

키워드 빈도 및 중심성 분석 기반의 머신러닝 헬스케어 연구 동향 : 미국 · 영국 · 한국을 중심으로

이택균*

Research Trend on Machine Learning Healthcare Based on Keyword Frequency and Centrality Analysis : Focusing on the United States, the United Kingdom, Korea

Lee Taekkyeun

〈Abstract〉

In this study we analyze research trends on machine learning healthcare based on papers from the United States, the United Kingdom, and Korea. In Elsevier's Scopus, we collected 3425 papers related to machine learning healthcare published from 2018 to 2022. Keyword frequency and centrality analysis were conducted using the abstracts of the collected papers. We identified keywords with high frequency of appearance by calculating keyword frequency and found central research keywords through the centrality analysis by country. Through the analysis results, research related to machine learning, deep learning, healthcare, and the covid virus was conducted as the most central and highly mediating research in each country. As the implication, studies related to electronic health information-based treatment, natural language processing, and privacy in Korea have lower degree centrality and betweenness centrality than those of the United States and the United Kingdom. Thus, various convergence research applied with machine learning is needed for these fields.

Key Words : Machine Learning, Healthcare, Centrality, Frequency

I. 서론

우리나라의 빠른 고령 인구의 증가로 인해서 헬스케어에 대한 관심이 증가하고 있다. 최근 헬스케어 분야는 기존의 환자 치료에서 벗어나 건강인을 대상으로 한 건강 관리 및 Wellbeing 그리고 고령층의 질병 관리 분야로 확대되고 있다[1].

헬스케어 분야에서는 환자의 상태를 모니터링하고 모니터링된 정보의 전송을 위하여 사물인터넷(Internet of Things) 기술을 이용한 연구[2, 3]가 진행되었다. 또한 환자의 헬스케어 데이터[4, 5]에 존재하는 복잡한 패턴을 찾아내기 위하여 머신러닝(Machine Learning) 기법을 적용하였으며 이를 통해서 질병 치료 및 진단에 머신러닝을 이용하는 연구가 진행되었다[6]. 그뿐만 아니라 심장병, 당뇨병 등의

* 아주대학교 다산학부대학 조교수(단독저자)

다양한 질병 예측에 머신러닝을 적용하는 연구[7, 8] 그리고 환자 치료를 위한 신약 개발에 머신러닝을 이용한 연구[9, 10]도 진행되었다. 환자의 헬스케어 정보는 관련된 여러 기관에서 공유되고 있으며 이러한 데이터의 보안을 위해서 머신러닝에 기반한 블록체인 기법을 적용한 연구[11, 12] 및 머신러닝에 기반하여 헬스케어 시스템에 대한 보안 침해 발생을 탐지하는 연구[13, 14]도 진행되었다.

기존의 머신러닝 헬스케어 관련 동향 분석을 살펴보면, 우선 국내 동향 분석에서 김영식[15]은 머신러닝에 기반한 헬스케어를 위한 국제 정책 동향, 시장 규모 그리고 산업체 기술 동향에 대해서 소개를 하였다. 배재권[16]은 헬스케어 분야에서 인공지능의 역할에 관해서 기술하고 국내외 인공지능 헬스케어 산업체 동향을 소개하였다. 이새봄[17]은 헬스케어 산업에 이용되는 인공지능 활용 범위를 구분하고 활용 범위 별로 이용된 인공지능 기술을 국내외 사례를 중심으로 기술하였다.

국의 동향 분석에서 Shailaja[18]는 헬스케어 분야에 대한 머신러닝의 필요성에 관해서 설명하였다. 또한 머신러닝이 적용될 수 있는 여러 헬스케어 분야의 소개와 분야들에 적용될 수 있는 다양한 머신러닝 기법에 관해서 기술하였다. Dhillon[19]은 다양한 헬스케어 데이터를 종류별로 분류하고 각 종류의 데이터 분석을 위해서 이용될 수 있는 머신러닝 기술들을 기술하였다. Ghosal[20]은 클라우드를 기반으로 하는 헬스케어 데이터에 존재하는 보안 문제들을 해결하기 위해서 적용될 수 있는 여러 머신러닝 기법을 소개하였다.

기존의 동향 분석 선행 연구에서 국내 동향 분석에서는 인공지능 헬스케어에 관한 국내 동향 분석[16, 17]이 주로 진행되었으며 인공지능의 많은 세부 분야 중의 하나인 머신러닝에 기반한 헬스케어 동향 분석은 김영식[15]을 제외하면 없는 편이다. 그뿐만 아니라 국내 동향 분석[15-17]에서는 산업체 기술을 중심

으로 기술 동향을 소개하였다. 그러나 연구 논문을 기반으로 하여 동향 분석을 하지 못하였다. 또한 국내와 국외 머신러닝 헬스케어 동향 분석[15-20]에서는 머신러닝 헬스케어 연구를 진행하는 주요 국가 간의 비교를 통해서 시사점을 제시하지 못하는 한계를 가진다.

따라서 본 연구는 연구 논문을 기반으로 하여, 머신러닝 헬스케어 연구를 진행한 주요 국가인 북미 지역의 미국과 유럽의 영국에서 진행된 머신러닝 헬스케어 연구와 비교를 통해서 한국에 집중적으로 연구가 필요한 분야를 확인하고 시사점을 제시하고자 한다.

본 연구에서는 2018년부터 2022년까지 Elsevier의 Scopus에서 머신러닝 헬스케어 관련 미국, 영국 그리고 한국의 논문 총 3425편을 수집하고 수집된 논문의 논문 초록들을 이용하여 키워드 빈도 및 중심성 분석을 진행하였다. 구해진 키워드 빈도에서 빈도가 높은 키워드를 확인하며 중심성 분석에 의해서 중심적인 연구 키워드를 파악하고자 하였다. 중심성 분석에는 대표적인 연결 중심성과 매개 중심성을 사용하였다. 키워드 빈도와 중심성 분석에 기반하여 국가별 비교하였으며 또한 한국의 머신러닝 기반의 헬스케어 관련 연구 방향성을 제시하기 위해서 한국 관련 시사점을 도출하였다.

본 연구의 구성을 보면, 관련 연구로 머신러닝 헬스케어 동향 분석 그리고 중심성에 대한 내용을 2장에서 소개하였으며 연구 질문, 자료 수집, 자료 전처리 및 분석 방법은 3장에서 기술을 하였다. 키워드 빈도, 중심성 분석 결과는 4장에서 그리고 5장에서 분석 결과에 대해서 논의하였으며 6장에는 결론으로 마무리하였다.

