

Healthcare에서 빅데이터의 활용

정성원

데이터솔루션 마케팅총괄이사

Sungwon Jung

Marketing Director, Data Solution Inc

초 록

최근 디지털세상이 열리면서, 헬스케어 분야에도 데이터가 폭발적으로 증가하고 있다. 이와 같은 헬스케어 빅데이터를 잘 활용하면 보건의료분야에 투여되는 막대한 비용을 절감할 수 있을 것으로 전망된다. 특히 헬스케어 빅데이터와 머신러닝이 결합된 다양한 스마트 의료기기의 출현과 스마트폰과 웨어러블 디바이스의 보급확대에 따른 최적의 모바일 헬스 환경이 조성되면서, 환자 개인별로 특화된 질병진단 및 치료서비스는 물론이고 만성질환 관리 서비스 및 질병예방 서비스 등으로 의료서비스의 혁신을 이룰 수 있다.

주제어: 빅데이터, 머신러닝, 딥러닝, 스마트 의료기기, 모바일헬스

서 론

헬스케어 데이터가 폭발적으로 증가하고 있다. IBM에 의하면 세계 16,000개 병원이 환자 데이터를 수집하고 있다고 하며, 80%의 헬스케어 데이터는 텍스트나 이미지, 영상과 같은 비구조화 데이터라고 한다. 전 세계 490만명의 환자가 원격모니터링 디바이스를 사용하고 있으며, 원격모니터링 디바이스를 사용하는 환자가 최근 5년 년평균 18%의 증가율을 보이고 있다고 한다. 개별 환자모니터링장비는 평균 초당 1,000개의 수치를 측정하고 있으며, 이는 환자 1인당 하루에 86,400개 수치가 생성되는 것을 뜻한다. 이렇게 생성된 헬스케어 빅데이터의 크기는 미국 국회도서관에 의하면 약 370 TB에 달한다고 한다.

세계적인 비즈니스컨설팅 전문업체인 McKinsey는 2013년 연구보고서를 통해 미국 보건 의료부문이 빅데이터 활용의 용이성이 높고 경제에서 차지하는 비중이 커서 빅데이터 기대효과가 큰 것으로 전망하였으며, 빅데이터를 통해서 미국 보건의료부문에서만 연간 최대 \$190 billion의 비용절감을 실현시킬 수 있다고 내다봤다.

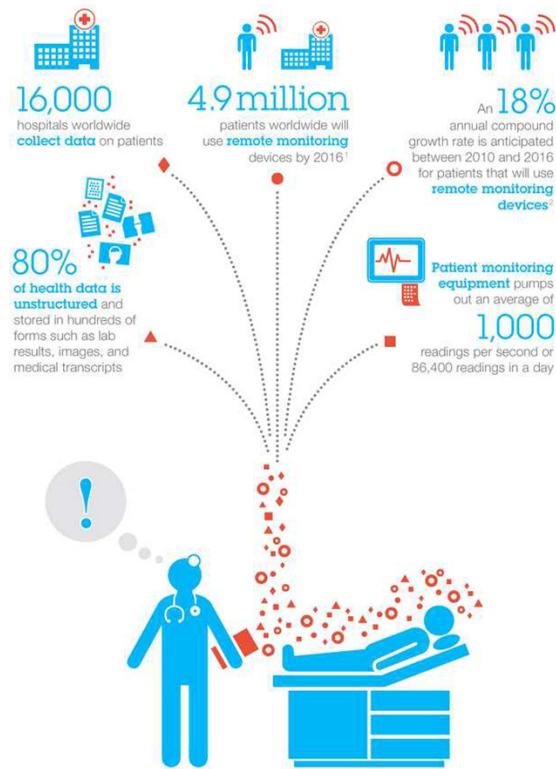
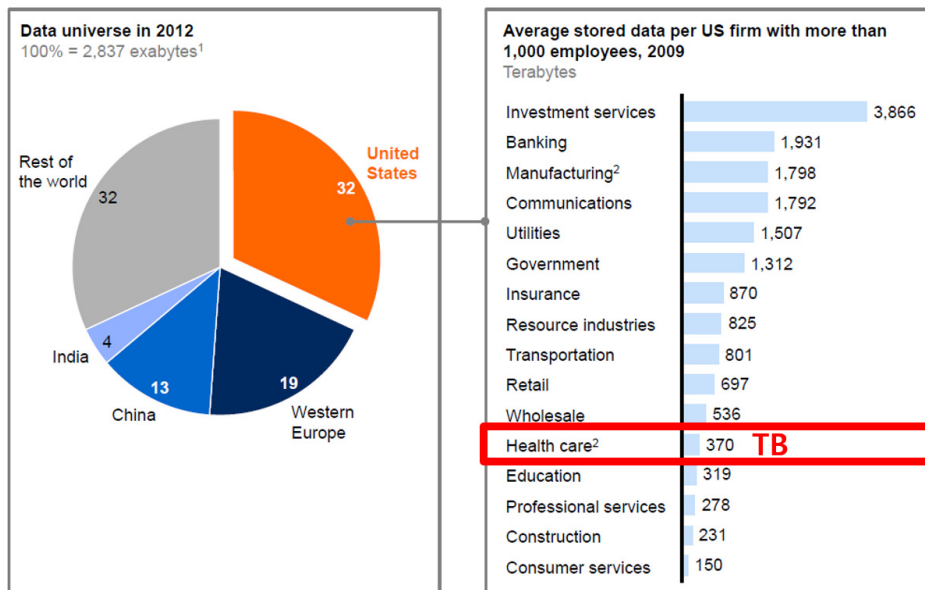


