

의료 빅데이터-AI-융합 기술의 발전 패러다임과 활성화 정책 이슈

박세환 기술법인 엔팜 전문위원

1. 머리말

의료 빅데이터가 AI와 융합되면서 의료 및 임상 데이터베이스에서 제공되는 정보들이 신약 개발, 질병 치료 등 바이오 의학에 크게 기여하고 있다. 이에 따라, 마이크로소프트(Microsoft), 아마존(Amazon), 애플(Apple), 구글(Google) 등 의료 빅데이터 기술시장의 선두주자인 글로벌 ICT 메이저 기업들은 의료 빅데이터-AI-융합 기술 개발에 주력하고 있다. 유전체학 기술과 보건 정보에 의해 생성된 의료 빅데이터를 효과적으로 분석하고 공유하기 위해서다.

이를 효과적으로 구현하기 위해선 의료 빅데이터에 대한 구조적 분석 기술이 필수적으로 요구된다. 앞서 소개한 기업들은 바이오의학, 디지털 치료제, 의료-AI 융합 기술 등을 개발해 의료정보 과학(맞춤형 의학 프로그램)에 적용할 수 있는 기반을 마련하고 있으며, 만성 질환 및 난치성 질환 치료에 직접적 효과를 거두고 있다.

국내에선 질병관리청, 건강보험심사평가원, 국민건강보험공단 등 건강관리기관이 보유한 의료 빅데이터가 기업이나 의료기관 등의 임상 데이터와 결합돼 개인 맞춤형 의료 서비스 및 신약 개발에 활용되고 있다. 특히, 돌봄 및 진단 AI 분야에선 음성 인식, 자연어 처리 등의 기술을 적용해 의료 이미지 분석, 의학 문헌 학습 등 진단예측 분야의 혁신을 견인하고 있다[1].

가명 정보기술¹⁾을 기반으로 한 의료 빅데이터 활용은 의료 서비스 패러다임 변화를 통해 다양한 개인 맞춤형 의료 서비스 산업을 활성화해 사회적 비용을 감소시키고 있다. 특히, 가명 정보기술은 개인정보보호 규제로 인해 데이터를 자유롭게 활용하기 어려운 경우 이를 해결해 데이터를 안전하게 활용할 수 있는 기반을 제공하면서, 데이터 경제 활성화에 필수 요인으로 부각되고 있다[2]. 의료 빅데이터-AI-융합 기술을 이용한 선진 의료 서비스가 효과적으로 정착되기 위해선 민감 의료 정보의 안전한 관리와 함께, 데이터 활용의 윤리적 책임의식에 대한 사회적 인식이 제고돼야 한다[3]. 2017년 11월 1일 기준, 한국은 65세 이상 노인인구가 712만 명으로 전체의 14%를 초과하는 고령사회가 됐다²⁾. 이에 4차 산업혁명 시대 바이오(의료)-ICT 융합 기술을 기반으로 웰빙 마인드를 지향하는 개인 맞춤형 의료 서비스가 절실하다. 이를 효과적으로 구현하기 위해선 최적화된 의료 서비스를 지향하는 의료 빅데이터-AI-융합 기술(의료 빅데이터 구축 및 활용 기술, 의료 빅데이터 분석 기술, 의료 빅데이터-AI-융합 기술 등)이 필수적으로 요구된다.

이번 원고에선 4차 산업혁명 시대 의료 산업 육성의 필수요소이자 의료 현장에서 새로운 솔루션

1) 개인정보 일부를 삭제하거나 대체해 추가 정보를 결합하지 않고서는 특정 개인을 알아볼 수 없도록 설계된 데이터를 의미함. 데이터 분석과 활용의 유용성을 유지하면서 개인정보보호를 강화할 수 있는 큰 장점이 있음

2) UN에선 노인인구 비중이 7% 이상이면 고령화 사회, 14% 이상이면 고령 사회, 20% 이상이면 초고령 사회로 구분.

으로 인식되는 의료 빅데이터-AI-융합 기술 발전 패러다임에 대해 설명한다. 이는 의료 빅데이터 구축 기술, 의과학(의료 정보학·바이오 정보학·바이오 통계 등) 분야 의료 빅데이터 활용 기술, 비정량화 의료 빅데이터를 안전하게 분석하는 데 필요한 의료 빅데이터 분석 기술을 모두 포함한다. 아울러, 의료 빅데이터-AI-융합 기술 활성화를 위한 데이터 3법 이슈에 대해 설명하고, TRM(기술 로드맵, Technology Road Map)을 제시한다.

