

텍스트 마이닝 기법을 활용한 인공지능과 헬스케어 융·복합 분야 연구동향 분석

윤지은* · 서창진**

Research Trend Analysis by using Text-Mining Techniques on the Convergence Studies of AI and Healthcare Technologies

Jee-Eun Yoon* · Chang-Jin Suh**

■ Abstract ■

The goal of this study is to review the major research trend on the convergence studies of AI and healthcare technologies. For the study, 15,260 English articles on AI and healthcare related topics were collected from Scopus for 55 years from 1963, and text mining techniques were conducted. As a result, seven key research topics were defined : "AI for Clinical Decision Support System (CDSS)", "AI for Medical Image", "Internet of Healthcare Things (IoHT)", "Big Data Analytics in Healthcare", "Medical Robotics", "Blockchain in Healthcare", and "Evidence Based Medicine (EBM)". The result of this study can be utilized to set up and develop the appropriate healthcare R&D strategies for the researchers and government. In this study, text mining techniques such as Text Analysis, Frequency Analysis, Topic Modeling on LDA (Latent Dirichlet Allocation), Word Cloud, and Ego Network Analysis were conducted.

Keyword : Artificial Intelligence(AI), Healthcare, Text Mining, Topic Modeling, Ego Network Analysis, Word Cloud

1. 서 론

최근 전 세계적으로 다양한 학문분야에서 최첨단 혁신 기술 간의 융·복합 연구가 활발히 추진되고 있다. 이러한 융·복합 연구를 통해 개발된 새로운 혁신 기술들은 전 산업 분야에 걸쳐 파괴적 혁신(Disruptive Innovation)¹⁾을 불러오고 있고, 전혀 새로운 비즈니스 모델들이 부상하는 등 산업 전반에 걸쳐 거대한 변화의 바람이 일고 있다. 이러한 변화는 최근 화두가 되고 있는 4차 산업혁명²⁾의 핵심기반이기도 하다. 즉, 4차 산업혁명은 기존에 존재하지 않았던 새로운 차원의 기술들이 개발되면서 사회 및 과학기술에서의 난제를 극복하는 ‘융·복합기술시대’의 출현(Cho et al., 2017)이라고도 볼 수 있다.

한국은 2016년 3월 이세돌과 알파고(AlphaGo)의 대결 이후 인공지능(Artificial Intelligence; AI)에 대한 국민들의 인식이 높아졌고, 2017년 정부의 지능정보사회(Intelligent Information Society)³⁾ 종합대책 발표에 이어 5월 대통령 선거를 거치면서 4차 산업혁명에 대한 정책적 관심 또한 매우 높은 상황이다(Choi and Oh, 2017). 정부⁴⁾는 미래 헬스

케어 산업을 선도하기 위해 6대 핵심 프로젝트로 ① 헬스케어 빅데이터 쇼케이스 구축, ② 헬스케어 산업 생태계 조성, ③ 스마트 임상시험센터 구축, ④ 인공지능 활용 신약개발, ⑤ 스마트 융·복합 의료기기 개발 및 제도개선, ⑥ 체외진단기기 시장진입 촉진을 선정하여 집중 논의하고 있다(Health-care Special Committee · Relevant Department Joint, 2018). 의료계는 인공 지능 기술과 빅데이터를 활용하여 임상, 의료서비스, 병원행정 등 병원에서 발생하는 모든 정보를 통합 관리하는 차세대 병원 정보시스템(Hospital Information System, HIS) 구축 등 다양한 융·복합 프로젝트를 적극적으로 추진하고 있다(Heo, 2017). 정밀의학 분야를 선도하고 있는 서울삼성병원은 2016년 차세대 HIS인 DARWIN(Digitalized Analysis & Research Window for Integrated kNowledge)을 도입했고, 서울아산병원은 2017년 차세대 HIS인 AMIS 3.0 (Asan Medical Information system 3.0)을 가동시켰다(Shin et al., 2018). 고려대학교 의료원은 2017년 차세대 EMR(Electronic Medical Record)인 꿈(KUEM)을 구축하였고, 2018년 에이브릴 항생제 어드바이저(Aibril Antibiotics Advisor; 3A⁵⁾)를 개발했다(Lee, 2018). 이를 종합해보면 헬스케어 분야에서 인공지능과 빅데이터에 기반한 혁신적인 기술의 부각은 경제적·사회적 파급효과가 매우 클 것이라고 요약해 볼 수 있다.

이와 같이 정부, 의료계(병원), 산업계(기업)에서 인공지능과 헬스케어 융·복합 분야에 관심을 갖고 보건의료 정책 및 R&D를 추진하고 있으나 아직 도입단계에 있으며, 세부적인 동향과 보건의료 R&D 전략기획을 지원하기 위한 연구는 미흡한 실정이다. 따라서 본 논문에서는 헬스케어를 포함한 전 산업분야에서 괄목할 만한 혁신을 유도하며 빠르게 성장해 가고 있는 인공지능 기반 기술과 헬스케어 관련 기술 분야의 융·복합 연구개발 동향을 체계적으로 파악해 보고자 한다. 이를 위해

- 1) 크리스텐슨(Christensen) 교수가 제시한 이론으로 현재 산업에서 중요시하는 속성을 축소시키는 대신 기존 중요시하지 않았던 다른 속성을 제공함으로써 새로운 시장을 창출될 수 있다는 변혁의 동인(動因)을 의미함(Christensen et al., 2015).
- 2) 2016년 1월 스위스 다보스(Davos)에서 개최된 세계경제포럼(WEF, World Economic Forum)에서 클라우스 슈밥(Klaus Schwab) 교수가 4차 산업혁명을 화두로 던지면서 세계적으로 주목받음(Song, 2016).
- 3) 지능정보사회는 기업들과 국민들에게 초지능 기술과 초연결 기술의 결합인 지능정보기술에 대한 도래를 정확히 알리고 적극 대비할 수 있게 하려는 것. 지능정보 사회가 도래한 2030년에는 의료(110조 원), 제조업(92조 원), 금융(48조 원), 유통(28조 원)순으로 경제효과가 클 것으로 분석됨(Kim, 2017).
- 4) 2017년 12월에 산·학·연·관계 전문가로 구성된 4차 산업혁명위원회 산하 헬스케어 특별위원회는 ‘건강수명 연장 및 헬스케어 일자리 창출’을 목표로 매월 운영되고 있음(Healthcare Special Committee · Relevant Department Joint, 2018).

- 5) 인공지능(AI)에 기반한 항생제 추천 프로그램.

Scopus에서 수집한 논문 초록을 대상으로 텍스트 마이닝 기법과 네트워크분석 기법을 적용해서 세계적으로 이 분야 연구자들의 관심이 집중되고 있는 핵심 연구주제들을 파악하고 이를 심층적으로 분석하였다. 본 논문의 연구결과는 향후 이 분야 연구자들의 트렌디한 핵심 연구주제 선정과 정부의 보건의료분야 R&D사업 추진 전략 수립 등을 위해 활용될 수 있을 것이다.