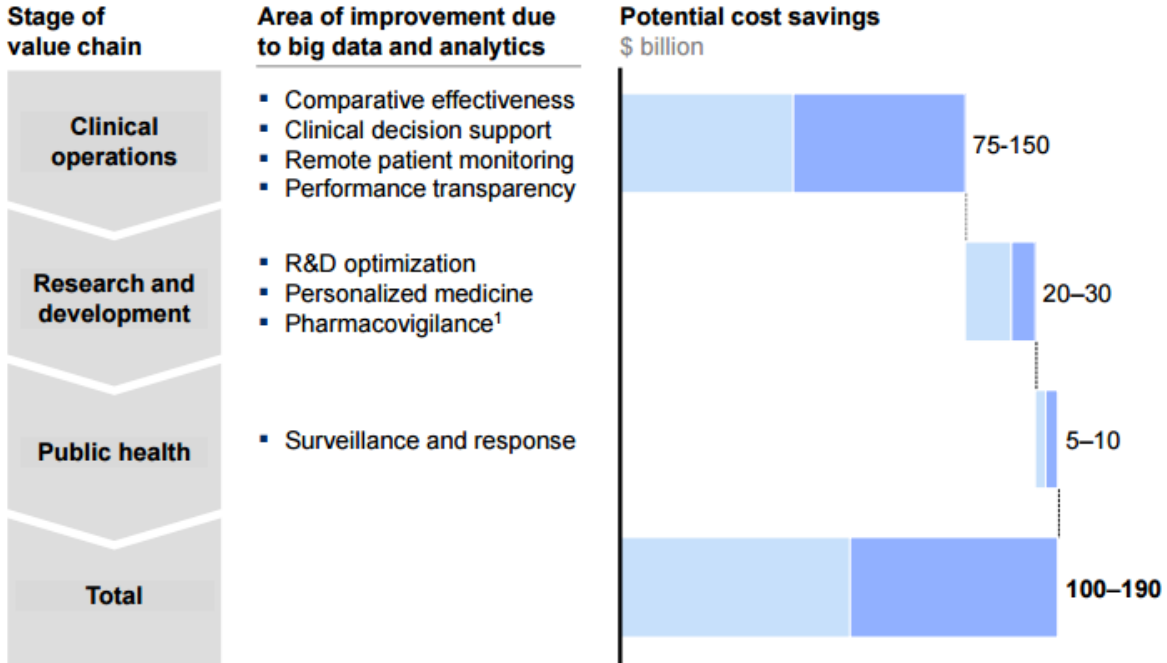
그림 1. Bigdata in healthcare, IBM



1 One exabyte = 1,024 terabytes, nearly 2.5 times as large as US Library of Congress web archive (as of May 2013).
2 The large number of firms in the manufacturing and health-care sectors reduces the available storage per company.
SOURCE: IDC; US Bureau of Labor Statistics; US Library of Congress; McKinsey Global Institute analysis

그림 2. 전세계에 축적된 데이터의 크기, McKinsey Global Institute, 2013

Big data analytics can generate up to \$190 billion annually in health-care cost savings by 2020



¹ Also known as drug safety, "pharmacovigilance" is defined by the World Health Organization as the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem.

그림 3. 미국 보건의료부문에서 빅데이터 활용효과, McKinsey, 2013

헬스케어 빅데이터 활용분야

헬스케어 빅데이터는 다양한 분야에 활용될 가치를 가지고 있다. 대표적인 활용분야는 다음과 같다.

- 현장현시 의료 의사결정 최적화(Better point-of-care decisions)
- 보건의료기관의 고객유지(Engage consumers in their healthcare)
- 재입원을 감소(Reducing Readmissions)
 - 예측모델을 사용하여 고위험 환자와 저위험 환자를 사전에 식별하고 심각한 환자에 대해 알람을 울림으로써 재입원율을 획기적으로 개선함
- 국민건강관리(Population health management)
- 보건의료기관 운영향상(Operational improvements)
- 질병치료연구 발전(Clinical research advancement)
- 정보기술비용 감소(Lower Information Technology costs)

실질적으로 환자/국민, 의료기관 및 기업을 대상으로 다양한 빅데이터 활용서비스가 제공된 해외사례가 있으며, 국내에서도 다양한 분야에 헬스 빅데이터를 활용하고자 하는 시도가 이어지고 있다.