2. 의료 빅데이터-AI-융합 기술 발전 패러다임

2.1 의료 빅데이터 구축 및 활용 기술

2.1.1 의료 빅데이터 구축 기술

AI, 클라우드 컴퓨팅 등을 접목한 의료 데이터 수집은 4차 산업혁명 시대 의료 산업 육성의 필수 요소다. 데이터 3법 개정은 특정 개인의 가명 정보를 이용한 의료 정보 공유를 가능케 했다. 이로써 기존 의료 데이터 활용 범위가 확대되고, 빅데이터 기반 제품 및 서비스 질 개선, 신규 비즈니스 모델 등장이 기대된다. 나아가, 기업, 의료기관 등이 보유한 임상 데이터를 결합한다면 임상 의사결정 지원, 개인 맞춤형 의료 서비스(하드웨어·상품·플랫폼·솔루션 등) 등 매우 다양한 서비스를 제공할 수 있을 전망이다.

이에 따라, 해외 주요국과 규제기관들은 의료 빅데이터 활용을 증가시켜 가고 있으며, 국내에서도 의료 빅데이터 활용을 강화해 100만 명 규모의 바이오 빅데이터가 구축되고 있다. 의료 빅데이터 활용강화 분야로는 개인 맞춤형 의료 서비스, 의료 서비스 질과 안전성 향상, 임상 의사결정 지원 등이 있다. 2018~2020년 말까지 국내 주요 의료 빅데이터 구축사업을 요약하면 다음과 같다[3].

- 2020년 12월까지 EMR(전자 의무기록, Electronic Medical Record) 표준화, 네트워크 재구축 등 바이오 헬스 빅데이터 구축 사업을 39개 의료기관과 7개 기업에 추진
- 의료 데이터를 국제표준 CDM(공통데이터 모델, Common Discovery Manager)³⁾로 표준화하고 데이터 분석 소프트웨어를 개발해 플랫폼을 구축
- CDM 데이터⁴⁾를 개방해 코로나19 관련 후유증, 이상 반응 등의 연구를 지원 (건강보험심사평가원)
- 100만 명 규모의 바이오 빅데이터를 구축, 2029년까지 데이터 중심 병원 5곳을 대상으로 5대 의료 빅데이터 플랫폼(공공기관·AI 신약개발·병원 임상·피부·유전체)을 구축

2.1.2 의료 빅데이터 활용 기술

의료정보학·바이오 정보학·바이오 통계 등의 의과학 분야에서 의료 빅데이터 분석 기술이 AI 기술과 융합돼 만성 질환, 난치성 질환, 유전체학(Genomics)등에 적용되고 있다. 덕분에 인간 기대수명도

3) 자원별 의미, 길이 및 값의 제한사항을 식별해 고객의 비즈니스 자원과 구성요소 간 데이터 교환을 위해 사용되는 정의 언어로서 DMTF(Distributed Management Task Force) CIM(Common Information Model) 표준, 비즈니스 프로세스 표준, 도메인 특정 표준을 포함하고 있음

4) 제공하는 CDM 데이터는 2021년 한 해 동안 의료 서비스를 이용한 전체 환자에서 총별 20%(약 1천만 명) 표본 추출, 대상 환자의 2018년 1월~2022년 4월까지의 청구 데이터를 CDM으로 변환한 데이터들임

획기적으로 늘어나고 있는데, 한 사람의 기대수명주기 동안 생산되는 의료 데이터는 1,106.4TB에 달한다⁵⁾[4].

이러한 의료 빅데이터-AI-융합 기술은 특정 만성질환 및 난치성 질환자로부터 각종 질병 예방의 학 등을 탐색하고 분석하는 능력을 제공하고 있다. 이를 통해, 보건의료와 생명과학 산업 활성화에 크게 기여하고 있다. 의료 빅데이터-AI-융합 기술 관련 전후방 글로벌 기술시장을 주도하고 있는 메이저 기업으로는 애플스트리(Appistry), 베이징 게놈연구소(Beijing Genome Institute), CLC 바이오(CLC Bio), 컨텍스트 매터스(Context Matters), DNA넥서스(DNAnexus), 게놈 인터내셔널 코퍼레이션(Genome International Corporation), GNS 헬스케어(GNS Healthcare), 넥스트바이오(NextBio), 패스파인더(Pathfinder) 등이 있다. 이들 기업은 임상 정보 관련 -omics 데이터의 수요 분석, 소유자 알고리즘을 이용한 빅데이터 분석·고속화·고품질화, 웹 기반 앱을 이용한 제약·바이오 기술회사 지원 등을 진행 중이다. 이를 통해, 보건의료 및 바이오 산업 분야에서 맞춤형 소프트웨어와 앱을 개발해 기술혁신을 견인하고 있다[5].

