network for each research topic. (Lee, 2012; Yoon et al., 2018; Jeong, 2010).

3.1 Analysis of Research Trends by Period

In this research, text analysis and frequency analysis were conducted to identify keywords that appeared in the paper for about 55 years from 1963 to 2018. Of the 11,008 words extracted from a total of 15,260 papers collected, the top 10 words with high frequency were extracted by period for 3,949 words that appeared more than TF-IDF 0.3 and more than 8 papers.

In this research, LDA-based topic modeling was performed on 3,949 words that appeared in more than 8 papers. As a result of examining the similarity between research subjects to determine the independence of the analyzed main research subjects, it was found to be less than 0.05, suggesting that there is no problem with similarity between research subjects. Also, as a result of analyzing the 2-mode network as shown in Figure 2 for the research topic and

the top 7 words with high assignment probability by research topic, there were no duplicate words, so it is judged that the main research topic was analyzed independently.

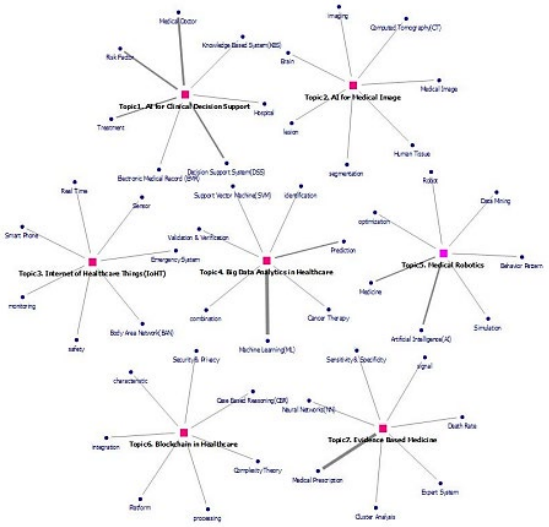

Figure 2. Result of 2-Mode Network (Topic-Word)

As shown in Table 2, the results of the analysis of major research topics organized three keywords with high allocation probability for each research topic in order and presented the weight of documents with the most representative research topics. The main research topic name was selected in consideration of the connection between major keywords with a high probability of allocation by research topic. Among the 15,260 papers, Artificial Intelligence (AI) for Clinical Decision Support Systems (CDSS) (17%) and Artificial Intelligence (AI) for Medical Image (17%) were found to have the highest share.

Table 2. Result of Topic Modeling

Topic		word	Doc.	%
Topic 1	AI for Clinical Decision Support System(CDSS)	Medical Doctor, Decision Support System(DSS), Risk Factor, Treatment, Knowledge Based System (KBS), Hospital, Electronic Medical Record(EMR)	2586	17%
Topic 2	AI for Medical Image	segmentation, Medical Image, Brain, Human Tissue, Computed Tomography (CT), imaging, lesion	2582	17%
Topic 3	Internet of Healthcare Things(IoHT)	Sensor, monitoring, Real Time, Body Area Network (BAN), safety, Emergency System, Smart Phone	2339	15%

3.2 Analysis Results by Key Research Topic

3.2.1 Topic 1. Artificial Intelligence(AI) for Clinical Decision Support Systems (CDSS)

Artificial Intelligence (AI) for Clinical Decision Support System (CDSS) (hereinafter referred to as 'research topic 1'), which had the highest share of research topics in healthcare, artificial intelligence convergence, and complex fields, was 2,586 (17%) out of 15,260 papers. (Lee et al., 2016).

Research topic 1 refers to an expert system that helps medical personnel make decisions when diagnosing and treating diseases based on patient clinical information. In other words, research topic 1 aids clinical decision-making so that doctors can make precise diagnoses and provides information on patients' drug use in the treatment and prescription stages to prevent drug side effects, helping to improve the quality of medical judgment and medical care. The number of papers by year in Research Topic 1 is increasing as shown. The Compound Annual Growth Rate and CAGR for the period 1975 to 2018 were found to be 14%. (Figure 3).

