

헬스케어 분야에서의 빅데이터 활용 활성화 방안

문자화

단국대학교

Activation of Health Care Big Data

Ja-hwa Moon

Dankook University

E-mail : jahwada@gmail.com

요 약

데이터의 폭발적인 증가와 함께 데이터를 통한 새로운 가치와 인사이트 도출에 주력하는 '빅데이터 시대'가 도래했다. 데이터 분석 기술의 발전에 따라 진단 및 치료 및 예방 분야에서 데이터 분석 및 활용의 중요성이 확대되고 있으며, 헬스케어 분야에서는 빅 데이터 활용이 부각되고 있다.

또한 2020년 1월 데이터 3법(개인정보 보호법, 정보 통신망법, 신용정보법)이 통과되면서 가명 정보를 통한 다양한 빅데이터 활용이 가능해졌다.

그러나 다양한 정책과 규제, 일관되지 않은 데이터 품질, 전문 인력 부재 등으로 인해 헬스케어 분야의 빅데이터 활용은 여전히 어려움을 겪고 있다.

본 연구에서는 헬스케어 분야의 빅데이터 활용 현황을 살펴보고 헬스케어 빅데이터 활용 활성화를 위한 해결과제, 해외사례, 방안 및 기대 효과를 분석한다.

ABSTRACT

With the explosive increase in data, the 'big data era' has arrived, focusing on deriving new values and insights through data.

With the development of data analysis technology, the importance of data analysis and utilization in the field of diagnosis and treatment as well as prevention is expanding, while the use of big data is emerging in the healthcare field.

Moreover, as the three data-related laws (Personal Information Protection Act, Information and Communication Network Act, and Credit Information Act) were passed in January 2020, it became possible to use a wide range of big data through pseudonym information.

However, the use of healthcare big data is still struggling due to various policies and regulations, inconsistent data quality, and the absence of specialized personnel.

Therefore, in this study, examines the current state of use of big data in the healthcare field, and analyzes the challenges, overseas cases, plans, and expected effects for activation of healthcare big data.

키워드

Big Data, Healthcare, Data analysis, Medical data

I. 서 론

4차 산업혁명에서 빅데이터는 다양한 산업의 성장에 기여하며, 새로운 가치를 창출하는 촉매제로

주목받고 있다. IDC(2018)에 따르면, 글로벌 데이터 규모는 2018년 33ZB에서 2025년 175ZB로 연평균 61% 증가할 것이라고 전망했다[1]. 그 중 헬스케어는 유전체 정보와 헬스케어 기기 등 개인으로

부터 얻는 데이터가 폭발적으로 증가하면서 빅데이터가 만들어지고 있는 대표적인 분야다. IBM에 따르면 전세계적으로 16,000개 병원이 환자의 데이터를 수집하고 있으며, 환자 1인당 하루에 86,400개의 데이터가 생성되고 있다고 발표하였다[2]. 이렇듯 헬스케어 분야의 데이터가 폭발적으로 증가하고 있는 가운데, 최근 헬스케어 산업 패러다임이 진단·치료 중심에서 사전예방중심으로 변화하면서 빅데이터 분석 및 활용의 중요성이 확대되고 있다. 더욱이 2020년 1월 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법)이 통과되면서 헬스케어 시장에 빅데이터 활용이 가능해졌으며, '디지털 헬스케어'의 시대가 도래했다고 볼 수 있다. 그러나 'K-뉴딜' 등 범정부 차원의 데이터 개방과 활용 요구가 커짐에도 불구하고 각종 정책 및 규제, 일관되지 않은 데이터 품질, 전문 인력 부재 등으로 인해 헬스케어 빅데이터의 활용까지는 여전히 주춤한 상황이다.

본 연구에서는 헬스케어 분야의 빅데이터 활용 현황을 살펴보고, 헬스케어 빅데이터 활성화를 위한 해결 과제와 해외사례, 방안 및 기대효과를 분석한다.

II. 국내 헬스케어 빅데이터 현황 및 해결과제

2.1 헬스케어 분야에서 활용되는 빅데이터 종류
헬스케어 분야에서 활용되는 빅데이터는 매우 다양하며, 생성 주체에 따라 표1과 같이 정의되고 분류될 수 있다[3]