2. 연구방법

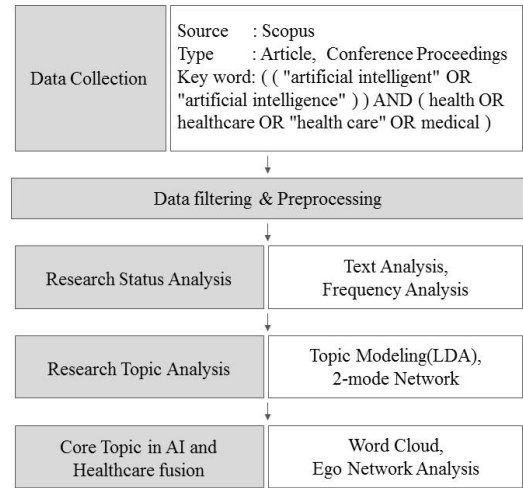
2.1 분석 대상

본 연구는 인공지능과 헬스케어 융·복합 분야의 연구동향을 파악하기 위해 Elsevier사의 Scopus에서 키워드⁶⁾ 검색결과 1963년부터 2018년까지 약 55년간 게재된 16,370편의 영어로 작성된 논문(Article, Conference Proceedings) 자료를 수집하였다. 텍스트 분석을 하기 위해 초록이 없는 논문 851편과 저자정보가 없는 196편, 중복 및 분석에 적합하지 않는 논문 63편을 제외하여 총 15,260편을 최종 분석 대상으로 선정하였다.

2.2 분석 방법

본 연구는 <Figure 1>과 같이 Scopus에서 수집한 ‘인공지능과 헬스케어 융·복합’ 분야의 영어 논문 초록을 대상으로 텍스트 분석, 시대별 주요 단어(핵심어) 빈도분석, LDA 기반 토픽모델링, 워드클라우드, 에고네트워크 분석을 시행하였다.

수집된 논문 초록을 텍스트 분석(text analysis) 하기 위해 소문자, 띄어쓰기 등을 표준화하고, 특수 문자 등 불필요한 정보는 전처리 과정을 통해 삭제하였다. 그리고 <Table 1>과 같이 약어, 단·복수형은 유의어 사전에 추가하고 15,260편의 논문 키워드를 기반으로 생성한 사용자 정의 사전(유의어



<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.

(Thesaurus), 복합명사(Defined Word))와 불용어(Stop Word) 사전을 적용하였다. 텍스트 분석에는 Netminer 4.4.2에서 제공하는 OpenNLP 형태소 분석기를 사용하였다.

LDA기반 토픽모델링은 비정형데이터인 텍스트 자료를 분석해서 의미 있는 주요 토픽(이하 ‘연구주제’)들을 찾아내기 위한 생성·확률모델(generative probabilistic model) 활용기법이다(Blei, 2012). 토픽모델링은 방대하고 구조화되지 않은 텍스트 자료로부터 맥락과 관련된 단어들의 패턴분석을 통해 의미 있는 연구주제들을 추론할 수 있다. 이러한 특성 때문에 토픽모델링은 논문 초록의 텍스트를 분석하여 주요 연구주제를 추출하고 시간의 흐름에 따른 연구동향을 파악하는 연구방법론으로 널리 활용되고 있다(Nam, 2016). 토픽모델링에서 연구자가 사전에 지정해 줄 수 있는 값은 연구주제의 개수,

6) ("artificial intelligent" OR "artificial intelligence") AND (health OR healthcare OR "health care" OR medical).

α 값, β 값, Gibbs sampling 값이 있다. Steyvers and Griffiths은 α 값을 $50/k$ (연구주제 수)로 값을 고정하고, 0.1, 0.01과 같이 β 값을 다르게 설정해서 최적의 값을 찾는 것을 권장했다(Steyvers and Griffiths, 2007). 본 연구에서는 LDA 기반 토픽모델링을 수행하기 위해 연구주제 수 7개,⁷⁾ α 값 7.0(50/7), β 값 0.1로 설정하고 Gibbs sampling⁸⁾을 1000번 반복했다.

네트워크 분석(Network Analysis)은 복잡한 관계를 보다 직관적으로 파악할 수 있도록 시각화하는 방법으로 특정 유형의 노드(node)와 노드들 간 연결 관계(link)로 정의된다. 네트워크 분석은 네트워크 전체를 분석할 수도 있지만, 네트워크 내 개별 노드를 에고로 설정해서 분석할 수 있다. 따라서 분석목적에 따라 특정 노드를 에고(ego)로 설정하여 직접 연결 관계를 형성하는 이웃 노드들을 분석하는 것을 에고 네트워크 분석이라고 한다. 에고 네트워크는 에고노드와 연결된 네트워크의 부분적 특성을 관찰하는데 유용하게 사용될 수 있다(Lee, 2012).

네트워크의 구조적 특성을 파악하기 위한 중요한 지표로는 연결중심성(Degree Centrality), 매개중심성(Betweenness Centrality) 등이 있다. 연결중심성은 네트워크 내 노드들이 얼마나 많은 연결을 가지고 있는지 측정하는 지표(Lee, 2012)로 연결중심성이 높은 노드일수록 이와 관련된 연구가 활발히 일어나고 있다고 볼 수 있다. 매개중심성은 한 노드가 네트워크 내의 다른 노드들 사이에서 중개자의 역할을 얼마나 수행하는지 측정하는 지표이다. 매개중심성이 높은 노드일수록 연구의 흐름을 통제하는 영향력을 가지며 중개자, 문지기 역할을 한다고 볼 수 있다(Lee, 2012). 네트워크 분석

방법은 미래를 예측하거나 트렌드를 반영하기 어려운 기존의 정성적인 분석을 보완하는 방법 중 하나로 제시되고 있다(Yoon and Suh, 2018; Jeong, 2010). 본 연구에서는 ‘인공지능과 헬스케어 융·복합’ 분야의 연구동향을 보다 효과적으로 분석하기 위해 주요 연구주제를 구성하고 있는 할당확률이 높은 단어 상위 100개를 대상으로 워드클라우드를 분석하고, 각 연구주제별 연결중심성과 매개중심성이 가장 높은 핵심어를 대상으로 에고 네트워크를 분석하여 동시출현(Co-occurrence) 단어⁹⁾들을 살펴보았다.

3. 연구결과

3.1 시대별 연구동향 분석결과

본 연구에서는 1963년부터 2018년까지 약 55년간 논문에 나타난 핵심어를 파악하기 위해 텍스트 분석, 빈도분석을 실시했다. 수집한 논문 총 15,260편에서 추출된 단어 11,008개 중 TF-IDF 0.3 이상, 논문 8편 이상 등장한 단어 3,949개를 대상으로 고빈도 상위 10개를 <Table 2>와 같이 시기별로 추출하였다.