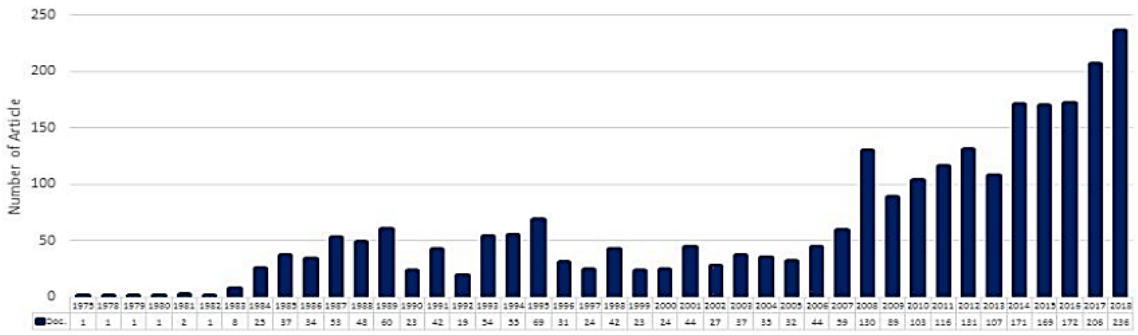

Figure 3. Changes in the Number of Articles (Topic 1)

Figure 4. Result of Word Cloud(Topic 1)

To identify key keywords related to research topic 1, as a result of analyzing word clouds for the top 100 allocation probabilities, Medical Doctor, Decision Support System (DSS), Risk Factor, Treatment, and Knowledge-Based System (KBS) were found to be the main. To examine research topic 1 in more depth, the ego network was analyzed focusing on the Medical Doctor, who showed the highest connection centrality (0.7677) and mediation centrality (0.1532) among the top 100 words of allocation probability.(Figure 4).

As a result of the analysis, the top five words of allocation probability were fully connected graphs, and the Medical Doctor was linked to Nursing, Clinical Decision Support System (CDSS), Medical Prescription, Hospital, Decision Making, and Recommender System, indicating that research related to diagnostic support, treatment decision support, prescription support, information search, and risk.

In addition, it is connected to Technology Acceptance and consultation with low connection centrality and mediation centrality, indicating that research related to technology acceptance of research topic 1 is underway. According to P&S Market Research (2016), the market size of Research Topic 1 was found to have a value of $4.105 billion in 2015, and the CAGR is expected to reach 21.5% between 2016 and 2022.

As such, the market related to research topic 1 is growing rapidly due to increased demand due to the development of medical information systems (HIS) and increased investment in medical information technicians (HIT). Major global companies related to Research Topic 1 include IBM Watson Ontology, Cerner Corporation, McKesson Corporation, General Electric Company, Zynx Health Corporation, and Simens AG.(Healthcare Special Committee, 2018; Kim, 2017).

3.2.2 Topic 2. Artificial Intelligence(AI) for Medical Image

The share of the research topic (hereinafter referred to as 'research topic 2') was 17% (2582) of the total. Research topic 2 means extracting features from medical images, which are unstructured data such as CT, MRI, X-ray, endoscopy, and ultrasound, and converting them into structured numerical data. (Jung, 2018). Although it was analyzed based on anatomical knowledge and clinical experience in the past, the development of deep learning technology in artificial intelligence has enabled Accuracy, Consistency, and Scalability analysis. to show a sharp

increase in the period 2000-2008 It was analyzed that it decreased from 2009 to 2011 and has been increasing again since 2012. The CAGR from 1964 to 2018 was found to be 14%.(Simon,2018).

As a result of analyzing the word cloud targeting the top 100 allocation probabilities to identify the main keywords related to research topic 2, segmentation, MedicalImage, Brain, Human Tissue, and Computed Tomography (CT) showed high importance. Text Network of Research Topic 2 In order to examine the structure in more depth, the ego network) was analyzed as shown in Figure 5, focusing on the segmentation with the highest connection centrality (0.8182) and mediation centrality (0.0904) among the top 100 words in the allocation probability.

As a result, the top 5 words in the allocation probability appeared as a fully connected graph, and segmentation was strongly connected with lesion, organ, vessel, lung, and FeatureExtraction. In addition, since Brain and Computer Tomography (CT), Magnetic Resonance Imaging (MRI), scan, resonance, and MixedReality are closely connected, various medical images are being applied to analysis, and research related to augmented reality is in progress. According to Signify Research, a global medical technology industry consulting firm, the market size of research topic 2 is predicted to reach $2 billion by 2023. (Figure 5).