II. 관련 연구

2.1 머신러닝 헬스케어 동향 분석

머신러닝 헬스케어에 관한 국내 동향 분석에서, 김영식[15]은 머신러닝 헬스케어에 대한 주요국의 정책 및 시장 동향을 설명하였다. 향후에는 머신러닝 헬스케어의 시장 규모가 급격하게 증가할 것으로 예상하였으며 머신러닝 헬스케어 기술을 개발하는 글로벌 선도 기업들의 기술 동향에 관해서도 소개하였다.

배재권[16]은 미국 및 유럽의 주요 국가에서 추진하는 주요 헬스케어 산업 동향과 국내 헬스케어 산업 동향에 대해서 소개하였다. 또한 유전체 정보, 임상 정보, 생활환경 정보 및 습관 정보 등을 인공지능으로 분석하여 환자 개인을 위한 최적의 맞춤형 의료 서비스를 제공하는 정밀 의료에 대한 사례연구를 제시하였다. 이를 통하여 정밀 의료 분야에 대한 투자 필요성도 강조하였다.

이새봄[17]은 헬스케어 산업을 주요 분야로 구분하고 분야별로 인공지능이 적용된 국내외 활용사례를 소개하였다. 즉, 병원, 개인, 보험, 의약품을 관련 주요 분야로 구분하였으며 분야별 활용사례로 인공지능과 의료데이터를 결합한 의료 진단 보조 솔루션, 스마트폰으로 환자의 생체 정보를 분석하여 건강 관리를 돕는 앱, 생명보험에서 보험계약 심사를 위한 인공지능 시스템, 유전체 빅데이터 및 인공지능 플랫폼을 이용한 신약 개발 등을 소개하였다.

머신러닝 헬스케어에 관한 국외 동향 분석에, Shailaja[18]는 다양하며 방대한 데이터로 구성된 헬스케어 데이터에서 복잡한 패턴을 찾기 위하여 머신러닝 기술의 필요성을 제시하였다. 또한 머신러닝이 적용될 수 있는 여러 헬스케어 분야를 소개하고 분야별로 정확한 질병의 진단을 위해서 적용될 수 있는 여러 머신러닝 기술을 소개하였다. 구체적으로 심장병, 당뇨병, 갑상선 등의 질환의 진단을 위해서 머신

러닝이 적용된 연구들도 기술하였다. Dhillon[19]은 헬스케어 데이터를 임상데이터, 유전자 데이터, 단백질 데이터 등으로 구분하고 각 구분된 데이터의 분석을 위해서 이용될 수 있는 머신러닝 기법들을 소개하였다.

Ghosal[20]은 클라우드 환경의 헬스케어 데이터에 존재하는 여러 보안 문제들을 해결하기 위한 다양한 머신러닝 기법들을 소개하였다. 즉, 클라우드 환경의 데이터에 대해서 데이터 유출, 데이터 손실 등의 여러 데이터 보안 문제들이 존재하며 이러한 보안 문제들의 해결을 위해서 적용될 수 있는 다양한 머신러닝 기법들에 관해서 소개하였다.

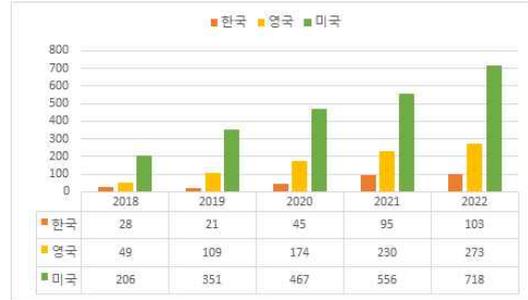
2.2 중심성

중심성(Centrality) 분석이란 텍스트로부터 추출한 키워드를 사용하여 키워드 사이의 관계를 네트워크로 표현하고 이러한 네트워크에서 키워드 사이의 관계를 분석하는 기법이다[21]. 이러한 키워드 네트워크는 링크 및 노드로 구성되며 노드는 키워드를 나타내고 링크는 키워드 사이의 관계를 표현한다. 중심성 분석에 기반하여 키워드 네트워크로부터 중심적인 키워드를 파악하고 이러한 중심적인 키워드 파악을 통해서 중심성 분석은 연구 동향과 트렌드 분석에 많이 이용된다[22, 23].

연결 중심성(Degree Centrality) 및 매개 중심성(Betweenness Centrality)을 본 연구의 중심성 분석을 위해서 사용하였다. 연결 중심성은 키워드 네트워크에서 노드의 연결 정도를 나타내는 지표이며 하나의 노드가 다른 노드들과 얼마나 연결되었는지를 나타낸다. 따라서 키워드 네트워크에서 노드가 다른 노드들과 많이 연결될수록 해당 노드의 연결 중심성은 높아지며 다른 노드들과 많은 연관성이 존재함을 나타낸다.[24].

매개 중심성은 키워드 네트워크에서 하나의 노드

가 다른 노드 사이의 매개적인 역할을 하는 정도를 나타내는 지표이다. 따라서 하나의 노드가 다른 노드 사이의 최단 경로 중에 존재할수록 해당 노드의 매개 중심성은 높아지며 매개 중심성이 높을수록 한 연구에서 다른 키워드들과 같이 고려되는 정도가 높아진다[25, 26].



<그림 1> 수집 논문 현황

III. 연구 방법

3.1 연구 질문

머신러닝을 활용한 헬스케어 관련 연구 동향 파악을 위해서 본 연구에서는 아래와 같은 연구 질문을 설정한다.

연구 질문 1: 머신러닝을 활용한 헬스케어에 대한 키워드 빈도 결과는 어떠한가?

연구 질문 2: 머신러닝을 활용한 헬스케어에 대한 연결 중심성과 매개 중심성 분석 결과는 어떠한가?

연구 질문 3: 머신러닝을 활용한 헬스케어에 대한 연결 중심성과 매개 중심성 분석 결과로부터 도출되는 한국 관련 시사점은 무엇인가?

3.2 자료 수집

자료 수집은 ‘machine learning’ 및 ‘healthcare’를 검색 키워드로 사용하여 2018년부터 2022년까지 발표된 논문을 대상으로 하여 Elsevier의 Scopus에서 수집하였다.