표 1. 헬스케어 빅데이터 종류

구분	종류
진료 데이터	전자 의무기록, 전자건강기록 정보, 처방정보, 입퇴원 기록, 의료영상자료 등 데이터
임상연구 데이터	의약품, 임상시험 데이터, 의료기기, 유전자 연구 데이터, 인체유래물 연구 데이터, 조사관찰 연구 데이터, 개인정보를 직·간접적으로 활용한 연구 데이터
공공기관 데이터	자격 및 보험료 관련 데이터, 진료내역, 건강검진 결과, 사망정보 등 공공기관에서 수집·보관·관리하는 데이터
기기기반 데이터	의료기기 및 환자 모니터링 장치 기반 데이터
오믹스(omics) 데이터	유전체, 전사체, 단백질체, 대사체, 후성유전체, 지질체 등 다양한 분자 수준에서 생성된 여러 데이터
라이프로그(Lifelog) 데이터	대체로 웨어러블, 홈 모니터링 장치, IoT, 모바일 앱 등에서 수집된 데이터로, 한 개인의 일상생활 활동에 관한 모든 데이터(체중, 심박수, 혈당, 몸무게 등)

앱·소셜 미디어 데이터	건강 포털, 의사 포털, SNS에서 수집된 다양한 헬스케어 관련 데이터
--------------	---

2.2 국내 헬스케어 빅데이터 현황

한국의 헬스케어 데이터 보유량과 IT인프라는 세계적으로 높은 수준으로 평가받는다. 양질의 공공데이터뿐 아니라 이를 효율적으로 활용할 수 있는 IT인프라 또한 높은 수준으로 구축되어 있기 때문이다. 특히 정부 주도적 빅데이터 양산을 통해 건강보험공단과 건강보험시사평가원이 보유한 빅데이터는 각각 3조 4,000억 건과 3조 건에 달하며, 전자의무 기록 보급률 또한 92%로 세계 1위 수준을 기록하고 있다. 의료 선진국인 유럽의 전자기록 보급률이 84%라는 것을 비교해 볼 때 이는 매우 높은 수치다. 이런 환경에도 불구하고 데이터 활용에 제약이 따른다는 지적에 따라 정부는 '가명 정보' 활용 등을 내용으로 하는 데이터 3법을 개정·시행 하고 보건 의료 데이터 가이드라인도 발표하는 등 데이터 활용을 위한 법·제도 기반을 마련했으며, 4차산업혁명위원회 내 헬스케어특별위원회 역시 헬스케어 빅데이터 활용 방안을 마련했다.

축적된 헬스케어 빅데이터는 다양한 주체와 목적에 의해 활용 될 수 있으며, 활용 가치 또한 높다. 보건 의료 측면에서 빅데이터를 기반으로 질병과 사회병리현상에 대한 관리 및 예측을 통해 국민 의료 서비스 질 향상과 비용절감 효과를 가져올 수 있으며, 보건정책에 대한 평가와 모니터링을 통해 개선점을 도출할 수 있다. 의료산업 측면에서 제약사들은 신약 개발에 소요되는 임상 기간을 단축시키거나, 맞춤형 약제 서비스를 제공하는 데 활용 할 수 있으며 보험사는 신규 보험 상품 및 헬스케어 서비스를 개발을 통해 고객 니즈에 대한 선제적 대응을 펼칠 수 있다.

국내 헬스케어 빅데이터 활용현황은 표 2와 같다.

표 2. 헬스케어 활용 현황

회사명	활용내용
헬스커넥트(Health-Connect)	Health On 앱을 통해 수집된 건강 기록 데이터, 병원 임상 데이터, 계열사를 통해 획득한 계층 데이터 등 빅데이터를 기반으로 만성 질환 관리, 질병 예측 등 개인 맞춤형 정밀 의료 서비스 제공
비바이노베이션(Viva-Innovation)	건강보험시사평가원에서 제공하는 기업 맞춤형 의료 빅데이터를 기반으로 문진서비스 제공 - 착한의사 어플리케이션을 통해 AI가 증상을 분석하고 증상 정보를 의료인에게 제공하는 'AI 증상체크 서비스' 제공 - 전국 7 만개 병원의 진료 시간, 병원별 평가 등급, 병원비 정보 등을 제공
M20	EMS 장비로부터 환자가 생성한 건강 데이터를

	수집하여 환자에게 적합한 근육 강화, 체지방 분해 등 최적화된 운동 프로그램을 제공한다
Dr. Noah Biotech	약물 빅데이터 분석을 통한 희귀 질환 치료제(신약) 개발
KB손해보험	빅데이터 전문 기업과의 협업을 통해 고객 맞춤형 보험 상품 개발

2.3 헬스케어 빅데이터 해결과제

한국은 현재 공공, 민간 모두 방대한 헬스케어 빅데이터를 보유하고 있고, 그 경제적 가치는 최대 2조 원으로 추정되지만, 일부 코호트 개방 및 공익적 연구에만 아주 제한적으로 활용되고 있는 실정이다[4]. 그동안 개인정보보호법 등 각종 규제로 인해 헬스케어 빅데이터를 공익적 목적으로만 활용 가능하도록 제한되었고, 사실상 민간기업 또는 산업으로의 활용이 봉쇄되었다. 그러는 사이 헬스케어 분야의 혁신적인 발전을 이루고 있는 미국 등 선진국과 한국간의 격차는 점점 벌어졌다. 최근 데이터 3법 개정과 의료보건산업 규제 완화, 보건 의료 데이터 활용 가이드라인 마련 등으로 가치창출의 발판이 마련됐으나, 여전히 각종 제도적 문제와 일관되지 않은 데이터 품질, 전문 인력 부족 등 해결해야 할 과제가 산적해있다.