표 1. 빅데이터 활용 서비스 해외사례

수요군	빅데이터 활용서비스	해외사례
환자/국민	개인의 유전체 분석을 통해 미래 발병가능성이 높은 질환을 예측하고 적합한 병원 의료진을 추천해주는 서비스	Genome Health Solutions Myriad Genetics
의료기관	초고속 유전체 분석 서비스 및 관련 분야의 새로운 학술정보와 치료법 등을 의료진에게 통합 제공하는 서비스	Genome Health Solutions
	초고속 유전체 분석 서비스 및 의료진이 관심 있는 데이터를 추출, 시각화, 맵핑하여 클라우드 기반으로 제공하는 서비스	DNAexus Appisty Inc.
	대규모 비정형 의료 데이터(녹음, 동영상, 사진 등)를 병원, 의료진 등이 원하는 정형 데이터 형태로 가공해주는 서비스	Health Fidelity
	각 의료기관들로부터 임상, 병원경영 및 재무성과 등 방대한 자료를 수집 및 가공하여 병원, 의료진들이 원하는 데이터를 제공하는 서비스	Explorys Inc.
	제약<의료기기 기업, 의료기관 등이 보유한 임상 데이터를 결합하여 연구자들에게 데이터 pool로 제공하는 서비스	NextBio
기업	동일 질병을 앓고 있는 환자 간에 축적되는 데이터를 기반으로 임상 또는 의약품, 의료기기의 수요예측 정보를 생성하여 제공하는 서비스	Patientslikeme
	개인, 기업, 단체 등 다양한 집단으로부터 의약품, 의료기기 기업등이 원하는 데이터를 수집하여 익명화, 표준화 등의 과정을 거쳐 제공하는 서비스	Qualcomm Life

머신러닝/딥러닝의 개념

머신러닝(Machine Learning)이란 인공지능(AI)의 한 분야로 데이터로부터 배울 수 있는 시스템을 개발하고 연구하는 것으로 예측분석(Predictive Analytics)을 지원한다. Arthur Samuel은 “머신러닝은 컴퓨터에게 배울 수 있는 능력, 즉 코드로 정의하지 않은 동작을 실행하는 능력에 대한 연구분야.”라고 정의하였으며, Tom Mitchell은 “머신러닝이란 어떠한 태스크(T)에 대한 꾸준한 경험(E)을 통하여 그 태스크(T)에 대한 성능(P)를 높이는 컴퓨터프로그램(알고리즘)을 연구하는 것이다.”라고 정의하였다. 머신러닝은 상당히 전문적인 용어였으나 2016년 3월 인공지능 알파고가 프로바둑기사 이세돌 9단과 대국하여 4승 1패로 승리하게 됨으로써 요즘 초등학교도 머신러닝이란 단어를 언급하고 있다고 한다. 알파고는 딥마인드라는 회사에서 만든 인공지능 시스템인데, 2010년 영국의 데미스 허사비스가 설립하였고 2014년에 4억달러(한화 4,800억원)로 구글에 인수되었다.

알파고는 실제로 딥러닝 알고리즘을 통해 수많은 기보를 학습하였으며, 많은 바둑프로그램과 프로기사들과 대국을 통해서 그 기량을 높여 나갈 수 있었다. 딥러닝(Deep Learning)이란 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화를 시도하는 머신러닝 알고리즘의 집합으로 정의되며, 큰틀에서 사람의 사

고방식을 컴퓨터에게 가르치는 머신러닝의 한 분야라고 이야기 할 수 있다. 즉, 딥러닝은 심층적(Deep) 머신러닝(Machine Learning)이라 할 수 있다.

딥러닝 알고리즘에는 심층 신경망 (Deep Neural Network, DNN), 합성곱 신경망(Convolutional Neural Network, CNN), 순환 신경망(Recurrent Neural Network, RNN), 제한 볼츠만 머신 (Restricted Boltzmann Machine, RBM), 심층 신뢰 신경망 (Deep Belief Network, DBN), 심층 Q-네트워크 (Deep Q-Networks) 등의 알고리즘들이 있으며 이미지/영상 인식, 음성인식, 자연어 처리 등에서 인상적인 성과를 내고 있다.

빅데이터와 머신러닝이 몰고오는 의료분야의 혁신

빅데이터와 머신러닝 기술이 결합된 스마트 의료기기는 보다 효과적이고 개인화된 솔루션을 제공함으로써 질병예방, 진단 및 치료 등의 의료분야의 혁신을 가져올 것으로 전망된다.

General Electric Company는 2015년 외부행사에서 GE가 앞으로 더욱 의료분야에 집중할 뜻을 밝히면서 “보이지 않은 것을 보이게(Make the Invisible, Visible)”란 모토와 같이 GE가 의료분야 혁신에 앞장설 것이라고 전망하였다. 이를 접한 청중들은 SF영화 프로메테우스에서 나온 미래의 의료 디바이스 Medpod 개발을 GE가 목표로 하고 있음을 알게 되었다.

미국 “인리틱”은 각종 의학 영상자료를 토대로 환자의 암세포를 식별하고 5년 이내 완치율, 사망률 등을 알려준다. 인리틱의 스마트 의료기기가 방사선 전문의 보다 더 정확하게 악성 폐종양을 찾아낼 수 있다는게 인리틱의 주장이다. 또한 미국 “23andMe”는 의사의 도움 없이 99달러 비용으로 개인 유전자 분석을 통해 질병 가능성 등 250여개의 의료정보를 제공하는 서비스를 출시했다.

시카고아동병원(CMH: Children's Memorial Hospital in Chicago)은 뇌종양 연구소 책임자인 Eric Bremer 박사의 주도로 뇌종양 gene-expression 데이터베이스 및 진단시스템을 구축하여 아동 뇌종양 진단 및 치료방법에 관하여 가장 성공적인 병원으로 평가 받고 있다. 시카고 아동병원의 뇌종양 진단시스템은 뇌종양 표본 1개당 7,000~30,000개의 측정치를 축적하고 머신러닝 기술을 접목하여 뇌종양을 분류하고, 뇌종양 관련 의학논문에 텍스트 마이닝 기술을 접목하여 뇌종양과 유전자 관계패턴을 도출하여, 243개 인공신경망 예측모델로 구성된 뇌종양 진단시스템이다.