2.2. 의료 빅데이터 분석 기술

정량화되지 않은 대부분의 의료 빅데이터를 기존 DBMS(Data Base Management System)로 분석하는 데는 어려움이 있다. 이에 연구자가 안전한 방식으로 대량의 개인의료 데이터를 분석하기 위해선, 데이터 저장장치, 대용량 서버, 데이터 처리 등의 기능을 갖춘 최적의 의료 빅데이터 분석기법이 요구된다.

이를 위해 개발된 AI 기반 컴퓨터 도구들이 현재 바이오의학, 생명과학 등 다양한 분야⁶⁾에서 인터넷 기반 무작위 수집 데이터 지식을 가공하고 있다. 대표적 사례로 1)공간적인 지도를 이용해 질병을 추적하는 빅데이터 분석기법인 GT(Google Trends), 2)대용량 의료 정보를 분석해 진단을 결정하는 GT 질병 추적 알고리즘 등이 있다. 의료 빅데이터 분석 기술 시장에서 주목받고 있는 GT의 특징은 다음과 같다[6].

- 바이오의학 분야 빅데이터 분석을 통해 질병 유행지역에서의 질병 추적과 모니터링 기능
- 구글 탐색 요구의 급격한 증가추세를 빅데이터 분석기법을 통해 검증(특정 지역의 응급환자 수 증가 전 플루(Flu) 증세와 플루 치료 등)

2.3 의료 빅데이터-AI-융합 기술

보건의료 분야에서 데이터 기반 AI 활용은 새로운 솔루션으로 인식되고 있다. 이는 효과적인 진단 및 치료, 의료진 업무경감, 의료 서비스 효율과 질 향상, 의료 현장 미충족 수요 해결을 가능케 한다.

최근엔 생성형 AI의 확산 등으로 디지털 전환(Digital Transformation)이 가속되면서, 안전성 및 신뢰성 측면의 장애요인(보건의료 데이터 활용의 난제 등)이 대두됐다[7].

5) - 의료데이터: 인적 정보/건강보험 정보/진료 정보/진료관리요약 정보/사망기록 정보 등
- 1,106.4TB의 외생 데이터: 6TB의 유전체 데이터, 0.4TB의 임상 데이터 등 행태적/사회경제적/환경적 요소로 구성
6) AI의 전후방기술인 NLP(자연언어처리, Natural-Langue Processing), 패턴인식 및 기계학습(Machine Learning) 등

- 진단오류 시 법적 책임 문제
- 민감 정보 처리 문제
- 데이터 권리주체의 불명확성
- 데이터 활용에 관한 윤리 문제
- 학습 데이터 품질 문제
- 시스템과 SW 간 데이터 연계 및 호환 문제 등

이로 인해 데이터 활용성이 저하되고 편향성이 생기는 문제가 발생하고 있으며, 관련 장애요인 해소에 주력할 필요가 있다[7].

특히, 보건의료 데이터는 다양한 정형 및 비정형데이터(텍스트, 사진, 동영상 등)가 혼재돼 있어 고도의 데이터 분석 및 처리 기술이 요구된다. 아울러, 보건의료 데이터를 효과적으로 활용하기 위해선 법·제도, 연구 윤리를 이해하고 가명 처리, 분석 등의 기술력을 갖춘 전문인력이 필수적이다. 특히 가명 정보 활용 시엔 법, 규정, 표준 등을 준수하는 가이드라인이 필요하다.

3. 의료 빅데이터-AI-융합 기술 활성화 정책 이슈

3.1 데이터 3법 이슈

데이터 3법⁷⁾이 개정되면서 산업적 목적, 과학적 연구와 통계작성, 공익적 기록보존 등 특정 목적으로 각 건강관리기관에 분산된 개인의료 빅데이터를 연계-통합한 후, 비식별화(De-Identification)를 통해 민간 연구자에게 제공하는 것이 가능해졌다. 이를 통해, 기업과 기관이 안전하게 가명 정보를 활용할 수 있는 방법과 기준 등을 수립할 수 있을 전망이다. 데이터 3법에선 '정보통신망 이용촉진 및 정보보호 등에 관한 법률'의 우선 적용 규정을 삭제함으로써 '개인정보보호법'과의 중복 규제 문제를 해결했다. 다만, 사람을 대상으로 하는 연구에는 '생명윤리 및 안전에 관한 법률⁸⁾'이 우선 적용되기 때문에 이 법률에서 정하지 않은 경우에 대해선 '개인정보보호법'을 적용받게 된다. 이는 '개인정보보호법'에서 규정하고 있는 가명 정보처리 특례 규정과는 무관하다고 볼 수 있다. 이에 개인의 민감성 의료정보를 가명 처리하기 위해 당사자 서면동의가 필수사항인 '생명윤리 및 안전에 관한 법률'을 보다 유연하게 개정할 필요성이 대두되고 있다.