전 기간 가장 고빈도로 등장한 연구 핵심어는 Medical Prescription, Artificial Intelligence(AI), Medical Doctor, Medicine, Risk Factor로 나타났다. 1963년부터 1999년까지 부각된 연구 핵심어는 Expert System과 knowledge Based System(KBS)으로 나타났으나 2000년 이후 빈도수가 급격히 떨어져 관심도가 낮아졌음을 알 수 있다. 1990년 이후에는 Decision Support System(DSS)이 새로운 연구 핵심어로 급부상했으며 2010년 이후에도 빈도수가 상승하여 관심도가 증가하고 있음을 알 수 있다. 2000년 이후에는 Machine Learning(ML), segmentation, Medical Image가 주요 연구 핵심어로 새롭게 등장했으며, 2010년 이후에는 Prediction, Support Vector Machine(SVM) 같은

7) 연구주제 수는 다양한 수의 토픽모델링 분석 결과 중 연구주제의 해석가능성이 가장 높은 7개로 설정함.

8) 복잡한 값을 샘플링하기 위해 설계된 Markov chain Monte Carlo(MCMC) 알고리즘 중 하나로 단어에 할당확률을 부여하여 임의의 연구주제에 할당함. Gibbs sampling을 100-1000번을 시행한 결과 연구주제별 핵심어가 가장 잘 할당된 1000번으로 설정함.

9) 초록에서 함께 등장한 단어를 의미함.

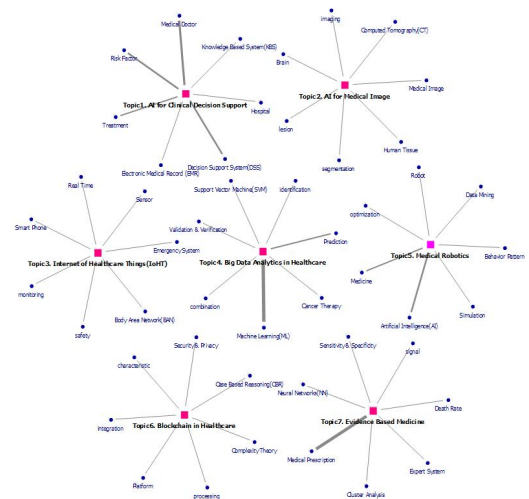
〈Table 2〉 Result of Frequency Analysis

	1963~1989	1990's	2000's	2010's
Doc.	753	1,235	4,001	9,271
Word	1,726	2,471	3,717	3,936
1	Expert System (297)	Knowledge Based System (KBS) (306)	Medical Prescription (583)	Machine Learning (ML) (3,013)
2	Medical Prescription (195)	Medical Prescription (244)	Artificial Intelligence (AI) (561)	Artificial Intelligence (AI) (1,737)
3	Knowledge Based System (KBS) (191)	Medical Doctor (239)	Medical Doctor (508)	Medical Doctor (1,411)
4	Artificial Intelligence (AI) (168)	Artificial Intelligence (AI) (232)	Machine Learning (ML) (469)	Medical Prescription (1,705)
5	Medical Doctor (149)	Expert System (201)	Risk Factor (451)	Decision Support System (DSS) (1,327)
6	Medicine (101)	Medicine (177)	Decision Support System (DSS) (357)	Risk Factor (1,231)
7	Risk Factor (83)	Risk Factor (151)	Medicine (317)	Prediction (1,135)
8	Treatment (55)	Decision Support System (DSS) (110)	segmentation (311)	Treatment (960)
9	Decision Making (55)	Electronic Medical Record (EMR) (100)	Neural Network (NN) (301)	Medicine (779)
10	consultation (52)	Neural Network (NN) (97)	Medical Image (286)	Support Vector Machine (SVM) (758)

핵심어들이 급상승했다. 이러한 텍스트 빈도분석을 통해 시대별로 주요 연구주제에 상당한 변화가 있었음을 파악할 수 있다.

3.2 주요 연구주제 분석 결과

본 연구에서는 논문 8편 이상 등장한 단어 3,949개를 대상으로 LDA기반 토픽모델링을 수행하였다. 분석된 주요 연구주제의 독립성을 판단하기 위해 연구주제 간 유사도를 검사한 결과 0.05 이하로 나타나 연구주제 간 유사도 문제는 없는 것으로 보인다. 또한 연구주제와 연구주제별 할당 확률이 높은 상위 7개 단어를 대상으로 <Figure 2>와 같이 2-mode 네트워크를 분석한 결과 중복되는 단어가 없는 것으로 나타나 주요 연구주제는 독립적으로 분석되었다고 판단되어진다.



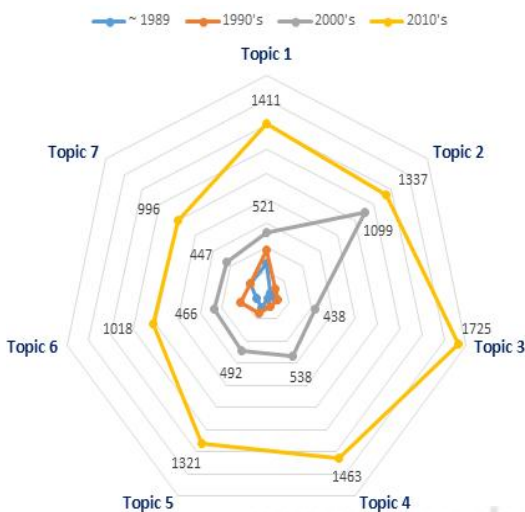
〈Figure 2〉 Result of 2-Mode Network(Topic-Word)

주요 연구주제 분석결과는 <Table 3>과 같이 연구주제별로 할당확률이 높게 나타난 주요 핵심어 7개를 순서대로 정리하고, 연구주제가 가장 대표적으로 나타난 문서 비중을 제시하였다. 주요 연구주제명은 연구주제별 할당확률이 높은 주요 핵심어들 간 연계성을 고려하여 선정하였다. 논문 15,260편 중 연구주제 점유율은 Artificial Intelligence(AI) for Clinical Decision Support System (CDSS)(17%)와 Artificial Intelligence(AI) for Medical Image(17%)가 가장 높은 것으로 나타났다.