UNC 헬스케어 (University of North Carolina Health Care)는 비영리 통합 의료기관으로 매년 37,000명 이상의 환자를 수용하는 최첨단 시설로 고품질의 의료서비스를 제공한다. UNC 헬스케어는 데이터분석을 통한 통찰력 확보를 위해 IBM의 텍스트 분석 서비스를 도입하여서 정형데이터 및 비정형데이터를 모두 분석하여 환자들의 입원 원인을 사전에 파악하여 환자 재입원 비용을 줄이는데 활용하였다.

환자의 영상과 텍스트 데이터에서 비정상(Abnormal) 정보를 추출해 내는 일은 의료진의 많은 시간과 노력을 필요로 하는데, UNC헬스케어는 유방 촬영술(mammography screenings)과 자궁경부세전 검사(Pap Smear)에 있어 콘텐츠 분석(Content Analytics)과 자연어 처리(Natural Language Processing)를 활용한 비정형 의료 데이터 분석을 수행하고 기계판독과 자동처리 알고리즘을 통해 비정상 부문을 자동 추출하여 의료진

의 시간과 노력을 획기적으로 절감하였다. UNC헬스케어의 빅데이터를 활용한 여성 암 진단시스템은 유방암과 자궁경부암 부문에서 암 진단 건수를 10% 이상 증가시켰으며, 결장암(colon cancer)과 같은 타 암의 진단에 확대 적용하고 있다.

구분	전반적 정확도	정밀도	민감도(리콜)	특이도	양성예측도
진단	78%	90%	80%	68%	90%
사후관리	79%	95%	74%	91%	95%

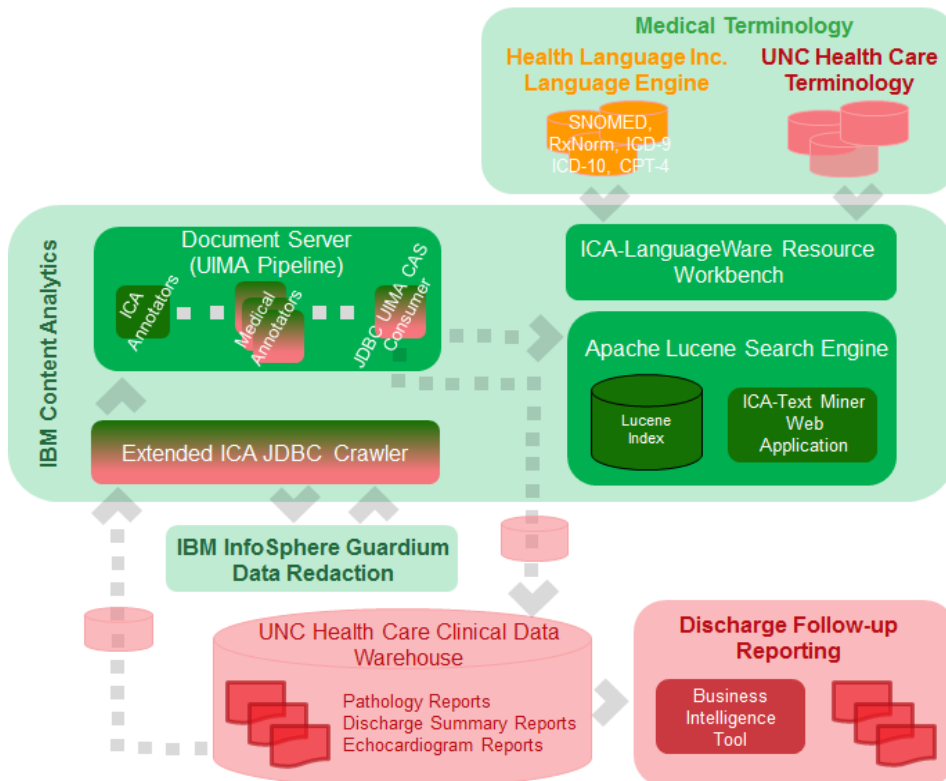


그림 4. 미국 UNC헬스케어의 여성암진단시스템 성과

2015년 8월 19일 글로벌 반도체 칩셋 제조사인 인텔(Intel)은 암환자의 유전자 및 임상 데이터 공유가 가능한 클라우드 플랫폼을 발표하였다. 이 플랫폼 명칭은 “CCC(Collaborative Cancer Cloud)”로 인텔과 미국 오레곤 의과대학 산하 나이트 암 연구소 (Knight Cancer Institute) 가 공동 개발 중이다. 인텔은 유전자 배열 분석에서 발생하는 대규모의 데이터 자원을 “CCC”를 통해 수집/공유하여 맞춤형 치료를 구현할 수 있을 것으로 예상하고 있으며, 나이트 암 연구소 소장 브라이언 드러커 역시 미래의학의 핵심과제는 모든 질병의 유전자 배열을 분석하는 것이라고 하였다.