3.2 의료 빅데이터-AI-융합 기술 TRM

의료-ICT 융합 기술을 효과적으로 구현하기 위해선 앞서 설명한 최적 의료 서비스 지향 3대 의료 빅데이터-AI-융합 기술 핵심기술(의료 빅데이터 구축 및 활용 기술, 의료 빅데이터 분석 기술, 의료 빅데이터-AI-융합 기술), 그리고 관련 세부 기술 분야에 대한 R&BD가 필수적으로 요구된다. 이를 위해, 2016년부터 추진돼온 스마트 헬스케어 TRM에 따라 핵심기술을 개발하는 데 주력할 필요가 있다[3][7].

7) 데이터의 이용 활성화를 위한 개인정보보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률(정보통신망법), 신용정보의 이용 및 보호에 관한 법률(신용정보법)등 3가지 법률의 통칭

8) 생명윤리 및 안전에 관한 법률(생명윤리법): 인간과 인체유래물 등을 연구하거나, 배아나 유전자 등을 취급할 때 생명윤리 및 안전을 확보하고 국민의 건강과 삶의 질 향상에 이바지함을 목적으로 마련된 법률(2019.04.23 일부개정, 2019.10.24 시행)

- 2016~2018년: 스마트 헬스케어(SHC) 서비스 시범사업 추진, SHC 플랫폼과 연계한 서비스 실증사업 추진, SHC 비즈니스 모델 발굴
- 2019~2020년: SHC 신시장 창출, SHC 산업 가치 확산
- 2020~2024년: 의료 빅데이터-AI-융합 기술 패러다임 전개
- 2025년 이후: 의료 빅데이터-AI-융합 기술 산업의 가치 확산

이러한 TRM 추진을 통해 바이오-ICT 융합 글로벌 기술시장 선점을 위한 전략을 수립하고, 방향성에 적합한 기술개발이 이뤄져야 한다. 특히, 1) 맞춤형 건강진단 및 질병 예방을 위한 웨어러블 기기 플랫폼 설계기술 개발, 2)웨어러블 기기 플랫폼과 스마트폰 연동 기술 개발, 3)정밀 의료 서비스를 위한 ICT 활용 유전체분석 기술 개발 등에 주력할 필요가 있다.

4. 맺음말

한국은 데이터 3법 개정을 바탕으로 가명 정보 개념을 도입해 개인정보 판단기준을 명확화하고, 개인정보 처리자의 책임을 강화했다. 특히, 개인정보가 포함된 민감성 의료 빅데이터를 가공해 특정인을 알아볼 수 없도록 함으로써, 의료개인정보를 강화했다. 아울러, 이를 미래 의료 서비스에 활용할 수 있는 법적근거도 마련했으며, 이를 통해 디지털 의료 서비스 산업 발전이 크게 성장할 전망이다. 특히, 건강관리기관 및 의료기관 등에서 다양한 목적으로 폭넓게 개인정보를 활용할 수 있게 됨으로써, 의료 빅데이터 관련 산업의 동반 성장을 견인하고 있다.

관련 기술의 효과적인 정착을 위해, 산업 현장 및 연구기관 등이 민감성 의료정보 활용에 따른 리스크를 최소화하는 데 주력할 필요가 있다[5]. 정보 주체의 사전 동의 없이 처리가 가능한 가명 정보처리는 의료 빅데이터 플랫폼 사업의 중요한 전환요소다. 개별추가 서면동의 없이 바이오-의료 정보 중심으로 연계 및 통합이 가능하기 때문에, 단기간 내 대규모 바이오 빅데이터를 구축할 수 있기 때문이다. 이에 데이터 3법 개정안의 해석 및 적용을 통해 규제완화 정책과 개인정보보호 정책 간 대립을 방지할 수 있는 방안이 필요하다. 이는 특정 개인의 유전자분석 정보나 임상 정보 등 민감성 정보의 수집·연계·공유 방법에 대한 가이드라인을 마련할 필요가 있음을 시사하고 있다.