〈Table 3〉 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%
Topic 4	Big Data Analytics in Healthcare	Machine Learning(ML), Prediction, Support Vector Machine(SVM), Cancer Therapy, Validation & Verification, combination, identification	2107	14%
Topic 5	Medical Robotics	Artificial Intelligence(AI), Medicine, Behavior Pattern, Robot, Data Mining, Simulation, optimization	2060	13%
Topic 6	Blockchain in Healthcare	processing, Case Based Reasoning(CBR), Security & Privacy, Platform, integration, Complexity Theory, characteristic	1801	12%
Topic 7	Evidence Based Medicine(EBM)	Medical Prescription, Neural Network(NN), Expert System, signal, Cluster Analysis, Death Rate, Sensitivity & Specificity	1785	12%

〈Figure 3〉에서는 시기별 주요 연구주제의 변화를 쉽게 비교해 볼 수 있다. 연구주제 2는 2000년대 논문 수의 비중이 크게 증가한 이후 지속적으로 관심도가 있는 것을 알 수 있으며, 나머지 연구주제들은 2010년대 논문 수의 비중이 55%가 넘는 것으로 나타나 연구주제 관심도가 2010년 이후 급격히 증가한 것을 알 수 있다.



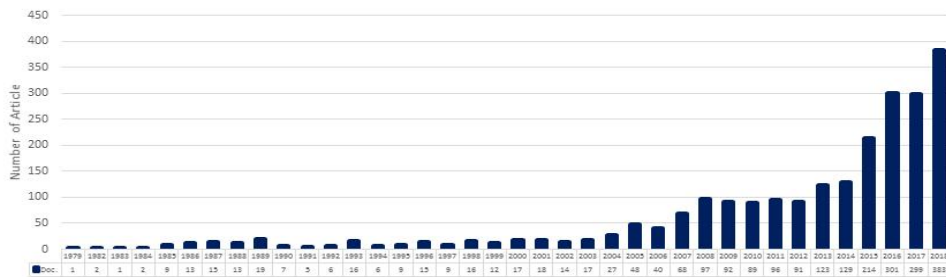
〈Figure 3〉 Changes in Number of Topic

3.3 주요 연구 주제별 분석 결과

3.3.1 Topic 1. Artificial Intelligence(AI) for Clinical Decision Support System (CDSS)

전체기간 중 헬스케어와 인공지능 융·복합 분야의 연구주제 점유율이 가장 높게 나타난 Artificial Intelligence(AI) for Clinical Decision Support System(CDSS)(이하 ‘연구주제 1’)은 전체 논문 15,260편 중 2,586편(17%)으로 나타났다. 연구주제 1은 환자의 임상정보를 기반으로 의료인이 질병을 진단하고 치료할 때 의사결정을 돕는 전문가 시스템을 말한다. 즉, 연구주제 1은 의사가 정밀한 진단을 할 수 있도록 임상 의사결정을 보조하고, 약물 부작용이 발생하지 않도록 치료와 처방단계에서 환자의 약물 복용정보를 제공하여 의학적 판단과 의료의 질 향상에 도움을 준다(Lee et al., 2016).

연구주제 1의 연도별 논문 수는 〈Figure 4〉에서 보는 바와 같은 증가추세를 보이고 있다. 1975년~2018년 기간 동안의 연평균증가율(Compound Annual Growth Rate, CAGR)은 14%인 것으로 나타났다.



〈Figure 10〉 Changes of the Number of Article(Topic 3)

되어 있어 다양한 의료영상이 분석에 적용되고 있으며 증강현실과 관련 연구가 진행되고 있음을 알 수 있다. 글로벌 의료기술 산업 컨설팅 업체인 Signify Research에 의하면 연구주제 2의 시장규모는 2023년까지 20억 달러에 이를 것으로 예측하고 있다(Simon, 2018). 연구주제 2와 관련된 주요 글로벌 기업으로는 IBM Watson Health, Aidence, Aidoc, Brainomix, DEEP-NOID Int., Enlitic, EnvoyAI, Optellum 등이 있다.

3.3.3 Topic 3. Internet of Healthcare Things (IoHT)

Internet of Healthcare Things(IoHT)(이하 ‘연구주제 3’)의 연구주제 점유율은 전체의 15%(2,339편)로 나타났다. 사물인터넷(Internet of Things, IoT)은 4차 산업혁명의 주요 기술 중 하나로 다양한 플랫폼과 상호 연결된 기술을 기반으로 한 사물(서비스, 장소, 제품)과 인간의 관계로 설명할 수 있으며(Song, 2016), 하버드 비즈니스 리뷰(Harvard Business Review)에 의하면 280억 가지의 “사물(things)”들이 인터넷과 연결될 것으로 예상하고 있다. IoT 기반의 솔루션인 연구주제 3은 의료계, 산업계에서 빠르게 받아 들여지고 있으며, 원격의료 모니터링(Remote Healthcare Monitoring), 스마트 폰 기반의 의료 솔루션(Healthcare Solutions Using Smartphones), Ambient Assisted Living(AAL),¹³⁾ 웨어러블 기기(Wearable Devices) 등을 지원할 수 있다(Rodrigues et al., 2018). 연구주제 3의 논문 수는 <Figure 10>에서



〈Figure 11〉 Result of Word Cloud(Topic 3)

보는 바와 같이 꾸준히 증가추세를 보이고 있으며, 1979년~2018년 동안의 CAGR은 14%로 나타났다.

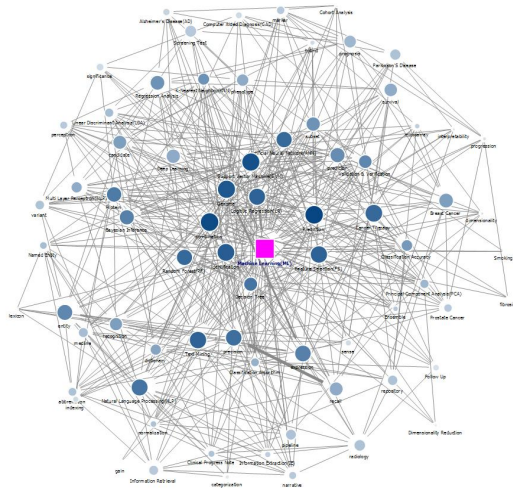
연구주제 3과 관련된 주요 키워드를 파악하기 위해 <Figure 11>과 같이 할당확률 상위 100개 단어를 대상으로 워드클라우드를 분석한 결과 Sensor, monitoring, Real Time, Body Area Network (BAN), safety가 주요도가 높게 나타났다. 연구주제 3을 보다 심도 있게 분석하기 위해 중앙중심성(0.6162)과 매개중심성(0.1255)이 높은 Sensor를 대상으로 <Figure 12>와 같이 에고네트워크¹⁴⁾를 분석했다. 분석 결과 Sensor는 할당확률이 높은 상위 4개 단어와 완전연결 그래프로 나타났으나 safety와는 연결되지 않은 것으로 나타났으며, Smart Phone, Accelerometer, Internet of Things(IoT), Failure, Human Activity Recognition(HAR)과 긴밀하게 연결되어 있는 것으로 나타나 센서 네트워크 기술을 활용하여 실시간 모니터링을 통해 개인맞춤형 건강관리 연구가 진행되고 있음을 알 수 있다.