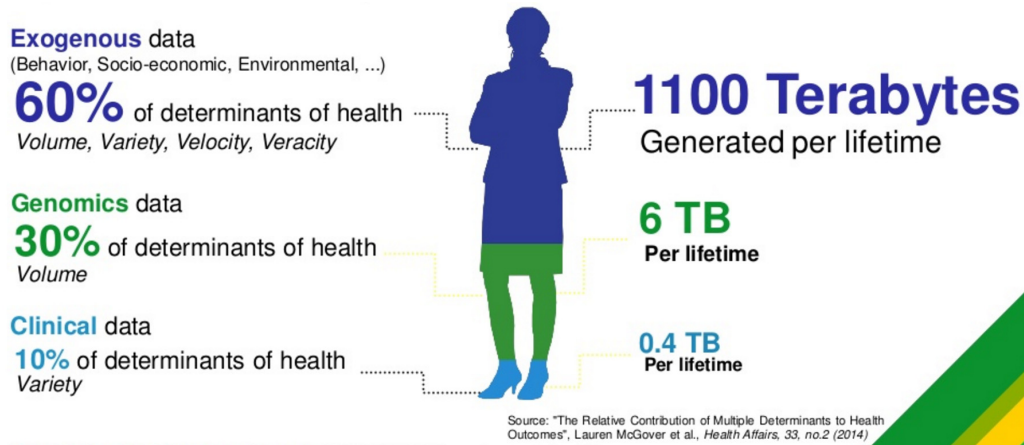


그림 5. 인간이 평생 만들어내는 데이터의 종류와 크기, IBM 2014

IBM은 2014년 “IBM Health and Social Programs Summit”에서 우리 인간이 만들어 내는 데이터는 의료데이터, 유전체데이터, 그 외에 외부적인 활동 데이터 등의 세가지 종류로 분류하고 인간이 한평생 동안에 만들어내는 이들 세 종류 데이터의 크기를 살펴보면 의료데이터가 0.4TB, 유전체 데이터가 6TB인데 반해서 그 외 외부적인 활동데이터는 무려 1100TB에 이르는데 이들 세 종류의 데이터가 우리 인체의 건강에 미치는 영향도는 각각 10%, 30%, 그리고 60%로 차이가 난다는 사실을 발표하였다. IBM은 이어 1500만 페이지 분량의 의료논문과 200여개의 의학서적과 300여개의 의료저널을 온토로지 기반으로 학습하여 여기에 머신러닝을 결합하여 암정복을 향한 도전을 시작하였다고 밝혔다. IBM은 2016년 2월에 건강정보데이터 업체 트루벤헬스 애널리틱스를 26억달러(한화 약 2조원)에 인수한 것을 비롯하여 지난 1년간 헬스케어 분야에 4번의 인수를 단행하였으며 총 40억 달러의 돈을 쏟아 부었다. 현재 IBM 왓슨헬스 분야의 직원은 5,000여명에 이르고 3억명분의 환자정보를 보유하게 되었다.



IBM Watson Vs. Cancer: The Race Is On

By mastering vast and fast-growing masses of data produced by cracking the genetic code, cognitive computing aims to help doctors fight cancer one patient at a time.

그림 6. IBM Watson의 암정복에 도전시작

IBM은 2014년 11월 미국 샌디에고의 개인유전정보 분석기업인 Pathway Genomics에 투자하기도 했는데, 이는 개인의 유전정보라는 궁극적인 개인화된 데이터까지 기존의 헬스/의료 데이터와 통합하겠다는 야심찬 계획입니다. Pathway Genomics는 웨어러블 디바이스 등에서 얻은 개인의 헬스케어 데이터와 개인의 유전정보를 통합하려는 “파노라마(Panorama)”라는 서비스를 개발하고 있습니다. Pathway Genomics가 이 데이터들을 통합하기 위해 사용하는 것이 IBM Watson의 인지컴퓨팅이다.

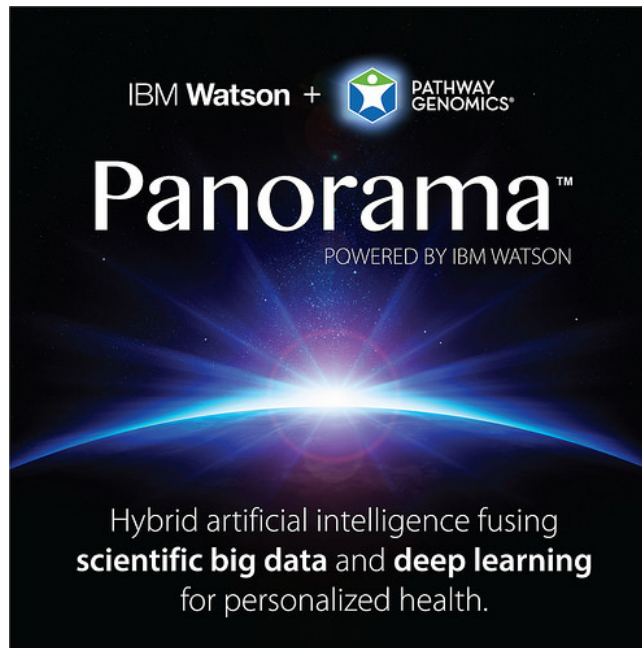


그림 7. IBM Watson과 Pathway Genomics의 Panorama 서비스

스마트폰과 웨어러블 디바이스의 보급과 함께 태어난 모바일헬스

스마트폰 및 다양한 센서 등의 발전에 따라 이제는 개인의 다양한 건강/의료 데이터가 측정되기 시작했다. 단순히 걸음 수, 활동량, 체지방률 등뿐만 아니라 체온, 혈압, 혈당, 심박수, 심전도, 산소포화도 등의 중요한 의료데이터도 다양한 모바일 디바이스와 스마트폰으로 실시간으로 연속적으로 측정할 수 있게 된 것입니다.