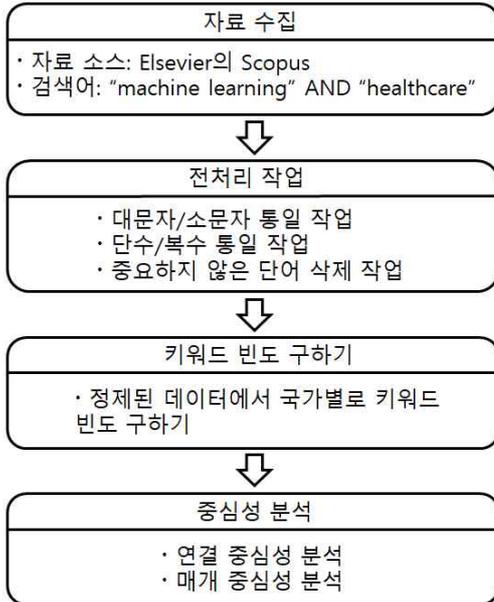
국가별로 수집된 논문 현황을 <그림 1>에 나타내었다. 국가별로 수집된 논문 편수는 2,298편(미국), 835편(영국), 292편(한국)이다. 수집된 논문 편수에서 미국이 발표한 논문 편수가 가장 많은 편이며 한국이 발표한 논문 편수가 가장 적은 편이다. 미국이 발표한 논문은 2018년부터 2022년까지 지속적으로 증가

였으며 영국이 발표한 논문 편수도 2018년부터 2022년까지 점진적으로 증가를 하였다. 그에 비해서 한국이 발표한 논문 편수는 2019년에 감소하였으나 2020년부터 2022년까지 점진적으로 증가하였다. 영국과 한국이 연도별로 발표한 논문 편수는 미국이 연도별로 발표한 논문 편수의 절반보다 적다. 연도별로 수집된 논문 편수는 283편(2018년), 481편(2019년), 686편(2020년), 881편(2021년), 1,094편(2022년)이며 논문 편수는 2018년부터 계속해서 증가하였다.

3.3 자료 전처리 및 분석 방법

본 연구와 관련된 전체적인 절차는 <그림 2>에 나타내었다. 자료 수집을 통해서 수집 논문들의 논문 초록으로부터 단어(키워드)를 추출하고 추출된 키워드를 대상으로 전처리 작업을 진행하였다.

전처리 작업에서 영어 대문자와 소문자를 일치시키는 작업이 진행되었으며 같은 의미의 단어 및 복수 단어를 통일하는 작업도 진행되었다. 그뿐만 아니라 숫자, 기호, 구두점 등은 불필요한 내용이므로 제거되었고 부사 및 관사 등의 의미가 중요하지 않은 단어들도 제거되었다. 전처리 작업으로부터 얻은 데이터를 이용하여 키워드를 추출하였고 추출된 키워드부터 키워드 빈도를 구하였다. 키워드 빈도는 국가별로 구하였으며 빈도가 높은 상위 15개 키워드를 선별



<그림 2> 전체적인 절차

하였다.

본 연구에서는 중심적인 연구 키워드를 파악하기 위해서 행위자 간의 상호 작용에 의해서 구축되는 연결망을 이용하여 키워드 간의 관계를 분석하는 사회망분석(Social Network Analysis)을 적용하였으며 사회망분석에서 많이 이용되는 중심성 분석 기법을 이용하였다. 중심성 분석을 위해서 연결 중심성 및 매개 중심성을 이용하였다.

연결 중심성을 위해서 키워드 네트워크에서 하나의 노드(키워드)가 얼마나 많은 다른 노드와 연결되었는지를 측정하였다. 즉, 30개 노드로 구성된 키워드 네트워크에서 각 노드별로 다른 노드와 연결된 링크 수를 세고 그리고 노드별로 연결된 링크 수가 0 ~ 1 사이의 값이 되도록 정규화된 값으로 나타내었다. 매개 중심성은 키워드 네트워크에서 노드의 매개적 역할을 하는 정도를 나타낸 것이며 30개 노드로 구성된 키워드 네트워크에서 해당 노드가 다른 두 노드 사이에 얼마나 많이 등장하는지를 측정하여 매개 중심성

을 구하였다. 따라서 키워드 네트워크를 통해서 30개 키워드에 대한 키워드별로 연결 중심성 값과 매개 중심성 값을 구하였으며 30개 키워드 중에서 연결 중심성이 높은 15개 키워드를 국가별로 분석하였다. 또한 매개 중심성이 높은 상위 15개 키워드를 국가별로 분석하였다. 전처리 과정, 키워드 빈도, 중심성 분석을 위해서 프로그래밍 언어 R을 사용하였다.

IV. 분석 결과

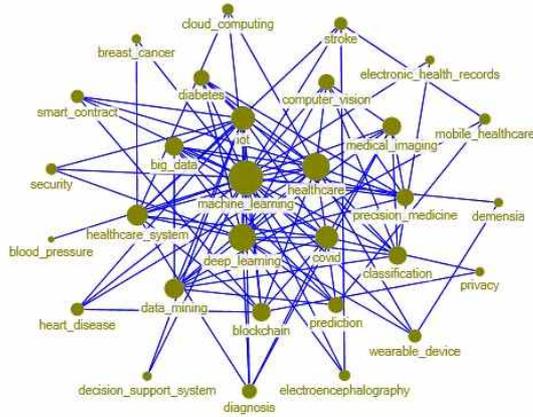
4.1 키워드 빈도 결과

키워드를 국가별로 추출하였으며 추출된 키워드를 이용하여 키워드 빈도를 구하였다. 빈도가 가장 높은 상위 15개 키워드를 <표 1>에 나타내었다.

<표 1>에서 상위 5개 키워드를 살펴보면, 각 국가에서 키워드 'machine_learning', 'deep_learning',

<표 1> 국가별 상위 빈도 키워드

순위	미국	영국	한국
1	machine_learning	machine_learning	machine_learning
2	deep_learning	deep_learning	deep_learning
3	healthcare	healthcare	healthcare
4	covid	covid	covid
5	iot	iot	iot
6	natural_language_processing	big_data	big_data
7	big_data	classification	blockchain
8	electronic_health_records	electronic_health_records	data_mining
9	predictive_analytics	precision_medicine	healthcare_system
10	data_mining	data_mining	classification
11	prediction	diagnosis	precision_medicine
12	precision_medicine	digital_health	prediction
13	classification	mental_health	diabetes
14	diabetes	natural_language_processing	dementia
15	diagnosis	prediction	diagnosis



<그림 5> 한국의 키워드 네트워크

<그림 5>는 한국의 논문 초록으로부터 추출된 키워드 중에서 빈도가 높은 30개 키워드로 구성된 키워드 네트워크이다. <그림 5>의 네트워크를 사용하여 각 키워드의 연결 중심성 및 매개 중심성을 구하였다.