III. 해외사례

한국보다 먼저 헬스케어 빅데이터에 개방적 태도를 보여왔던 미국·영국·중국·일본의 빅데이터 정책과 활용사례는 표 3과 같다.

표 3. 해외 헬스케어 빅데이터 정책 및 활용 사례

국가	정책	활용사례
미국	2015년 1월 '정밀 의료 이니셔티브'를 발표하고 국가 우선 과제로 헬스케어를 선정함 - 'All of US' 프로젝트를 통해 10년 동안 100만 명의 데이터를 기부하여 유전 정보, 생체 정보, 전자 의료 기록을 수집함	- '메인 헬스인포넷'은 미국 국민의 전자건강기록(EHR)을 분석하여 예상되는 고위험군 사람에게 위급상황의 예방법 및 응급실 방문의 필요성 등의 정보를 제공한다 - 'Cancer Moonshot Program'은 암 환자들의 치료, 회복에 관한 데이터를 활용하여 암치료에 성공률이 가장 높은 치료법을 추출하는 연구를 진행함
영국	'국가보건서비스(NHS)' 주도의 국가 정책 수립 - 여러 의료 서비스에 분산된 데이터는 보건복지 정보센터	- Healthera는 NHS에서 인증 및 협업하는 스마트폰 앱이며 환자는 병원의 방문없이 기존 처방전의 재발급을 신청하면 지정 약국에서 약을 받을 수 있는 서비스를 제공한다

	(HSCIC)로 전송되어 데이터베이스화됨 - NHS Digital을 통해 개인 의료 정보 빅데이터를 기반으로 의료 서비스를 개발 및 제공한다	- Google Verily는 만성질환 데이터와 사회 전반에 걸친 데이터를 분석하여 만성질환 발병 가능성이 높은 사람을 사전에 예측하고 원격 의료를 통해 건강관리 서비스를 제공한다
일본	2015년 '차세대의료 기반법'을 개정하고 익명가공정보 개념을 도입하여 개인정보를 활용할 수 있는 근거 마련. 또한 개인데이터저장소(PDS)와 데이터 거래시장의 개념에 데이터신탁 개념을 더하여 정보은행을 도입함	- '미쓰이시미토모은행'은 병원과 약국에서 수집된 의료데이터를 디지털화하고 축적된 데이터를 필요로 하는 자에게 제공한다. 이를 통해 정보제공자(환자)는 유·무형의 이익을 돌려받을 수 있음 - '후지쓰'는 클라우드 기반 모니터링 서비스가 허용됨에 따라 의료기관과 협업하여 원격진료서비스 제공 및 의료데이터를 저장 및 활용함
중국	2015년 '빅데이터 발전촉진 행동강령'을 시작으로 의료 빅데이터 발전 및 활용 정책을 발표함 - 의료고령화 사회에 대한 대비와 낮은 의료 서비스 접근성 등에 대한 해결책으로 모바일 헬스케어 산업을 적극적으로 지원	- '마오지엔강'은 인공지능과 빅데이터 기반 모바일 헬스케어 플랫폼을 개발했으며 가입자의 개인 건강상태를 수집한 생활데이터를 분석하여 건강관리 통합 솔루션을 제공한다 - '알리바바'는 모바일 의료시스템인 '미래병원'을 통해 환자가 진료를 받기 위해 병원에 방문할 경우에 발생하는 의료서비스 전반에 걸친 필요정보를 제공한다

IV. 헬스케어 빅데이터 활성화 방안 및 기대효과

4.1 가명화 정보 규제 완화

헬스케어 빅데이터는 정보의 민감성과 복잡성으로 활용이 쉽지 않은 영역이다. '데이터3법' 시행에 따라, 개인정보를 가명처리하여 활용할 수 있는 기반이 마련 됐지만, 가명화 정보라 할지라도 결합 활용하고자 할 경우 별도의 활용심의회와 결합 적정성 검토 등을 거쳐야 하는 번거로움이 있으며 지나친 비식별화 처리로 인해 데이터 활용에 어려움이 존재한다. 개인정보·민감정보에 대한 엄격한 관리의 분명 필요하다. 그러나 가명처리를 통해 개별적으로 식별이 불가능한 데이터까지 규제하는 것은 완화되어야 한다. 미국의 경우, 의료정보보호법(HIPPA)을 통해 식별자를 제거한 데이터를 전면적 규율 면제로 간주하고 정책 평가, 생명 과학 연