2014년 6월 2일 애플은 아이폰의 새로운 운영체제 iOS8을 발표하면서 디지털헬스케어 플랫폼 HealthKit과 어플리케이션 Health를 발표하였다. 헬스케어 관련 개발자를 위한 플랫폼인 HealthKit 및 일반 사용자들을 위한 대쉬보드 어플리케이션인 Health를 통해 외부의 3rd party 개발사들의 각종 헬스케어 웨어러블 디바이스 및 스마트 의료기기들이 하나의 플랫폼에서 통합적으로 관리될 수 있게 되었다.

애플은 디지털 헬스케어 플랫폼인 Health를 통해서 웨어러블 디바이스 및 휴대용 의료기기와 기존 병원의 의료시스템까지 통합하고자 Mayo Clinic과 협력을 하고 있다. Mayo Clinic은 1889년에 설립된 병원으로

매년 1.2백만명의 환자들이 방문하는 대형병원으로 디지털 헬스케어 분야의 혁신을 발 빠르게 받아들이고 있으며, 아이폰 기반의 의료 어플리케이션 Better와 파트너십을 통해 환자에게 원격의료서비스를 제공하고 있다.

애플은 미국전역에 1억명 이상의 환자에게 서비스되고 있는 미국 최대 EMR 회사인 Epic과 제휴하고 있는데, 스탠퍼드 병원, 존스홉킨스 의대, 클리블랜드 클리닉등 총 22개 유명병원들이 Epic의 EMR을 채택하고 있다. Epic은 MyChart라는 스마트폰 기반의 EMR/PHR 앱을 출시한바 있다는데, 이 MyChart 어플리케이션을 통해 환자들은 자신의 진료정보를 볼 수 있고, 의사들과 연락할 수 있으며 병원 예약관리 등 다양한 기능을 활용할 수 있다.

웨어러블 디바이스를 활용한 모바일헬스의 영역은 달리기와 걷기등과 같은 운동량 추적은 물론이고 음식 식단조절기능, 체중감량 어드바이스, 신체상태에 따른 운동지도, 기초체온관리, 수면패턴 분석 및 조언, 혈압 관리를 통한 스트레스 조절, 복용남용 및 부작용 방지에 이르기 까지 매우 다양하다. 또한 이와 같은 개인 건강정보는 의료기관의 의료정보시스템과 결합되면 자연스럽게 디지털 헬스케어 영역으로 확장될 수 있다.



그림 8. 웨어러블 디바이스 적용분야, 웨어러블 디바이스 산업백서, 2014

AT&T와 WellDoc이 개발한 당뇨병 환자 건강관리 어플리케이션 DiabetesManager는 최초로 미국 FDA의 의료기기 허가를 받은 모바일 헬스 관련 서비스이다. WellDoc은 단순히 당뇨병 환자의 건강상태를 수집하여 피드백하는 것에 그치지 않고 행동분석 (Behavioral Analysis)을 기반으로 환자의 상태를 종합적으로 평가 및 분류하여 의료전문가의 건강관리에 대한 체계적인 피드백을 환자에게 즉시(just-in-time) 제공하는 시스템을 구축하였다. 또한, 당뇨병만 아니라 약물순응도(drug adherence)부터 심혈관질환 및 정신질환까지 복잡하고 다양한 만성질환을 관리하는 mHealth 솔루션으로 확장되고 있다.