4.2.2 연결 중심성 분석

<그림 3>, <그림 4>, <그림 5>의 키워드 네트워크를 사용하여 국가별로 연결 중심성을 구하였고 또한 국가별로 연결 중심성이 높은 상위 15개 키워드를 <표 2>에 나타내었다.

<표 2>의 연결 중심성 순위는 각 키워드가 다른 연구 키워드와 얼마나 연관성이 높은 중심적인 연구 키워드인지를 국가별로 나타내며 이를 통하여 키워드 간의 연결 중심성을 비교하였다. <표 2>에는 다양한 분야의 키워드들이 포함되어 있다. 따라서 이러한 키워드들을 분야별로 파악하기 위해서 <표 2>의 키워드들을 관련 분야별로 구분하였다. 그래서 머신러닝 관련 분야(분야 A), 헬스케어 관련 분야(분야 B), 사물인터넷 및 빅데이터 관련 분야(분야 C), 질병 관련 분야(분야 D), 예측 진단 관련 분야(분야 E), 기타 분야(분야 F)로 구분하여 <표 3>에 정리하였다.

<표 2> 국가별 상위 연결 중심성 키워드

순위	미국	영국	한국
1	machine_learning (1.0)	machine_learning (1.0)	machine_learning (0.931)
2	deep_learning (0.931)	deep_learning (0.758)	deep_learning (0.586)
3	healthcare (0.931)	healthcare (0.758)	healthcare (0.586)
4	big_data (0.793)	covid (0.517)	iot (0.448)
5	covid (0.724)	big_data (0.517)	covid (0.413)
6	iot (0.620)	iot (0.413)	data_mining (0.310)
7	electronic_health_records (0.620)	electronic_health_records (0.413)	healthcare_system (0.344)
8	predictive_analytics (0.620)	precision_medicine (0.379)	big_data (0.275)
9	diagnosis (0.620)	diagnosis (0.379)	blockchain (0.275)
10	natural_language_processing (0.586)	classification (0.344)	classification (0.275)
11	data_mining (0.517)	data_mining (0.344)	medical_imaging (0.275)
12	prediction (0.517)	prediction (0.344)	precision_medicine (0.241)
13	diabetes (0.517)	privacy (0.344)	prediction (0.206)
14	digital_health (0.517)	digital_health (0.310)	diabetes (0.206)
15	public_health (0.517)	blockchain (0.310)	computer_vision (0.206)

<표 2>에서 각 국가에서 공통으로 보이는 상위 순위 5위 안에 포함되는 키워드는 'machine_learning', 'covid', 'deep_learning', 'healthcare'이며 이 키워드들은 각 국가에서 연결 중심성이 높은 키워드들이다. 이 키워드 중에서 'machine_learning', 'deep_learning', 'healthcare'가 각 국가에서 연결 중심성이 가장 높은 상위 순위 3위 안에 포함되는 키워드들이다.

<표 3>에서 머신러닝 관련 분야(분야 A)의 키워드를 보면, 'machine_learning', 'deep_learning'이 공통

<표 3> 분야에 따른 국가별 연결 중심성 키워드

분야	미국	영국	한국
A	machine_learning, deep_learning	machine_learning, deep_learning, classification	machine_learning, deep_learning, classification
B	healthcare, electronic_health_records, digital_health, public_health	healthcare, electronic_health_records, digital_health	healthcare, healthcare_system
C	big_data, iot, data_mining	big_data, iot, data_mining	iot, data_mining, big_data
D	covid, diabetes	covid	covid, diabetes
E	predictive_analytics, diagnosis, prediction	precision_medicine, diagnosis, prediction	precision_medicine, prediction
F	natural_language_processing	privacy, blockchain	blockchain, medical_imaging, computer_vision

으로 나타났으며 이를 통해서 국가마다 이 키워드와 관련된 연구가 이 분야에서 가장 연결 중심적으로 진행된 것으로 보인다. 또한 영국과 한국에서는 키워드 'classification'이 보이며 머신러닝 분야 중 하나인 분류(classification)와 관련된 연구가 공통으로 진행된 것으로 파악된다. 또한 이 분야 키워드들의 연결 중심성 순위를 <표 2>에서 보면, 키워드 'machine_learning'의 연결 중심성이 각 국가에서 가장 높은 것을 볼 수 있다.

<표 3>의 헬스케어 관련 분야(분야 B)의 키워드에서는 'healthcare'가 각 국가에서 공통으로 나타났다. 또한 'electronic_health_records', 'digital_health'가 미국과 영국에서는 공통으로 나타났으며, 따라서 미국과 영국에서는 전자 건강 기록(electronic health records) 관련 연구 및 정보 통신 기술을 융합한 디지털 헬스 케어 관련 연구가 중심적으로 진행된 것으로 보인다. 연결 중심성 순위를 <표 2>에서 보면, 'electronic_health_records'의 연결 중심성이 'digital_health'의 연결 중심성보다 높다.

사물인터넷 및 빅데이터 관련 분야(분야 C)의 키워드를 보면, 'big_data', 'iot', 'data_mining'이 공통으로 보이며 각 국가에서 이 키워드와 관련된 연구가 연결 중심적으로 진행된 것으로 파악된다. <표 2>에서 이 키워드들의 연결 중심성 순위를 살펴보면, 미국과 영국의 경우에는 'big_data'의 연결 중심성 순위가 'iot', 'data_mining'보다 높으며 한국의 경우에는 'iot'의 순위가 'big_data', 'data_mining'보다 높다. 이를 통해서 이 분야에서 미국과 영국에서는 빅데이터 관련 연구가 가장 연결 중심적인 연구로 진행되었으며 한국의 경우에는 사물인터넷 관련 연구가 가장 연결 중심적인 연구로 진행된 것으로 판단된다.

질병 관련 분야(분야 D)의 키워드에서는 'covid'가 공통으로 보이며 <표 2>에서 'covid'의 연결 중심성 순위를 보면 이 분야 키워드 중에서 연결 중심성이 가장 높다. 따라서 이 분야에서 covid 바이러스 관련 연구가 국가마다 가장 연결 중심적으로 진행된 것으로 보인다. 또한 미국과 한국에서는 키워드 'diabetes'가 공통으로 보이며 두 나라에서 당뇨병과 관련된 연구가 연결 중심적으로 진행된 것으로 파악된다.