13) 014년 능동형 생활지원(Active Assisted Living; AAL)으로 명칭 변경(Suh and Kang, 2018).

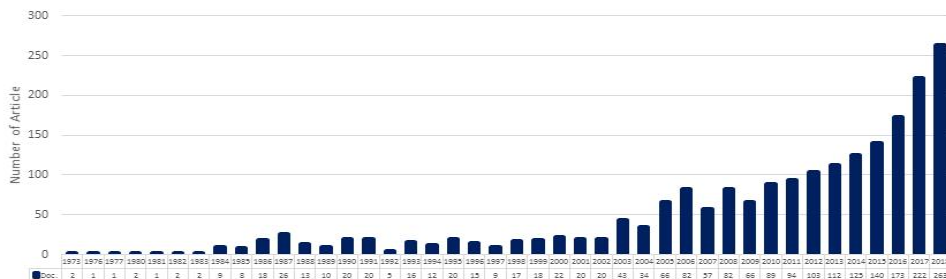
14) link reduction simulation 결과 컴포넌트가 분할되어, weight가 약하지만 중요한 노드와의 연결 끊김을 방지하기 위해 weight를 한정하지 않음.

것으로 나타났으며, Genome, Logistic Regression (LR), Feature Selection(FS), Decision Tree, Natural Language Processing(NLP)과 인접해 있는 것으로 나타나 연구주제 4는 유전자 및 의료 데이터, 텍스트 분석, 질병 예측분석 등과 관련된 연구가 진행되고 있음을 알 수 있다.

P&S Market Research(2018)에 의하면, 연구주제 4의 시장규모는 2018년~2023년 동안 연평균 증가율(CAGR)이 21.8%에 달할 것으로 예측하고 있다. 연구주제 3은 개인이 건강을 관리하고 의료를 선택하는 방식을 변화시켜 건강에 대한 결과와 의료비용 지출의 효율을 높이는 등 보건의료시스템 전반에 걸쳐 혁신 성장동력이 될 것으로 기대되고 있다(Kang, 2016).



<Figure 15> Ego Network Analysis(Topic 4)



<Figure 16> Changes of the Number of Article(Topic 5)

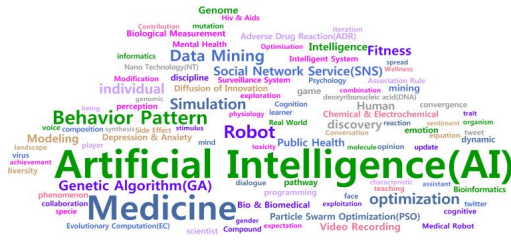
연구주제 4와 관련된 글로벌 회사로는 IBM Watson Health, Cerner, HealthCatalyst, HealthECTM, EPIC, AMITECH, Acmeaware, CONICER, Prognos, Optum, GE Healthcare, Siemens AG 등이 있다.

3.3.5 Topic 5. Medical Robotics

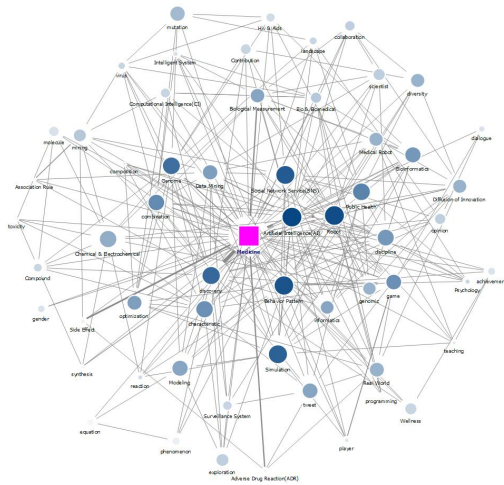
Medical Robotics(이하 ‘연구주제 5’)의 연구주제 점유율은 전체의 13%(2060편)로 나타났다. 최근 로봇과 인공지능이 급속히 발전하면서 기존에 사람들이 하던 일들을 로봇이 수행하거나 사람들이 수행하기 어렵거나 효과적으로 할 수 없던 일들을 로봇이 보조하기 시작했다. 연구주제 5는 로봇 응용 분야에서 가장 활발하게 연구가 이뤄지고 있으며, 다빈치(DaVinci) 수술 시스템과 같은 “수술 로봇”, 환자에게 식사와 약을 가져다주는 “일상작업 로봇15)”, 병원 및 일상생활 환경에서 사용되는 “개인관리(Personal care) 로봇” 등 다양하게 활용될 수 있다(GlobalTech Korea, 2016). 연구주제 5의 논문 수는 <Figure 16>과 같이 지속적인 증가추세이며, 1973년~2018년 CAGR은 11%인 것으로 나타났다.

연구주제 5와 관련된 주요 연구 키워드를 파악하기 위해 할당확률 상위 100개 단어를 대상으로 <Figure 17>과 같이 워드클라우드 분석한 결과 Artificial Intelligence(AI), Medicine, Behavior Pattern, Robot, Data Mining이 주요하게 나타났다.

15) 향후 의료 혹은 간병 업무에 종사할 수 있는 잠재력을 지니고 있음.



〈Figure 17〉 Result of Word Cloud(Topic 5)



〈Figure 18〉 Ego Network Analysis(Topic 5)

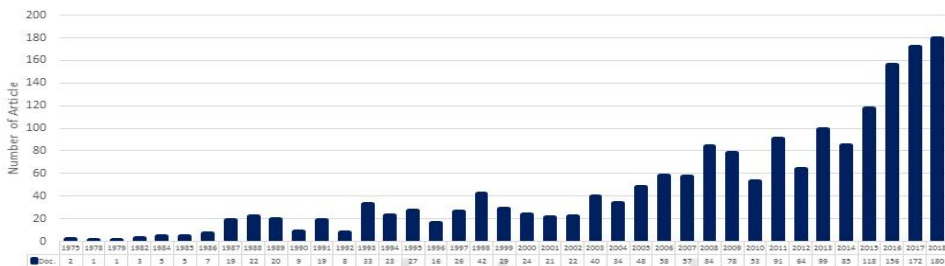
연구주제 5의 텍스트 네트워크 구조를 효과적으로 분석하기 위해 중앙중심성(0.5859)과 매개중심성(0.1808)이 가장 높은 Medicine을 중심으로 <Figure 18>과 같이 에고 네트워크를 분석했다. 분석 결과 Medicine은 할당확률 상위 5개와 모두 연결되어 있으며, discovery, Genome, Biological Measurement와 긴밀하게 연결된 것으로 나타나 다양한 지능형 로봇 연구들이 진행되고 있음을 알 수 있다. 또

한 매개중심성과 중앙중심성은 낮은 Side Effect, Adverse Drug Reaction(ADR)과 긴밀하게 연결된 것으로 나타나 연구주제 5의 부작용에 대한 연구가 진행되고 있음을 알 수 있다.