2012년 PwC Health Research Institute의 연구에 따르면 이와 같이 당뇨병환자를 위한 모바일헬스 솔루션이

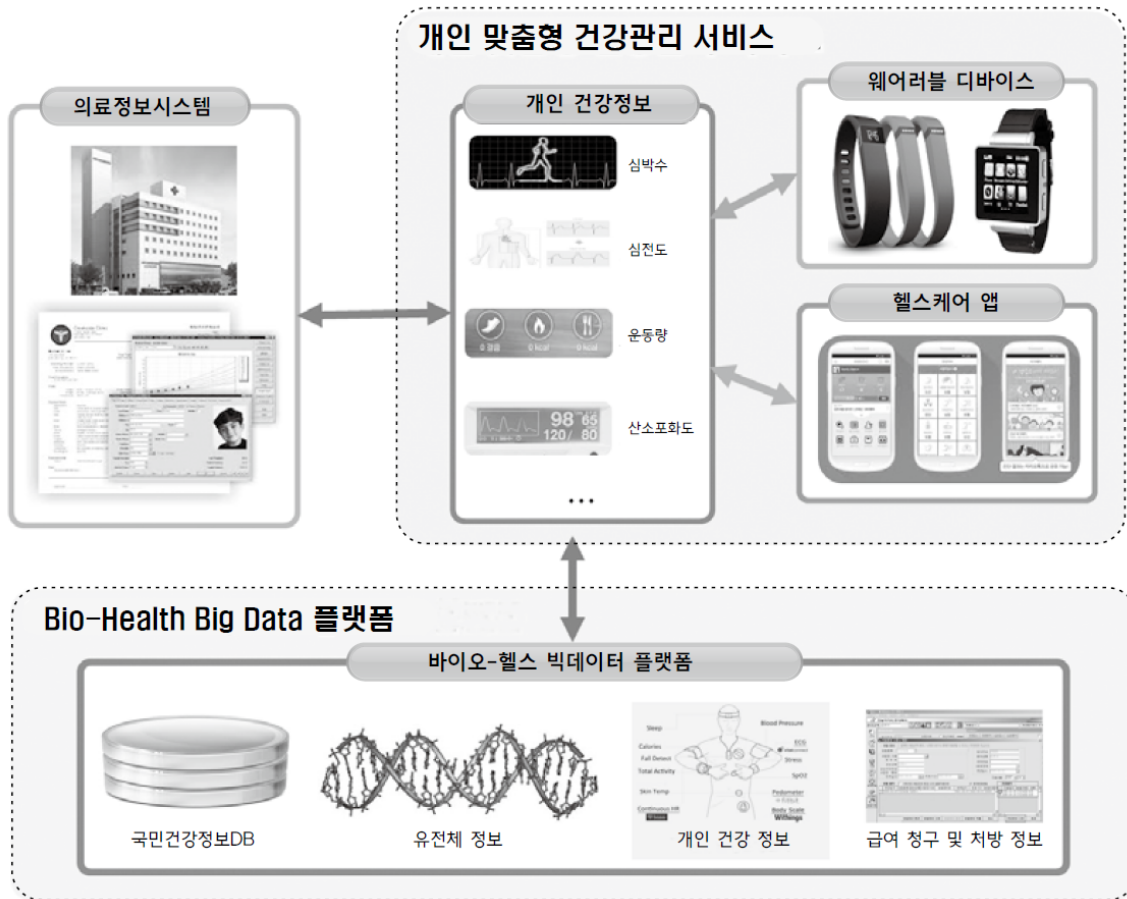


그림 9. 모바일헬스 빅데이터 플랫폼



그림 10. 당뇨병 환자 건강관리 모바일헬스 어플리케이션 DiabetesManager

충분히 활성화 될 경우 환자 당 연간 최대 10,000달러까지 절약할 수 있다는 결과가 제시된다. 이는 미국 2015년기준 당뇨병 환자수 33만명, 전당뇨(prediabetes) 환자 수 88만명 전체에 적용된다면 약 121억 달러의 절감효과를 거둘 수 있다고 한다.

국내에서도 보건복지부에서 고혈압 만성질환 집중관리 모델을 개발하고 이를 모바일헬스로 연결시키고자 하는 노력을 하고 있다.