예측 진단 관련 분야(분야 E)의 키워드를 보면, 국가마다 'prediction'이 나타났으며 이를 통해서 질병 예측과 관련된 연구가 연결 중심적인 연구로 국가마다 진행된 것으로 보인다. 또한 미국과 영국에서는 키워드 'diagnosis'가 공통으로 보이며 질병 진단에 관한 연구가 연결 중심적으로 진행된 것으로 판단된다. 영국 및 한국에서는 키워드 'precision_medicine'이 나타났으며 정밀의학(precision medicine) 관련 연구가 두 나라에서 연결 중심적인 연구로 진행된 것으로 보인다.

기타 분야(분야 F)의 키워드에서 키워드 'blockchain'이 영국과 한국에서 나타났으며 두 나라에서 헬스케어 정보 보안을 위한 블록체인과 관련된 연구가 연결 중심적인 연구로 진행된 것으로 판단된다. 또한 영국에서는 키워드 'privacy'가 보이며 헬스

케어 개인 정보 보호를 위한 연구가 연결 중심적인 연구로 이루어진 것으로 보인다.

4.2.3 매개 중심성 분석

<그림 3>, <그림 4>, <그림 5>에 있는 국가별 키워드 네트워크를 사용하여 매개 중심성을 구하였고 <표 4>에는 국가별로 매개 중심성이 높은 상위 15개 키워드를 나타내었다. <표 4>의 키워드 매개 중심성 순위는 각 키워드가 국가별로 얼마나 매개적인 연구 키워드인지를 의미하며 이를 통하여 키워드 간의 매개 중심성을 비교하였다.

<표 4>는 다양한 분야의 키워드들을 포함하므로 이러한 키워드들을 분야별로 파악하기 위해서 <표 4>의 키워드들을 관련 분야별로 구분하였으며 머신러닝 관련 분야(분야 A), 헬스케어 관련 분야(분야 B), 사물인터넷 및 빅데이터 관련 분야(분야 C), 질병 관련 분야(분야 D), 예측 진단 관련 분야(분야 E), 기타 분야(분야 F)로 구분하여 <표 5>에 정리하였다.

<표 4>에서 국가마다 나타난 상위 순위 5위에 포함된 키워드는 'machine_learning', 'deep_learning', 'covid', 'healthcare' 이며 이 키워드들은 매개 중심성이 높은 키워드들이다. 이 키워드와 관련된 연구는 높은 매개적인 역할을 한 연구들로 보인다.

<표 5>에서 머신러닝 관련 분야(분야 A)의 키워드를 보면, 키워드 'machine_learning', 'deep_learning' 이 각 국가에서 매개 중심성 키워드로 나타났으며 이 키워드들의 매개 중심성 순위를 <표 4>에서 살펴보면, 높은 매개 중심성을 보이며, 따라서 국가마다 머신러닝과 딥러닝 관련 연구가 높은 매개적인 역할을 하는 연구로 진행된 것으로 파악된다.

<표 5>의 헬스케어 관련 분야(분야 B)의 키워드에서는 'healthcare'가 국가마다 매개 중심성 키워드로 나타났다. 키워드 'electronic_health_records'가 미국과 영국에서 보이며, 따라서 두 나라에서는 전자 건

<표 4> 국가별 상위 매개 중심성 키워드

분야	미국	영국	한국
1	machine_learning (0.195)	machine_learning (0.607)	machine_learning (0.955)
2	deep_learning (0.167)	healthcare (0.213)	deep_learning (0.158)
3	healthcare (0.127)	deep_learning (0.192)	healthcare (0.154)
4	big_data (0.083)	covid (0.069)	covid (0.115)
5	covid (0.056)	big_data (0.060)	iot (0.086)
6	iot (0.053)	diagnosis (0.036)	data_mining (0.071)
7	predictive_analytics (0.045)	iot (0.028)	healthcare_system (0.060)
8	electronic_health_records (0.038)	precision_medicine (0.025)	blockchain (0.055)
9	diagnosis (0.035)	privacy (0.025)	prediction (0.033)
10	natural_language_processing (0.029)	mental_health (0.020)	stroke (0.031)
11	digital_health (0.029)	electronic_health_records (0.020)	classification (0.023)
12	data_mining (0.028)	dementia (0.019)	medical_imaging (0.012)
13	public_health (0.017)	classification (0.015)	big_data (0.008)
14	diabetes (0.017)	natural_language_processing (0.012)	precision_medicine (0.005)
15	prediction (0.015)	health_informatics (0.010)	wearable_device (0.004)

강 기록(electronic health records)과 관련된 연구가 매개적인 역할을 한 것으로 파악된다.

사물인터넷 및 빅데이터 관련 분야(분야 C)의 키워드를 보면, 각 국가에서 키워드 'big_data', 'iot'가 보이며 공통으로 사물인터넷 및 빅데이터 관련 연구가 공통으로 매개적인 역할을 한 것으로 보인다. 이 키워드들의 매개 중심성 순위를 <표 4>에서 보면, 미국과 영국의 경우에 'big_data'의 매개 중심성은 'iot'보다 높은 편이며, 한국의 경우에 'iot'의 매개 중심성이 'big_data'보다 높은 편이다. 미국과 한국에서는 키워

<표 5> 분야에 따른 국가별 매개 중심성 키워드

분야	미국	영국	한국
A	machine_learning, deep_learning	machine_learning, deep_learning, classification	machine_learning, deep_learning, classification
B	healthcare, electronic_health_ records, digital_health, public_health	healthcare, electronic_health_ records, mental_health, health_informatics	healthcare, healthcare_system
C	big_data, iot, data_mining	big_data, iot	iot data_mining, big_data
D	covid, diabetes	covid, dementia	covid, stroke
E	predictive_analytics, diagnosis, prediction	diagnosis, precision_medicine	prediction, precision_medicine
F	natural_language_ processing	privacy, natural_language_pr ocessing	blockchain, medical_imaging, wearable_device

드 'data_mining'이 공통으로 나타났으며 두 나라에서는 데이터마이닝 관련 연구가 공통으로 매개적인 연구로 진행된 것으로 파악된다.

질병 관련 분야(분야 D)의 키워드에서는 'covid'가 국가마다 나타났으며 covid 바이러스와 관련된 연구가 공통으로 매개적인 역할을 한 것으로 보이며, 또한 이 분야 키워드들의 매개 중심성 순위를 <표 4>에서 보면, 각 국가에서 'covid'의 매개 중심성이 가장 높다.

예측 진단 관련 분야(분야 E)의 키워드를 보면, 미국과 영국에서는 키워드 'diagnosis'가 나타났고 미국 및 한국에서는 'prediction'이 보이며 영국과 한국에서 'precision_medicine'이 공통으로 나타났다. 이를 통해서 미국과 영국에서는 질병 진단 관련 연구, 미국과 한국에서는 질병 예측 관련 연구 그리고 영국과 한국에서는 정밀 의학 관련 연구가 매개적인 역할을 한 연구로 진행된 것으로 판단된다.