MarketAndMarket의 시장 보고서(2018)에 의하면, 연구주제 5의 시장규모는 2017년 54.7억 달러에서 2023년 167.4억 달러로 2018년~2023년 동안 CAGR이 21%에 달할 것으로 예측하고 있다. 인구의 고령화와 개인 맞춤형 의료 수요의 증가와 전문 의료 인력 부족 등으로 인해 연구주제 5에 대한 수요가 증가하고 있다(GlobalTech Korea, 2016). 연구주제 5와 관련 회사로는 Intuitive Surgical, Hansen Medical, Medrobotics, Verb Surgical, Microbot Medical, Titan Medical, CyberKnife System이 있다.

3.3.6 Topic 6. Blockchain in Healthcare

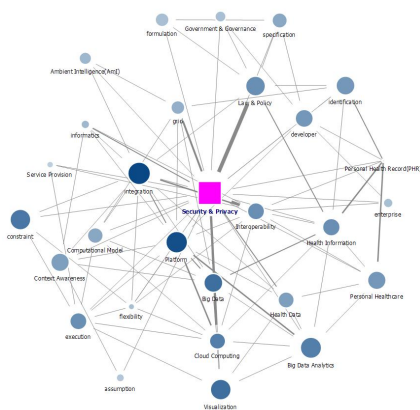
Blockchain in Healthcare(이하 '연구주제 6')의 연구주제 점유율은 전체의 12%(1801편)로 나타났다. 블록체인(Blockchain)은 중앙에서 통제되는 기존의 시스템과 달리 분산화 된 공개 거래장부로 특정한 주체에 의해 통제가 불가능하고 위조가 불가능한 구조이다(Kim and Song, 2018). 디지털화된 데이터의 안정성과 투명성을 높이기 위한 방안으로 의료분야에서 블록체인을 도입하고 있다. 연구주제 6은 개인 주도형 건강관리 지원, 보험 청구·심사 프로세스 적용, 의료기기·의약품 유통 채널 적용, 임상시험 안전성 향상과 데이터 공유, 개인 의료정보 및 건강정보의 보호 등 다양하게 활용될 수 있다(Choi, 2017). 연구주제 6의 논문 수는 <Figure 19>에서 보는 바와 같이 연도별 증가



〈Figure 19〉 Changes of the Number of Article(Topic 6)



<Figure 20> Result of Word Cloud(Topic 6)



<Figure 21> Ego Network Analysis(Topic 6)

추세를 보이고 있으며, 1975년~2018년 동안의 CAGR은 11%인 것으로 나타났다.

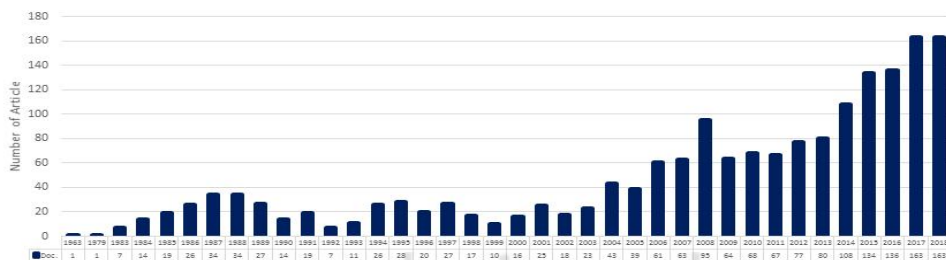
연구주제 6과 관련된 주요 연구 키워드를 파악하기 위해 할당확률 상위 100개 단어를 대상으로 <Figure 20>과 같이 워드클라우드를 분석한 결과 processing, Case Based Reasoning(CBR), Security & Privacy, Platform, integration이 주요하게 나타났다. 연구주제 6의 텍스트 네트워크 구조를 보다 심도 있게 분석하기 위해 중앙중심성(0.2828)

과 매개중심성(0.0526)이 가장 높은 Security & Privacy를 대상으로 <Figure 21>과 같이 에고 네트워크를 분석했다. 분석 결과 Security & Privacy는 Law & Policy, grid, Cloud Computing, Interoperability, Health Data가 긴밀하게 연결되어 있는 것으로 나타나 연구주제 6의 보안과 법률 및 정책 제도와 관련된 연구가 주를 이루고 있음을 알 수 있다.

MarketAndMarket의 시장보고서(2018)에 의하면 연구주제 6 시장 규모는 2017년 36.9백만 달러에서 2023년에는 8.29억 달러로 2018년~2023년 동안 CAGR이 72.8%로 급격히 성장할 것으로 예측하고 있다. 연구주제 6 관련 글로벌 기업으로는 Google Deepmind Health, IBM Watson Health-FDA, Gem Health, MedRec, BlockMedX, Robomed Network, Microsoft Blockpharma 등이 있다.

3.3.7 Topic 7. Evidence Based Medicine(EBM)

Evidence Based Medicine(EBM)(이하 ‘연구주제 7’)의 연구주제 점유율은 전체의 12%(1,785편)으로 나타났다. 연구주제 7은 임상 전문 지식을 체계적인 연구를 통해 얻은 가장 유용한 외부 임상 증거와 통합하는 것을 말하는 것으로, 개별 환자에게 진료 혹은 의료 서비스를 제공할 때 현재 나온 연구 중에서 가장 타당한 근거를 기반으로 의사결정 하는 것을 의미한다(Sackett et al., 1999). 연구주제 7의 논문 수는 <Figure 22>와 같은 연도별 증가 추세를 보이고 있다. 1975년~2018년 동안의 CAGR은 10%인 것으로 나타났다.



<Figure 22> Changes of the Number of Article(Topic 7)

이후 관심도가 급격히 증가했으며 나머지 연구주제들은 2010년 이후 연구자들의 관심도가 크게 증가한 것으로 나타났다.

셋째, 7개의 주요 연구주제들의 연구내용을 파악하기 위해 연결중심성과 매개중심성이 가장 높은 주요 핵심어 중심으로 에고 네트워크 분석을 실시하였다. 분석결과 ①연구주제 1은 임상정보를 기반으로 의료진단 의사결정 지원과 기술 수용 관련 연구, ②연구주제 2는 다양한 의료영상 분석과 증강현실 관련 연구, ③연구주제 3은 센서 네트워크 기술을 활용한 개인맞춤형 건강관리 관련 연구, ④연구주제 4는 유전자 및 의료데이터 분석 관련 연구, ⑤연구주제 5는 다양한 지능형 로봇과 부작용 관련 연구, ⑥연구주제 6은 보안과 법률 및 정책제도 관련 연구 ⑦연구주제 7은 기계학습 연구와 증의학, 간염, 치매 관련 연구가 진행되고 있음을 알 수 있었다.