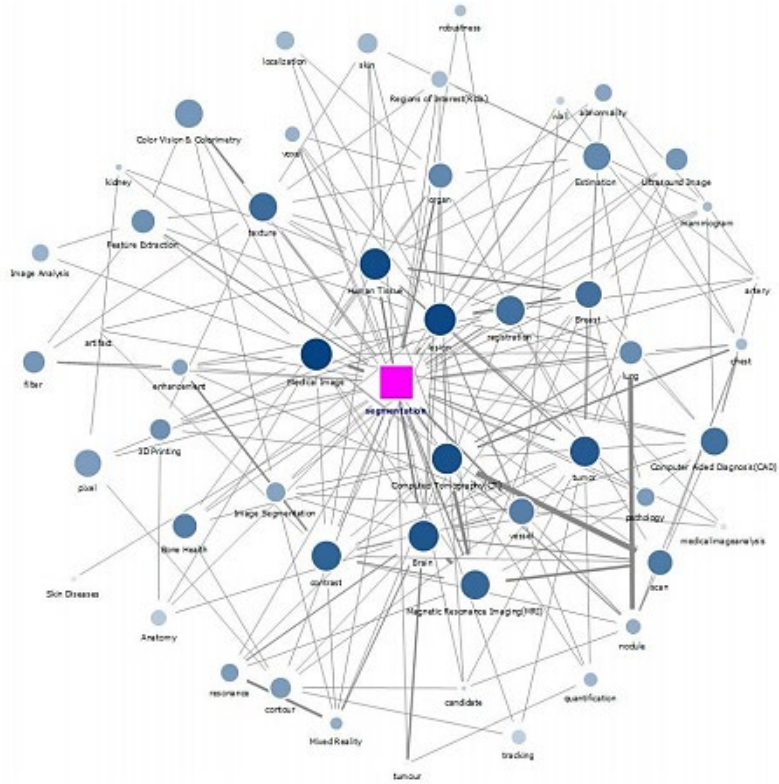

Figure 5. Ego Network Analysis (Topic 2)

3.2.3 Topic 3. Internet of Healthcare Things (IoHT)

The share of the research topic (hereinafter referred to as 'research topic 3') was 15% (2,339 pieces) of the total. Internet of Things, IoT is one of the main technologies of the Fourth Industrial Revolution, which can be described as the relationship between things (services, places, products) and humans based on various platforms, and according to the Harvard Business Review, 28 billion "things" are expected to connect to the Internet.(Song, 2016). Research Topic 3, an IoT-based solution, is quickly accepted by the medical industry and can support Remote Healthcare Monitoring, Healthcare Solutions Using Smartphones, Ambient Assisted Living (AAL) Wearable Devices, etc.(Healthcare Special Committee, 2018).

It is steadily increasing, and the CAGR from 1979 to 2018 was 14%. As a result of analyzing the word cloud for the top 100 words of allocation probability to identify the main keywords related to research topic 3, Sensor, monitoring, Real-Time, Body Area Network (BAN), and safety showed high majorities. To analyze research topic 3 in more depth, the ego network was analyzed for sensors with high central centrality (0.6162) and mediating centrality (0.1255).(Figure 6). The analysis shows that the Sensor is not connected to safety and that it is closely connected to Smart Phones, Accelerometers, the Internet of Things (IoT), Failure, and Human Activity Recognition (HAR) using sensor network technology to conduct personalized health care studies.(Kim, 2017; Rodrigues et al., 2018).

As a result of the analysis, the Sensor appeared as a graph fully connected with the top 4 words with high allocation probability, but it was not connected with safety. It can be seen that personalized health care research is in progress through real-time monitoring using sensor network technology.
MaKinsey predicts that the market size of research topic 3 will have an economic effect of more than USD 170 billion by 2020 due to the reduction of medical costs, improvement of the quality of life of chronic patients, and health monitoring for disease prevention according to research topic 3, MarketAND The market is predicted to be worth $163.2 billion (Rodrigues et al., 2018). Major global companies related to research topic 3 include IBM, Microsoft, Philips, Cisco, Cerner, Apple, and DeepMindHealth.(Rodrigues et al., 2018).