기타 분야(분야 F)의 키워드를 살펴보면, 미국과 영국에서는 'natural_language_processing'이 나타났

으며 두 나라에서 자연어 처리와 관련된 연구가 매개적인 연구로 진행된 것으로 보인다. 또한 영국의 키워드 'privacy'를 통해서 영국에서는 개인 정보 보호에 관한 연구가 매개적인 역할을 하는 연구로 진행되었으며 한국의 키워드 'blockchain'를 통해서 한국에서는 개인 의료 정보 보호와 연관된 블록체인에 관한 연구가 매개적인 역할을 하는 연구로 파악된다.

V. 논의

본 장에서는 연구 질문들에 관한 내용을 논의하고자 한다. 첫째, 머신러닝을 활용한 헬스케어에 관한 키워드 빈도를 구한 결과에서 'machine_learning', 'covid', 'iot', 'deep_learning', 'healthcare'는 국가마다 빈도 순위 5위에 포함되는 키워드이며 높은 관심을 가지는 연구 키워드들로 보인다. 그 외에도 상위 5개 키워드에 포함되지 않았으나 'big_data', 'prediction', 'precision_medicine', 'classification', 'data_mining', 'diagnosis'도 각 국가에서 공통으로 관심을 가지는 연구 키워드들이다.

둘째, 머신러닝을 활용한 헬스케어에 관한 연결 중심성 및 매개 중심성 분석 결과에 대해서 전체적으로 살펴보면, 'machine_learning', 'deep_learning', 'healthcare', 'covid'는 각 국가에서 상위 순위 5위 안에 포함되며 연결 중심성 및 매개 중심성이 높은 키워드들이다. 이를 통해서 머신러닝과 머신러닝 분야 중 하나인 딥러닝을 활용한 헬스케어 관련 연구가 가장 중심적인 연구로 진행되었다. 또한 머신러닝을 활용한 covid 바이러스 관련 연구도 높은 중심성을 가진 연구로 진행된 것으로 보인다.

분야별로 구분하여 논의하면, 머신러닝 관련 분야에서 키워드 'machine_learning', 'deep_learning'이 각 국가에서 상위 연결 중심성과 매개 중심성 키워드로 나타났다. 이를 통해서 머신러닝 및 딥러닝 관련

연구가 각 국가에서 가장 중심적이며 매개적인 연구로 진행된 것으로 파악된다. 또한 영국과 한국에서는 머신러닝 분야 중 하나인 분류(classification) 관련 연구가 공통으로 중심적이고 매개적인 역할을 하는 연구로 진행된 것으로 보인다.

헬스케어 분야의 키워드 'healthcare'가 국가마다 상위 연결 중심성 및 매개 중심성 키워드로 나타났으며 따라서 각 국가에서 헬스케어 관련 연구가 중심적이며 매개적인 연구로 진행된 것으로 파악된다. 또한 미국과 영국에서는 전자 건강 기록(electronic health records) 관련 연구가 중심적이며 매개적인 연구로 진행된 것으로 보이며 한국에서는 헬스케어를 위한 시스템 관련 연구가 중심적이고 매개적인 연구로 진행된 것으로 파악된다.

사물인터넷 및 빅데이터 관련 분야의 경우에 키워드 'iot', 'big_data'가 상위 연결 중심성 및 매개 중심성 키워드로 각 국가에서 나타났다. 따라서 사물인터넷과 빅데이터 관련 연구가 각 국가에서 중심적이며 매개적인 연구로 진행된 것으로 보인다. 이 분야에서 미국 및 영국의 경우에는 빅데이터 관련 연구가 그리고 한국에서는 사물인터넷 관련 연구가 더 중심적이며 매개적인 역할을 한 것으로 파악된다. 키워드 'data_mining'은 상위 연결 중심성 키워드로 각 국가에서 나타났으며 상위 매개 중심성 키워드로는 미국과 한국에서만 나타났다. 따라서 데이터마이닝 연구는 각 국가에서 연결 중심적인 연구로 진행되었으며 영국에서는 데이터마이닝 연구가 미국과 한국에 비해서 낮은 매개적인 역할을 하는 것으로 보인다.

질병 관련 분야에서 키워드 'covid'가 국가마다 상위 연결 중심성 및 매개 중심성 키워드로 나타났으며 머신러닝을 이용한 covid 바이러스 관련 연구가 이 분야에서 가장 중심적이고 매개적인 연구로 진행된 것으로 파악된다. 또한 미국의 경우에는 머신러닝 기반의 당뇨병 연구도 중심적이며 매개적인 역할을 하는 연구로 진행된 것으로 보인다. 임동식[27]에 따르

면 2015년 기준 미국 당뇨병 환자는 3030만 명으로 전체 인구의 9.4%이며 당뇨병 전 단계에 있는 인구도 8410만 명에 달했다. 즉, 미국 인구 중에서 3분의 1 이상이 당뇨병 또는 당뇨병 전단계인 것으로 나타났으며 앞으로 더욱 증가할 것으로 예상하였다. 또한 자료[28]에 의하면 2019년 기준 당뇨병은 미국인의 주요 사망원인으로 10위 안에 포함되었다. 따라서 이러한 이유로 미국의 경우에 머신러닝 기반의 당뇨병 연구가 중심적인 연구로 진행된 것으로 보인다.

예측 진단 관련 분야에서 미국과 영국에서는 머신러닝을 적용한 질병 진단 관련 연구, 영국 및 한국에서는 정밀 의학 관련 연구, 미국과 한국에서는 질병 예측과 관련된 연구가 중심적이며 매개적인 역할을 하는 연구로 진행된 것으로 파악된다. 박혜경[29]의 헬스케어 분야 현황에서, 한국의 경우에는 머신러닝을 포함한 인공지능 기술이 질병과 관련된 예측 연구에 많이 적용되어서 연구들이 진행되었다. 이러한 이유로 질병 예측 관련 연구가 한국에서는 중심으로 진행된 것으로 파악된다.

기타 분야 키워드를 보면, 미국에서는 자연어 처리를 위하여 머신러닝을 적용한 연구가 중심적이고 매개적인 역할을 하는 연구로 진행되었으며 영국에서는 머신러닝을 이용한 개인 정보 보호 관련 연구 그리고 한국에서는 머신러닝 기반의 블록체인 관련 연구가 중심적이고 매개적인 연구로 진행된 것으로 보인다.