본 연구가 주는 시사점은 다음과 같다.

첫째, 본 연구는 텍스트 분석과 토픽모델링, 에고네트워크 분석을 통해 4차 산업혁명이라는 화두 속에서 헬스케어 분야에 커다란 변혁을 가져오고 있는 인공지능과 헬스케어 융·복합 분야의 전세계 연구동향을 시대별로 체계적으로 파악해 볼 수 있었다는 점에서 학술적 의의를 가진다.

둘째, 본 연구에서 도출된 7개의 주요 연구주제는 인공지능 기술의 발달로 빠르게 혁신성장하고 있는 헬스케어 분야의 핵심 연구주제들이다. 글로벌 컨설팅 기업인 Accenture에 의하면 헬스케어 분야에서 인공지능 기술의 적용은 2021년까지 연평균 40%라는 엄청난 성장률을 보여줄 것으로 전망하고 있으며, 2016년 글로벌 회계 컨설팅 기업인 PwC가 EMEA(Europe, the Middle East and Africa)지역의 12개국 소비자 12,000명을 대상으로 설문조사 한 결과 4차 산업혁명 기술을 접목한 헬스케어 서비스를 이용할 준비가 되어 있고 인공지능 및 로봇진료도 기꺼이 받을 준비가 되어 있는 것으로 나타났다(Choi, 2018). 이와 같이 빠르게 성장 발전해 가고 있는 인공지능 기술과 밀접하게 연관된 헬스케어 융·복합 산업의 7개의 주요

연구주제들을 일목요연하게 살펴 볼 수 있게 도출했다는 점에서 학술적·실무적 의의를 가진다.

마지막으로 본 연구 결과는 향후 이 분야 연구자들이 그간의 연구동향을 체계적으로 살펴봄으로써 향후 연구방향을 설정하거나 정부의 보건의료 R&D 정책 방향과 전략수립에 중요한 기초자료로 활용될 수 있을 것이다. 향후에는 본 연구에서 도출된 연구주제의 국내·외 논문데이터, 특허 데이터, 뉴스 데이터에 대한 동향을 함께 비교 분석해 본다면 보다 유의미한 시사점을 도출 할 수 있을 것으로 기대된다.

References

- Blei, D.M., "Probabilistic Topic Models", *Communication of the ACM*, Vol.55, No.4, 2012, 77-84.
- Cho, Y.R., C.W Woo., and J.H. Choi, "Performance Analysis on Collaborative Activities of Multidisciplinary Research in Government Research Institutes", *Journal of Korea Technology Innovation Society*, Vol.20, No.4, 2017, 1089-1121.
- (조용래, 우청원, 최종화, "국가 출연연구소의 협업적 융합연구 성과 분석", *기술혁신학회지*, 제20권, 제4호, 2017, 1089-1121.)
- Choi, G.Y., "European Healthcare Policy and Industry Trends in the Fourth Industrial Revolution", *KITA Market Report*, 2018, 1-14.
- (최경윤, "4차 산업혁명 시대의 유럽 헬스케어 정책 및 산업 동향", *한국무역협회 해외시장 보고서*, 2018, 1-14.)
- Choi, H.J., "Application of blockchain in the healthcare industry", *KHIDI Brief* Vol.236, 2017, 1-26.
- (최한준, "헬스케어 산업에서의 블록체인 기술의 활용", *보건산업브리프*, 제236권, 2017, 1-26.)

- Choi, H.S. and M.A. Oh, "The Need and Direction for Data-Driven Health and Welfare Policies in the 4th Industrial Revolution", *Health and Welfare Forum*, Vol.250, 2017, 15-28.
- (최현수, 오미애, "4차 산업혁명에 대비한 보건복지 분야 데이터 주도 정책 추진 필요성과 방향", *보건복지포럼*, 제250호, 2017, 15-28.)
- Choi, Y.S., "Health artificial Intelligence", Cloud-nine, Korea, 2018.
- (최윤섭, "의료 인공지능", 클라우드 나인, 한국, 2018.)
- Christensen, C.M., J.H. Grossman, and J. Hwang, "A Disruptive Solution for Health Care 7th (Bae, S.Y. Trans.)", *The Korean Doctor's Weekly*, Korea, 2015.
- (클레이튼 M. 크리스텐슨, 제롬 H. 그로스만, 제이슨 황, "파괴적 의료혁신 7판(배성운 역)", 청년의사, 2015.)
- GlobalTech Korea, "Regulation of medical robot : Maximize opportunities while minimizing risk", *KIAT*, 2016, 1-27.
- (GlobalTech Korea, "의료 로봇의 규제 : 위험 요소를 최소화하면서 기회를 극대화하기", 한국산업 기술진흥원, 2016, 1-27.)
- Healthcare Special Committee · Relevant Department Joint, "4th Industrial Revolution Based Healthcare Development Strategy", *The 9th fourth Industrial Revolution Committee Resolution*, No.2, 2018.
- (헬스케어 특별위원회 · 관계부처 합동, "4차 산업혁명 기반 헬스케어 발전 전략", 제9차 4차산업혁명위원회 의결안건, 제2호, 2018.)
- Heo, J.E., "During the Fourth Revolution, Hospitals active in using Bigdata", 2017, Available at http://biz.chosun.com/site/data/html_dir/2017/01/08/2017010800284.html?Dep0=twitter.
- Jeong, G.H., "A study of foresight method based on textmining and complexity network analysis", *Korea Institute of Science & Technology Evaluation and Planning*, 2010, 1-138.
- (정근하, "텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구", *한국과학기술기획평가원*, 2010, 1-138.)
- Jung, G.H., "Trend of artificial intelligence-based medical image analysis technology", *ITFIND Weekly Technology Trends*, Vol.1863, 2018, 1-12.
- (정규환, "인공지능 기반 의료영상 분석 기술 동향", *정보통신기획평가원 주간기술동향*, Vol.1863, 2018, 1-12.)
- Kang, H.J., "National-Level Use of Health Care Big Data and Its Policy Implications", *Health and Welfare Forum*, Vol.238, 2016, 55-71.
- (강희정, "보건의료 빅데이터의 정책현황과 과제", *보건복지포럼*, 제238호, 2016, 55-71.)
- Kim, J.P. and E.G. Song, "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.
- (김종필, 송유진, "블록체인 기술 혜택의 효과가 블록체인 보험 서비스의 수용의도에 미치는 영향 : UTAUT 모형을 기반으로", *한국IT서비스학회지*, 제17권, 제4호, 2018, 163-189.)
- Kim, J.W., "Background and main contents of establishment of mid and long term comprehensive measures of intelligence information society", *KIET Industry Economy*, 2017, 74-77.
- (김정원, "지능정보사회 중장기 종합대책의 수립 배경과 주요 내용", *KIET 산업경제*, 2017, 74-77.)
- Kim, Y.H. and Y.J. Kim, *Social Network Analysis 4th*, Parkyoungsa, Korea, 2016.
- (김용학, 김영진, *사회연결망분석* 제4판, 박영사,