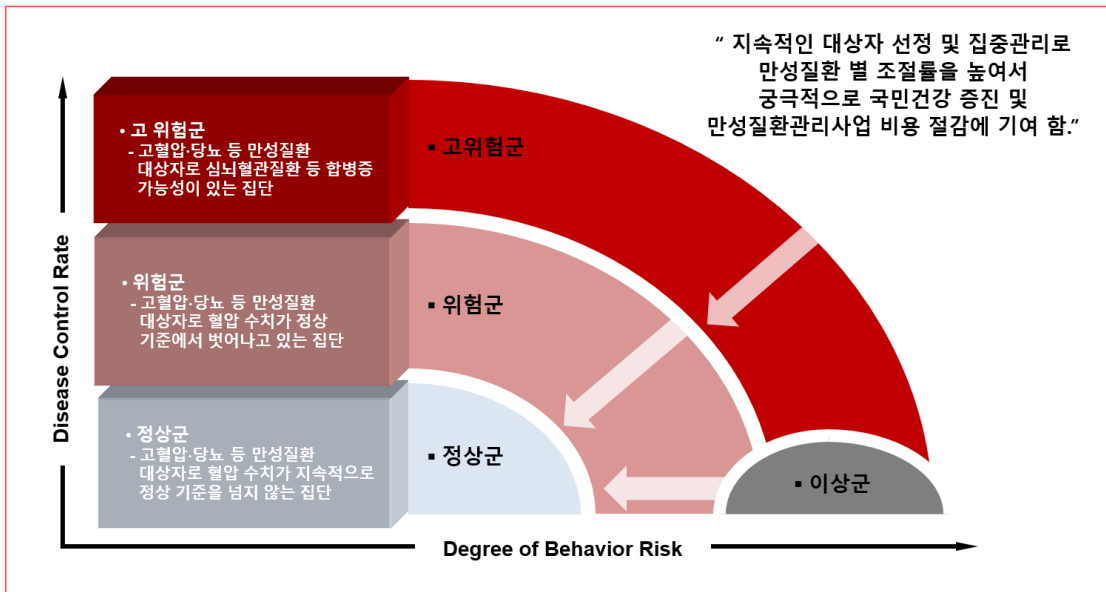


그림 11. 보건복지부 만성질환 집중관리 모델 개요

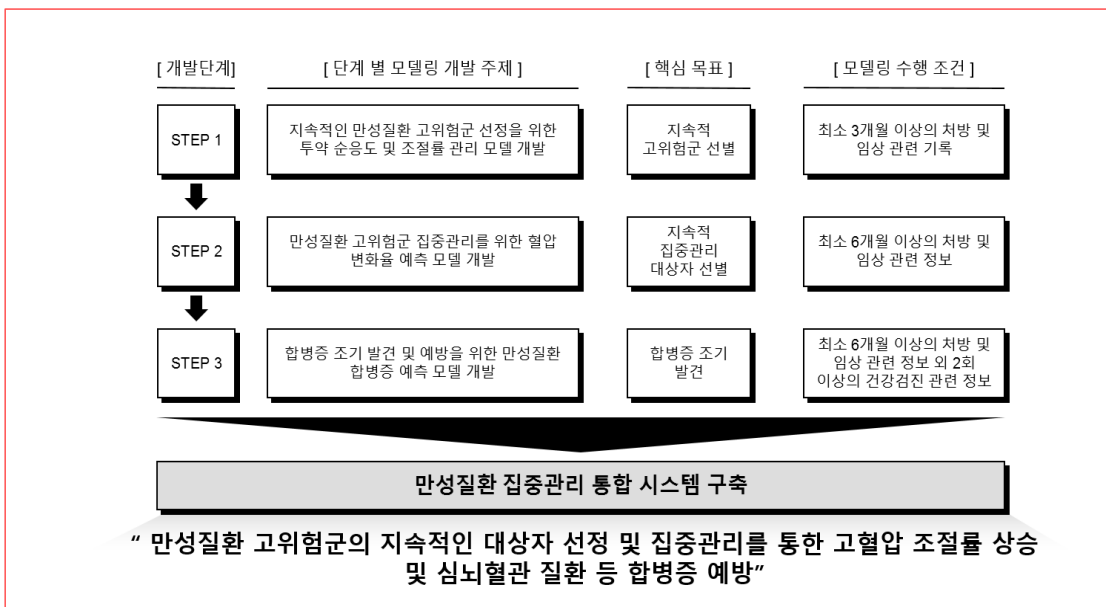


그림 12. 보건복지부 만성질환 집중관리 모델 개발전략

만성질환의 지속적·체계적·효율적인 집중관리를 위한 투약순응도 및 조절을 예측을 위해 만성질환 중 고혈압을 대상으로 하여 투약순응도 관리모델과 조절을 관리모델로 구분, 세부적으로 대상자 유형 별 특성을 파악하여 사전에 관리하기 위한 예측모델을 각각 개발하였다.

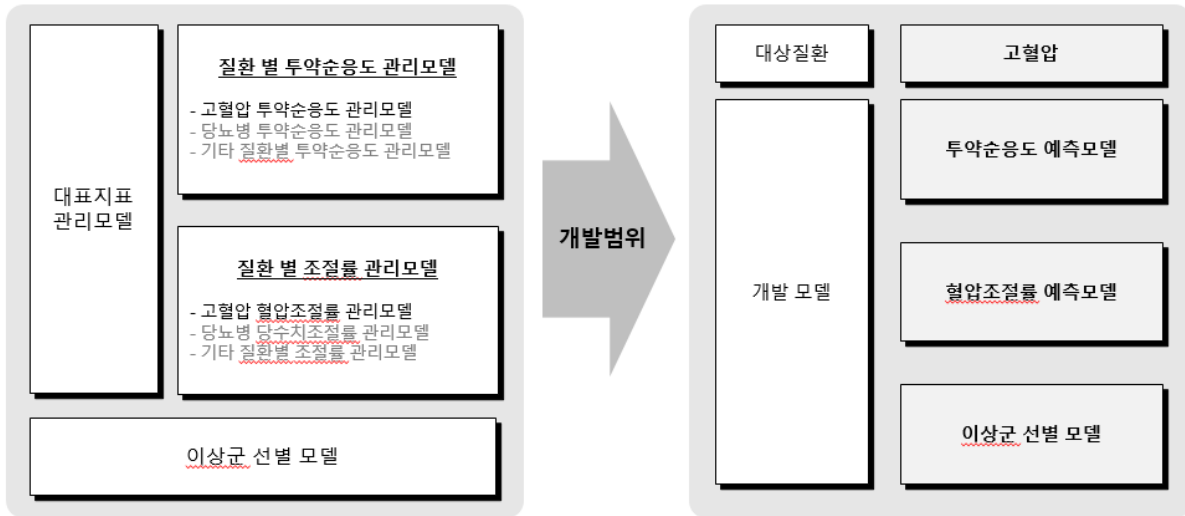


그림 13. 보건복지부 만성질환 집중관리 모델 개발범위

결론

Healthcare분야에서 빅데이터와 머신러닝의 활용은 이제 더 이상 선택이 아닌 필수로 자리잡고 있다. 선진국에서는 이미 이점을 간과하지 않고 집중투자를 하고 있으며 GE, IBM, Apple, Intel 등의 글로벌 IT기업들도 앞을 다투어 이 분야에 대한 연구개발 및 실무적용에 열을 올리고 있다. 대국민의 의료복지 수준을 획기적으로 개선하고 문화 콘텐츠 한류에서 시작한 브랜드 코리아를 의료분야로 확장시키기 위해서라도 헬스케어 빅데이터와 머신러닝에 많은 관심과 연구 및 투자가 이루어져야 한다고 생각한다.

참고문헌

1. IBM, “Bigdata in healthcare”, 2014
2. IBM, “Health and Social Programs Summit 2014”
3. McKinsey, “Game changers: Five opportunities for US growth and renewal”, 2013
4. General Electric Company “Make the Invisible, Visible” Report, 2015
5. IBM 빅데이터 적용성과, 2011
6. 웨어러블 디바이스 산업백서, 2014
7. 보건복지부 2010년 국민의료비 및 국민보건계정
8. 보건복지부, 만성질환 집중관리 통합시스템의 개요