셋째, 머신러닝을 활용한 헬스케어에 관한 중심성 분석 결과를 통하여 도출되는 한국과 관련된 시사점을 보면, 헬스케어 분야에서 'electronic_health_records'가 한국의 경우에 상위 연결 중심성 및 상위 매개 중심성 키워드로 포함되지 않았다. 이를 통해서 미국과 영국에 비해서 한국의 전자 건강 기록 관련 연구가 연결 중심적이며 매개 중심적인 연구로 진행되는 정도가 낮은 것으로 파악된다. 따라서 한국의 전자 건강 기록 관련 연구가 보다 중심적이며 매개적

인 연구가 되도록 하기 위해서 방대한 환자의 전자 건강 기록 정보를 머신러닝을 이용하여 분석하여 환자 개인별로 맞춤형 진료 및 치료를 위한 연구가 필요해 보인다.

질병 관련 분야에서 한국의 당뇨병 관련 연구가 미국에 비해서 매개적인 역할이 낮은 것으로 보이며 더욱 매개 중심적인 연구로 발전하기 위해서 다양한 분야와 융합 연구를 통해서 당뇨병과 관련된 집중적인 연구가 필요해 보인다. 특히, 머신러닝을 적용한 당뇨병의 예측, 진단 그리고 치료 관련 연구가 필요해 보인다. 또한 한국의 경우에는 영국에 비해서 치매 관련 연구가 연결 중심성과 매개 중심성이 낮은 연구로 파악되며 환자의 뇌 관련 데이터를 머신러닝 기법으로 분석하여 치매 조기 진단과 관련된 활발한 연구가 필요해 보인다.

기타 분야 키워드에서 'natural_language_processing'이 한국의 경우에 상위 연결 중심성 및 상위 매개 중심성 키워드로 포함되지 않았으며 미국에 비해서 자연어 처리 관련 연구가 연결 중심성과 매개 중심성이 낮은 연구로 파악된다. 따라서 머신러닝에 기반하여 자연어 처리 기술을 적용한 원격 진료 서비스에 관한 연구 그리고 음성 인식 기술을 적용한 의무 기록 작성 시스템에 관한 연구가 집중적으로 진행할 필요가 있어 보인다. 또한 영국에 비해서 한국의 개인 정보 보호 관련 연구가 연결 중심성과 매개 중심성이 낮은 것으로 보이며, 따라서 환자 개인 정보 보호를 위한 안전한 네트워크 및 시스템에 대한 연구 그리고 인증된 도메인 간에 환자 의료 정보 교환을 지원할 수 있는 보안 기술에 관한 연구도 필요해 보인다.

VI. 결론

본 연구는 Elsevier의 Scopus에서 2018년부터 2022년까지 머신러닝을 활용한 헬스케어 관련 논문의 연

구 동향을 미국, 영국, 한국 중심으로 파악하고자 하였다.

머신러닝을 활용한 헬스케어에 관한 연결 중심성과 매개 중심성 분석 결과에 의하면 각 국가에서 머신러닝, 딥러닝 그리고 헬스케어에 관한 연구가 가장 중심적이며 매개적인 역할을 하는 연구로 진행되었고 covid 바이러스 관련 연구도 높은 중심성 및 매개성을 가진 연구로 진행된 것으로 파악된다.

머신러닝을 활용한 헬스케어에 관한 중심성 분석을 통하여 도출되는 한국 관련 시사점을 보면, 헬스케어 분야에서는 환자의 전자 건강 기록 정보를 머신러닝을 통해서 분석하여 환자 맞춤형 진료 및 치료를 위한 연구가 필요해 보이며 질병 관련 분야에서는 당뇨병, 치매 등의 질병의 정확한 예측, 진단 그리고 치료를 위해서 머신러닝을 적용한 집중적인 연구가 필요하다. 구체적으로 학문적 그리고 실무적으로 나누어 살펴보면, 실무적으로는 환자의 상태를 지속적으로 모니터링할 수 있는 단말 시스템, 모니터링된 방대한 데이터를 전송할 수 있는 통신 시스템, 전송된 많은 데이터를 관리하고 검색할 수 있는 시스템의 구축이 필요하다. 학문적으로는 이러한 구축된 시스템으로부터 제공되는 방대한 건강 기록 정보를 다양한 머신러닝 기법들을 통해서 분석하여 정확한 환자 맞춤형 진단과 치료 방법을 위한 연구가 필요해 보인다.

기타 분야 키워드에서는 자연어 처리와 관련하여 실무적으로는 사용자의 음성을 텍스트로 변환하는 시스템, 변환된 정보를 저장하고 관리하는 서버 시스템의 개발 등이 필요해 보이며 학문적으로는 정교한 음성 인식을 위한 음성 인식 모델의 개발에 관한 연구가 필요해 보인다. 또한 개인 정보 보호와 관련하여 실무적으로는 안전한 개인 정보 보호를 위한 보안 시스템의 개발이 필요하며 학문적으로는 인증된 도메인 간에 환자 의료 정보를 교환할 수 있는 보안 기술에 대한 연구가 필요하다.

미래에는 다양한 헬스케어 분야에 머신러닝을 포함한 인공지능 기술이 더욱 많이 적용될 것으로 보이며 특히 인공지능(머신러닝 포함)을 기반으로 하여 의료 영상 처리를 통한 진단, 음성 인식과 같은 자연어처리를 활용한 전자 진료 기록 등의 분야와 관련된 연구가 집중적으로 진행될 것이며 이러한 연구와 관련된 키워드가 미래에는 많이 나타날 것으로 예상된다.

이렇게 도출된 시사점을 통해서 한국의 머신러닝 및 헬스케어 관련 연구의 방향성을 제시하고자 한다. 그리고 본 논문의 한계점은 머신러닝을 활용한 헬스케어에 관한 논문의 연구 동향 분석을 미국, 영국, 한국을 중심으로 진행하였으며 향후에는 활발하게 연구를 진행한 여러 국가를 포함하여 다양한 지역의 머신러닝을 활용한 헬스케어에 관한 연구를 동향 분석을 하고자 한다.