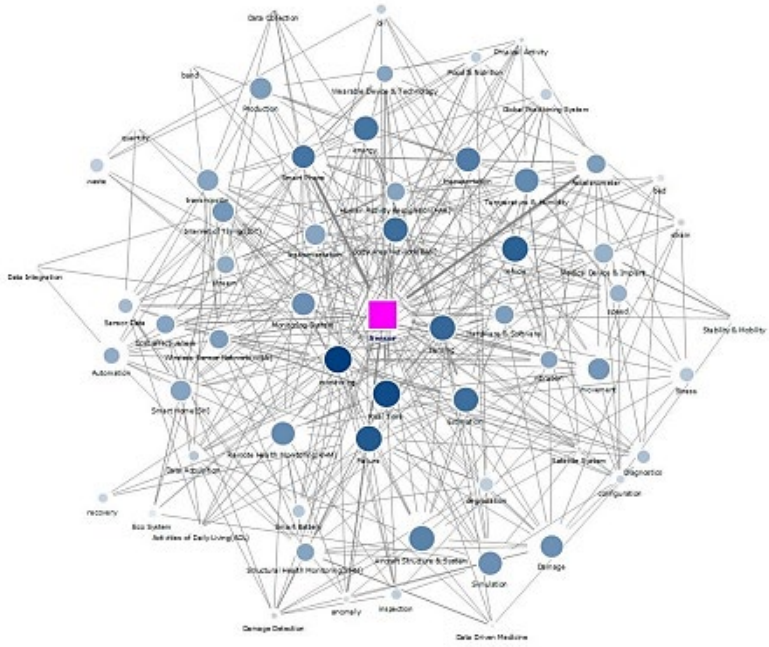

Figure 6. Ego Network Analysis (Topic 3)

3. Conclusion

The purpose of this study is to analyze major research topics and research trends in the era of artificial intelligence and healthcare convergence and to systematically understand the changes in the complex field to see the future direction of interdisciplinary convergence and complex research. To this end, the results of text analysis, frequency analysis by era, topic modeling, 2-mode network, word cloud, and ego network analysis for English thesis abstracts published by Scopus for about 55 years from 1963 to 2018 are as follows.

First, as a result of frequency analysis by era, research related to Expert Systems and Knowledge-Based Systems (KBM) was active before 2000, but interest has decreased since 2000. Interest in research related to Decision Support Systems (DSS) after 1990, Machine Learning (ML), segmentation, and Medical Images after 2000, and Prediction and Support Vector Machine (SVM) after 2010.

The analysis results of this paper show that there have been significant changes in research topics in artificial intelligence, healthcare convergence, and complex fields by era. Second, as a result of topic modeling analysis, it was possible to identify three major research topics in this field and research keywords for each research topic.
Among the three main research topics, 'AI for Clinical Decision Support Systems (CDSS) and 'AI for Medical Image' had the highest share of 17%, followed by 'The Internet of Healthcare Things (IoHT) (15%), 'Big Data Analytics in Healthcare (14%)', and 'Medical Robotics (13%), Healthcare (12BM)'.

As a result of analyzing seven major research topics by era, 'Research Trend Analysis by Text-Mining Techniques on the Convergence Studies of AI and Healthcare Technologies' in the 2000s showed that interest among researchers has increased significantly since 2010.

Third, to understand the research contents of the three major research topics, an ego network analysis was conducted focusing on the keywords with the highest connection-centeredness and mediation-centeredness. Analysis Results Research topic 1 is a study related to medical diagnosis decision support and technology acceptance based on clinical information, Research topic 2 is a study related to various medical image analyses and augmented reality, and Research topic 3 is a study related to personalized health management using sensor network technology.

The implications of this study are as follows.
First, this study is of academic significance in that it was possible to systematically grasp the global research trends in artificial intelligence, healthcare convergence, and complex fields, which are bringing great changes to the healthcare field through text analysis, topic modeling, and ego network analysis.
Second, the three main research topics derived from this study are key research topics in the healthcare field that are rapidly innovating and growing with the development of artificial intelligence technology.

According to Accenture, a global consulting firm, the application of artificial intelligence technology is expected to grow at an annual rate of 40% by 2021, and PwC, a 2016 global accounting consulting firm, surveyed 12,000 consumers in Europe, the Middle East, and Africa, and found they are ready to use healthcare services that incorporate the Fourth Industrial Revolution technology.(Choi, 2018).
It is of academic and practical significance in that it has drawn three major research topics in the complex industry, healthcare convergence closely related to artificial intelligence technology, which is growing and developing rapidly.

Finally, the results of this study can be used as basic data for setting future research directions or establishing the government's healthcare R&D policy direction and strategy by systematically examining research trends in this field. In the future, it is expected that more meaningful implications can be derived by comparing and analyzing trends in domestic and overseas paper data, patent data, and news data of research topics derived in this research.

References

Blei, D. M.,“Probabilistic Topic Models”, *Communication ofthe ACM*, Vol.55, No.4, 77-84, 2012.