의료 빅데이터(의료 정보)는 데이터 저장-이동-확보 비용보다, 그 전 단계인 생성-획득 비용이 더 적게 드는 특징이 있기에 매우 엄격한 공급망(Supply Chain, 생성-획득-저장-이동-확보-이용-폐기)이 요구된다. 아직은 의료 빅데이터 저장-이동-확보와 관련된 저렴하고 안전한 최적의 해결책은 제시되지 않은 상황이다. 이는 미래 바이오의학을 포함한 의료 빅데이터-AI-융합 연구가 상품·솔루션 등의 제품화까지 이어지기엔 많은 과제가 남아 있다는 점을 시사한다. 또한, 의료 빅데이터는 하드디스크를 통해 다른 곳으로 이동시킬 수 있어 이동 과정에서 데이터가 손실되지 않도록 데이터압축 소프트웨어 등이 필수적으로 요구된다.

최근엔 의료 빅데이터를 다 른 위치로 옮기면서 동시에 데이터 처리가 가능한 클라우드 컴퓨팅 기반 해결 방법이 개발됐다. 이는 데이터 이전과 잠재(Latency) 과정을 실행하는 도구로 주목받고 있으며, 저장 공간이 확보되지 않은 곳에서 생성되는 의료 빅데이터를 임의의 장소로 안전하게

이동할 필요가 있음을 시사하고 있다. 개인 의료 데이터의 보안 문제를 해결하는 것도 매우 중요한 과제다. 이를 해결하기 위해선 금융보안에서 이용하는 암호화 알고리즘(Encryption Algorithm)을 갖춘 보안 시스템이 요구된다. 아울러, 연구 참가자나 환자가 공개적으로 생성 데이터를 공유할 수 있는 역동적인 클라우드 컴퓨팅 기술을 적용해야 함을 시사한다.

의료 정보학·바이오 정보학·바이오 통계 등 의료-ICT 융합 산업 분야에서 의료 빅데이터-AI-융합 기술은 차세대 의료 서비스를 크게 향상시키며 맞춤형 의학 프로그램의 발전을 이끌고 있다. 이를 기반으로, 임상 데이터 및 다양한 -omics 데이터 관리, 개인의료정보 보안 등에 대한 통합관리 시스템이 구축되고 있다. 결과적으로, 만성질환 및 난치성 질환자 관리의 질을 향상하고, 나아가 의료비 절감이 기대된다.

이를 위해, 임상 및 유전 데이터 등 의료 빅데이터 공급망에 있어 개인 의료정보보안 등의 문제를 해결할 수 있는 법·제도적 기반이 구축돼야 한다. 아울러, 전 세계적으로 확대되고 있는 데이터 경제(Data Economy)로의 전환에 따른, 의료 빅데이터 분석 및 활용에 대한 법·제도적 기반도 필요하다.

본 연구는 국가과학기술인력개발원(KIRD)의 「2025년 과학기술인 학습조직(러닝랩)_학습주제: 국내 바이오산업의 디지털 전환을 위한 연구정책전환」 지원 결과임

[참고문헌]

[1] 의료AI와 바이오AI가 빚어내는 개인 맞춤형 의료(혁신가이드, 2023.11.27.)

<https://brunch.co.kr/@botongmarketer/531>

[2] 가명정보 기술: 데이터 활용과 개인정보 보호의 균형점(A3Blog, 2025.4.3.)

<https://a3soft.tistory.com/entry/pseudonymization>>

[3] 박세환, "데이터 3법을 계기로 본 의료 빅데이터와 디지털 의료 서비스", SecuN저널 162호, (주)테크월드, 2022.07. <https://www.epnc.co.kr/news/articleView.html?idxno=303048>

[4] 한국의 평균수명과 연령별 기대수명 분석(티스토리© Kakao Corp., 2025.2.18.)

<https://infouser.tistory.com/entry/한국의-평균수명과-연령별-기대수명-분석>>

[5] 한국보건산업진흥원 의료기기산업 종합정보시스템 자료종합.

<https://www.khidi.or.kr/board/view?linkId=48874427&menuId=MENU01499&schStartDate=&schEndDate=&categoryId=>

[6] Dugas, A.F, "Influenza forecasting with Google Flu trends", PLoS ONE 8, pe56176.

[7] 우혜경, "AI 시대, 보건의료 데이터를 이해하고 활용하기 위한 다섯 가지 핵심역량", 전자신문, 2024.12.10. <https://www.etnews.com/20241209000104>

※ 출처: TTA 저널 제219호