- 한국, 2016.)
- Lee, D.H., H.Y. Jung, M.H. Kim, M.E. Lim, D.H. Kim, Y.W. Han, Y.W. Lee, J.H. Choi, and S.H. Kim, "Trend of Clinical Decision Support System(CDSS)", *Electronics and Telecommunications Trends*, Vol.31. No.4, 2016, 77-85.
- (이동훈, 종호열, 김민호, 임명은, 김대희, 한영웅, 김영원, 최재훈, 김승환, "임상의사결정지원시스템(CDSS) 기술동향", *전자통신동향분석*, 제31권, 제4호, 2016, 77-85.)
- Lee, S.S., *Network Analysis Methodology*, non-hyeong, Korea, 2012.
- (이수상, 네트워크 분석 방법론, 논형, 한국, 2012.)
- Lee, M.J., "Aibril Health status and plan", 2018 Korea Society of IT Services spring conference, Korean Federation of Science and Technology Societies, 2018, 32-43.
- (이명진, "Aibril Health 현황 및 계획", 2018 한국IT서비스학회 춘계학술대회, 한국과학기술회관, 2018, 32-43.)
- MarketAndMarket, "Medical Robots Market by Product(Instruments & Accessories, Medical Robotic Systems(Surgical Robots, Rehabilitation Robots, Non-invasive Radiosurgery Robots)), Application(Laparoscopy, Pharmaceutical Applications), and Region-Global Forecasts to 2023", Jul. 2018.
- Nam, C.H., "Examination of possibility of utilization of topic modeling method in diary document research", *Cross-Cultural Studies*, Vol. 22, No.1, 2016, 89-135.
- (남춘호, "일기자료 연구에서 토픽모델링 기법의 활용가능성 검토", *비교문화연구*, 제22권, 제1호, 2016, 89-135.)
- P&S Market Research, "Global Clinical Decision Support Systems(CDSS) Market Size, Share, Development, Growth and Demand Forecast to 2020-Industry Insights by Product (Integrated and Standalone), by Application (Drug Dosing Support, Clinical Guidelines, Clinical Reminders, Drug-Drug Interactions, Drug Allergy Alerts, and Others), By Model (Knowledge-Based, and Non-Knowledge-Based), By Mode of Delivery(Web-Bsed, Cloud-Based and On-Premises), By Component(Software, Hardware, and Services), and By Healthcare Provider Entity Capacity", Aug. 2016.
- P&S Market Research, "Big Data Analytics in Healthcare by Component(Software[Electronic Health Record Software, Practice Management, Workforce Manangement], Hardware [Data Storage, Routers, Firewalls, Virtual Private Networks, E-Mail Service], Services), by Deployment Type(On-Demand, On-Premise), by Analytics Type(Descriptive, Predictive, Prescriptive), by Application(Financial Analytics[Claim Processing, Revenue Cycle Management, Risk Adjustment & Assessment, Payment Integrity and Fraud, Waste and Abuse], Clinical Data Analytics [Quality Control, Population Health Management, Clinical Decision Support, Reporting and Compliance, Precision Health], Operational Analytics, Population Health Analytics), by Region (U.S., Canada, U.K., Germany, France, Italy, China, Japan, India, Brazil, U.A.E.) - Global Market Size, Share, Development, Growth, and Demand Forecast, 2013~2023", Jun.2018.
- Rodrigues. J.J.P.C., D.B. De Rezende Segundo, H.A. Junqueira, M.H. Sabino, R.M. Prince, J. Al-Muhtadi, and V.H.C. De Albuquerque, "Enabling Technologies for the Internet of Health Things", *IEEE Access*, Vol.6, No.3,

- 2018, 13129-13141, doi : 10.1109/ACCESS.2017.2789329 (Downloaded Nov 6, 2018.)
- Sackett, D.L., W.M.C. Rosenber, J.A.M. Gray, R.B. Haynes, and W.S. Richardson, "Evidence based medicine : what it is and what is isn't", *BMJ*, 1996, Vol.312, 71-22.
- Simon, H., What's news for machine learning in medical imaging, Signify Research, 2018.
- Shin, Y.S., J.H. Lee, B.J. Kim, J.H. Lee, D.G. Hwang, S.U. Kim, and G.L. Park, The revision of the healthcare system in accordance with the Fourth Industrial Revolution, Korea Institute for Health and Social Affairs, Korea, 2017.
- (신영석, 이진형, 김범준, 이재훈, 이영희, 황도경, 김소운, 박금령, "제4차 산업혁명에 조응하는 보건의료 체계 개편 방안," 한국보건사회연구원, 한국, 2017.)
- Song, K.J., The fourth industrial revolution of Klaus Schwab, new present, korea, 2016.
- (송경진, 클라우드 슈바의 제4차 혁명, 새로운 현재, 한국, 2016.)
- Steyvers, M. and T. Griffiths, "Provaivistic Topic Models", In T. Landauer, D McNamara, S. Dennis, and W.Kintsch(eds), *Latent Semantic Analysis : A Road to Meaning*, Laurence Erlbaum, 2007, 1-15.
- Suh, H.M. and M.K. Kang, "AAL(Activity Assisted Living) Industry Status and Standardization Trend", *KATS Technical Report*, Vol. 112, 2018, 1-20.
- (서혜미, 강민구, "AAL(능동형 생활지원) 산업 현황 및 표준화 동향", *국가기술표준원 기술보고서*, Vol. 112, 2018, 1-20.)
- Yoon, J.E. and C.J. Suh, "Research Trend Analysis on Smart healthcare by using Topic Modeling and Ego Network Analysis", *Journal of Digital Contents Society*, Vol.19, No.5, 2018, 981-993.
- (윤지은, 서창진, "토픽모델링과 에고 네트워크 분석을 활용한 스마트 헬스케어 연구동향 분석", *디지털콘텐츠학회*, 제19권, 제5호, 2018, 981-993.)

◆ About the Authors ◆



Jee-Eun Yoon (health.jen@gmail.com)

Jee Eun Yoon is a Ph.D. candidate in the Dept. of Business Administration at Hanyang University. She received master degree in global healthcare from Hanyang University. Her current research interests include Healthcare, Healthcare Industry Policy, Big data, Text Mining, Network Analysis, and etc.



Chang-Jin Suh (sjsuh@hanyang.ac.kr)

Chang Jin Suh is professor in Department of Business Administration of Hanyang University. His Current research interests include Health Economics, Medical Management, Medical Policy, Healthcare Industry Policy.