Cho, Y. R., Woo , C. W. and . Choi, J. H., “Performance Analysison Collaborative Activities of Multidisciplinary Research in Government Research Institutes”, *Journal of Korea Technology Innovation Society*, Vol.20, No .4, 2017, 1089-1121.

Choi, G. Y., “European Helathcare Policy and Industry Trends in the Fourth Industrial Revolution”, *KITA Market Report,* 2018, 1-14.

Choi, H. J., “Apllication of blockchain in the healthcare industry”, *KHIDI Brief*, Vol. 236, 2017, 1-26.

Choi, H. S., and Oh, M. A., “The Need and Direction for Data-Driven Health and Welfare Policies in the 4th Industrial Revolution”, *Health and Welfare Forum*, Vol.250, 15-28, 2017.

Choi, Y. S., “Health artificial Intelligence”, *Cloud- nine*, Korea, 2018.

Nam, C. H., “Examination of possibility of utilization of topic modeling method indiary document research”*, Cross-Cultural Studies*, Vol. 22, No.1, 2016, 89-135.

Christensen, C. M., Grossman , J. H., and Hwang, J., “A Disruptive Solution for Health Care 7th, (Bae, S. Y. Trans.)”, *The Korean Doctor’s Weekly*, Korea, 2015.

Global Tech Korea,“Regulation of medical robot: Maximize opportunities while minimizing risk”,*KIAT*, 2016,1-27.

Healthcare Special Committee, Relevant Department Joint, “4th Industrial Revolution Based Heathcare Development Strategy”, *The 9th fourth Industrial Revolution Committee Resolution,* No.2, 2018.

Yoon, J. E. and Suh, C. J., “Research Trend Analysis on Smart health care by using Topic Modeling and Ego Network Analysis”, *Journal of Digital Contents Society*, Vol.19,No.5, 2018, 981-993.

Heo, J. E., "*During the Fourth Revolution, Hospital sactiveinusing Bigdata*", 2017, Available at http://biz.chosun.com/site/data/html_ dir/2017/01/08/2017010800284.html?Dep0=t witter.

Jeong, G. H., "A study of fore sight method based on text mining and complexity network analysis", *Korea Institute of Science & Technology Evaluation and Planning,* 2010, 1-138.

Jung, G. H., "Trend of artificial intelligence-based medical image analysis technology"*, ITFIND Weekly technology Trends, Vol*.1863, 2018, 1-12.

Kim, J. P. and Song, E. G., "The effects of BlockChain Technology Benefits on Acceptance Intentions of BlockChain Insurance Services : Based on the UTAUT Mode", *Journal of Information Technology Services,* Vol.17, No.4, 2018, 163-189

Kim, J. W., "Background and main contents of the establishment of mid and long-term comprehensive measures of the intelligence information society", *KIET Industry Economy*, 74-77, 2017,

Lee, D. H., Jung, H. Y., Kim, M. H., Lim, M. E., Kim, D. H., Han, Y. W., Lee, Y. W., Choi, J. H., and Kim, S. H., "Trend of Clinical Decision Suport System(CDSS)", *Electronics and Telecommunications Trends*, Vol.31. No.4, 2016, 77-85.

Rodrigues. J. C., Rezende, B. De, Segundo, H. A., Junqueira, M. H., Sabino, R. M,, Muhtadi, J. A., and Albuquerque, H.C.,"Enabling Technologies for the Internet of Health Things", *IEEE Access*, Vol.6, No.3, 13129-13141, 2018.

Simon, H., What's news for machine learning in medical imaging, *Signify Research*, 2018.

Shin, Y. S., Lee, J. H, Kim, J., Lee, J. H., Hwang, D. G., Kim, S. U., and Park, G. L., The revision of the healthcare system in accordance with the Fourth Industrial Revolution, *Korea Institute for Health and Social Affairs*, Korea, 2017.

Biographies

Kim See You is student in MY PAUL SCHOOL. He is interested in artificial intelligence, deep learning, cryptography, robots, healthcare, block chains, drones, autonomous vehicles, etc., and is conducting related research.

Shin Dong Ho is Professor and Teacher in MY PAUL SCHOOL. He obtained his Ph.D in semiconductor physics in 2000. He is interested in artificial intelligence, deep learning, cryptography, robots, block chains, drones, autonomous vehicles, the Internet of Things, metaverse, virtual reality, and space science, and is conducting related research.