구, 기타 다른 시도를 위한 2차 활용을 지원한다. 다른 데이터와의 결합으로 데이터 주체를 식별할 가능성이 없는 경우, 더 이상 개인의 민감정보로 간주하지 않는 것이다[5]. 이처럼 개인정보 뿐 아니라 비식별 처리된 정보의 활용을 일정 부분 보장하는 방식으로 규제 완화가 필요하다.

또한 분야마다 적용할 수 있는 개인정보 보호수준은 다를 수 있다. 따라서 비식별 처리된 데이터가 보건의료 분야에서의 활용가치를 발휘할 수 있도록 연구기관·기업의 활용 목적과 상황에 맞는 가명처리 기준 수립이 필요하다.

4.2 일관된 데이터 품질 확보

헬스케어 빅데이터를 효과적으로 활용하기 위해서는 데이터 질적수준 향상이 필수적으로 선행되어야 한다. 그러나 아직까지 국내 뿐 아니라 전세계적으로도 품질 관련한 명확한 기준이 없어, 기관별 품질 상이를 초래하고 있다. 공공기관·의료기관·개인·기업 등 다양한 주체에서 수집되고 있는 헬스케어 데이터의 개방과 활용을 위해 일관된 양질의 데이터를 생성할 수 있도록 정부 주도의 품질 검증 기준 및 인증 제도 마련이 필요하다. 이를 바탕으로 기관 간의 데이터 연계를 추진함으로써 사용자는 가치있는 정보로의 전환이 가능할 것이며, 데이터 활용성을 높일 수 있다.

4.3 전문인력 양성

스위스 국제경영개발대학원(2020)이 발표한 디지털 경제 순위 조사에 따르면 한국은 '빅데이터 사용 및 활용 능력'에서 15위를 차지했다[6]. 40위를 차지한 2019년도와 비교해봤을 때 괄목한만한 성과다.

그러나 헬스케어 산업의 시장 우위를 선점하기 위해서는 이 결과에 만족해서는 안된다. 미래 경쟁력을 확보하기 위해 이미 오래전부터 선진국들은 국가의 역량을 쏟아부어 빅데이터의 발전과 인력 양성에 매진해왔다. 향후 헬스케어 산업은 빅데이터를 제대로 활용할 수 있는 인재 양성에 달려 있다고 해도 과언이 아니다. 정부는 빅데이터를 중심으로 한 4차 산업시대를 이끌 인재 양성 문제를 중요한 국정과제로 선정해 이를 집중적으로 다루고 추진해야 한다.

요하다. 미국 등 선진국들은 이미 헬스케어 빅데이터 활용을 위한 다양한 정책을 운영 중에 있으며, 규제완화를 통해 많은 민간기업들의 산업 참여를 이끌어내고 있다.

헬스케어 데이터 활용 활성화를 위해 현재 정부가 추진하는 정책과 규제 완화가 데이터 연구 및 제품 개발에 실질적으로 활용될 수 있도록 지속적인 제도의 개선이 필요하다. 또한 일관된 데이터 품질을 확보하기 위한 기준 마련, 체계적인 전문인력 양성이 필요하다.

Acknowledgement

Thank you parents, sister, and Joo-wan Lee.

References

- [1] IDC, "Data Age 2025," November 2018.
- [2] IBM, "Bigdata in healthcare," 2014.
- [3] So-young Yoo "Healthcare Big Data Dilemma and Solution," *Kakao AI Report*, vol. 11, 24p, March 2018.
- [4] Bokuenews "Establishment of national strategy for utilization of health care data in earnest," March 2021.
- [5] ZDNet Korea, "Finding solutions for using health care data from 'users'," November 2020.
- [6] Ministry of Science and Technology Information and Communication, "2020 Swiss Institute for International Management Digital Competitiveness Evaluation Results Announced," October 2020.

V. 결 론

빠른 속도로 쌓여가는 데이터를 어떻게 활용하느냐가 미래 헬스케어 분야의 성패를 결정한다. 아직 헬스케어 분야에서 빅데이터 활용은 초기단계에 지나지 않으며, 많은 투자와 노력이 필요한 시기이다. 다만 헬스케어는 개인정보·민감정보를 다수 포함하기 때문에 민간주도로 활용 활성화를 도모하기 어려운 분야이므로 정부의 역할은 매우 중