참고문헌

- [1] 김기봉 · 한근희, “4차 산업혁명시대의 디지털 헬스케어 산업에 대한 연구,” 융합정보논문지, 제10권, 제3호, 2020, pp.7-15.
- [2] 서승범 · 이재상 · 구민정, “스마트폰을 이용한 병원 진료예약 및 진료시스템 연구,” 문화기술의 융합, 제3권, 제1호, 2017, pp.29-33.
- [3] 김다원 · 남희조 · 이승연 · 함유경 · 서오석 · 이형준, “웨어러블 디바이스와 키넥트 센서를 활용한 다중 사용자 홈트레이닝 헬스케어 시스템,” 한국통신학회논문, 제44권, 제4호, 2019, pp.719-727.
- [4] 박유상 · 최종선 · 최재영, “IoT환경에서 헬스케어 서비스 제공을 위한 기기종 센서데이터 수집 모델,” 정보처리학회논문지(소프트웨어 및 데이터 공학), 제6권, 제2호, 2017, pp.77-84.
- [5] 정윤수 · 한근희, “헬스케어 정보 수집을 위한 병원감 데이터 통합 모델 설계,” 한국융합학회논문지, 제9권, 제6호, 2018, pp.1-7.
- [6] Sukanya, M., “Applications of Big Data Analytics and Machine Learning Techniques in Healthcare Sectors,” International Journal of Engineering and Computer Science, Vol.6, 2017, pp.21963-21967.
- [7] Hazra, A., Mandal, S., Gupta, A., Mukherjee, A., and Mukherjee, A., “Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review,” Advances in Computational Science and Technology, Vol.10, No.7, 2017, pp.2137-2159.
- [8] Iyer, A., Jeyalatha, S., and Sumbaly, R., “Diagnosis of Diabetes Using Classification Mining Techniques,” International Journal of Data Mining & Knowledge Management Process (IJDKP), Vol.5, No.1, 2015, pp.1-14.
- [9] Ding, Y., Tang, J., and Guo, F., “Identification of Drug-Side Effect Association via Semisupervised Model and Multiple Kernel Learning,” IEEE Journal of Biomedical and Health Informatics, Vol.23, No.6, 2019, pp.2619-2632.
- [10] Saad, A., Omar, Y., and Maghraby, F., “Predicting Drug Interaction With Adenosine Receptors Using Machine Learning and SMOTE Techniques,” IEEE Access, Vol.7, 2019, pp.146953-146963.
- [11] Salim, M., Park, L., and Park, J., “A Machine Learning based Scalable Blockchain architecture for a secure Healthcare system,” 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), 2022.
- [12] Lakhan, A., Mohammed, M., Nedoma, J.,

- Martinek, R., Tiwary, P., and Kumar, N., "DRLBTS: deep reinforcement learning-aware blockchain-based healthcare system," *Scientific Reports*, Vol.13, No.1, 2023, pp.1-15.
- [13] Sundas, A., Badotra, S., Bharany, S., Almogren, A., Tad-EIDin, E., and Rehman, A., "HealthGuard: An Intelligent Healthcare System Security Framework Based on Machine Learning," *Sustainability*, Vol.14, No.19, 2022, pp.1-16.
- [14] Coelho, K., Nogueira, M., Vieira, A., Silva, E., and Nacif, J., "A survey on federated learning for security and privacy in healthcare," *Computer Communications*, Vol.207, 2023, pp.113-127.
- [15] 김영식, "헬스케어 분야 머신러닝 기술 활용 및 동향," *보건산업브리프*, 제289권, 2019, pp.1-12.
- [16] 배재권, "스마트 헬스케어 산업의 인공지능 기술 적용에 관한 연구," *로고스경영연구*, 제19권, 제4호, 2021, pp.180-197.
- [17] 이새봄 · 송재민 · 박아름, "헬스케어산업에서의 인공지능 활용 동향," *한국콘텐츠학회논문지*, 제20권, 제5호, 2020, pp.448-456.
- [18] Shailaja, K., Seetharamulu, B., and Jabbar, M., "Machine Learning in Healthcare: A Review," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018.
- [19] Dhillon, A., and Singh, A., "Machine Learning in Healthcare Data Analysis: A Survey," *Journal of Biology and Today's World*, Vol.8, No.2, 2019, pp.1-10.
- [20] Ghosal, P., Das, D., and Das, I., "Extensive Survey on Cloud-based IoT-Healthcare and Security using Machine Learning," 2018 Fourth International Conference on Research in Computational Intelligence and Communication (ICRCICN), 2018.
- [21] 유재호 · 전의찬 · 김하나, "텍스트 분석을 활용한 기후변화 연구 동향 분석: 한국기후변화학회지를 중심으로," *Journal of Climate Change Research*, 제10권, 제3호, pp.161-172, 2019.
- [22] 황고은 · 황동열, "빅데이터 기술을 활용한 인문콘텐츠 분야의 의미연결망 분석," *인문콘텐츠*, 제43호, 2016, pp.229-255.
- [23] 황고은 · 문신정, "영상콘텐츠분야 정권별 빅데이터 분석 - 상위 중심성 값의 변화를 중심으로," *디지털콘텐츠학회논문지*, 제18권, 제5호, 2017, pp.911-921.
- [24] 한지윤 · 신영준, "인공지능교육 관련 연구 동향 분석: 키워드 네트워크 분석," *인공지능연구 논문지*, 제1권, 제2호, pp.20-33, 2020.
- [25] 주재홍 · 송지훈, "키워드 네트워크 분석을 활용한 지식은폐 연구동향 분석," *지식경영연구*, 제22권, 제1호, pp.217-242, 2021.
- [26] 류기진 · 남형식 · 조상호 · 류동근, "사회연결망 분석을 이용한 컨테이너 정기선 항로 패턴 분석에 관한 연구 : 부산항을 중심으로," *한국항해학만학회지*, 제42권, 제6호, 2018, pp.529-538.
- [27] 임동식, "미국은 당뇨병 나라?...인구 3분의 1이 시름," *전자신문*, <https://www.etnews.com/20170721000375>, 2017.07.22.
- [28] 한국보건산업진흥원, "2022 국가별 보건산업현황 보고서," <https://www.khidi.or.kr/board/view?linkId=48881113&menuId=MENU00085>, 2022.09.20.
- [29] 박혜경, "인공지능(AI) 헬스케어산업 현황 및 동향," *한국과학기술연구원 융합연구정책센터*, 2019.06.24.

■ 저자소개 ■



이택균
(Lee Taekkyeun)

2016년 9월-현재
아주대학교 다산학부대학 조교수
2014년 3월~2016년 8월
아주대학교 소프트웨어학과
강의교수
2010년 9월 State University of New York at
Buffalo, Computer Science and
Engineering(공학박사)
1998년 2월 고려대학교 전산학과(이학석사)
관심분야 : 데이터 마이닝, 인공지능
E-mail : taeklee@ajou.ac.kr

논문접수일 : 2023년 8월 1일
수정접수일 : 2023년 8월 21일
게재확정일 : 2023년 8월 25일