

RIIA
Gangwon Regional Institute of Industrial Advancement

ISSUE PAPER



김성진
(주)제이엘파트너스 대표

기술개발

모바일 헬스케어 어플리케이션 개발에 있어 전주기적 데이터 수집 및 활용 방안

I. 서론

II. 본론

- 데이터의 정의
- 실사용데이터(Real World Data, RWD)

III. Dlab Flow의 활용

- 데이터의 수집 및 활용 방법

IV. 결론 및 제언

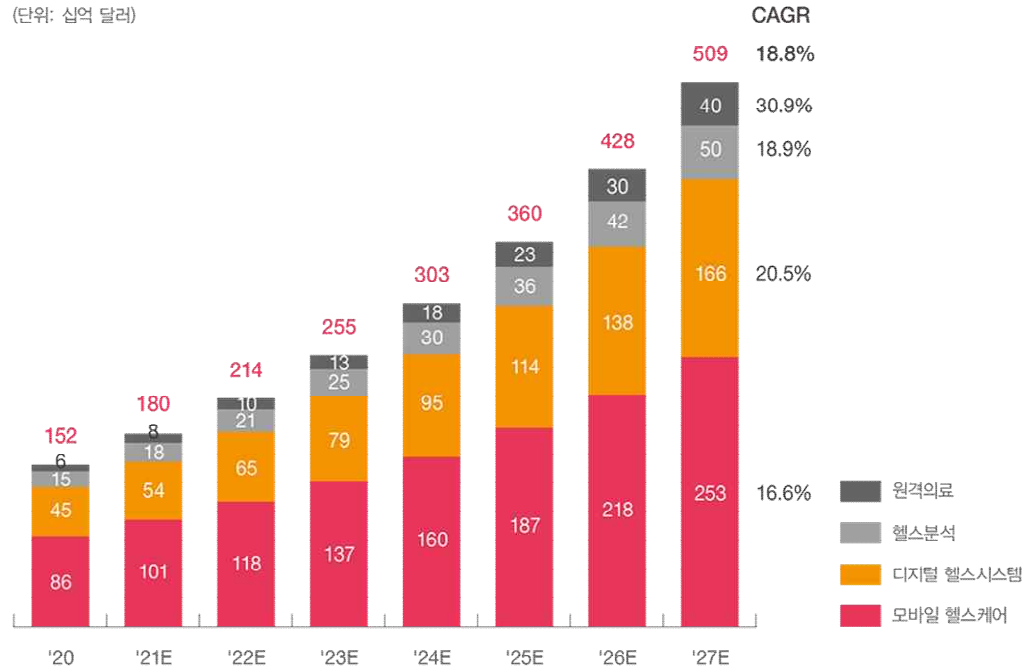
I 서론

- 정보통신 기술이 발달함에 따라 과거 대비 스마트 기기에서 수집되는 사용자 데이터가 많아지고 있으며, 해당 데이터 중 건강 관련 데이터를 활용하는 모바일 헬스케어 어플리케이션(이하 헬스케어앱)이 급증하고 있다.
- 과거에도 헬스케어 시장은 존재하였지만, 2000년대까지는 법령의 미비로 인하여 데이터는 의료기관을 통해서만 확보할 수 있었고, 사용범위도 매우 제한적이었기에 데이터 수집과 활용은 일부 대기업이나 상급종합병원(2012년 이전에는 제3차 의료급여기관으로 정의함)에서만 가능하였지만, 2000년 후반에서는 U-Health, 2010년대부터는 Smart Health로 발전함에 따라 일반 기업도 데이터를 확보하고 사업에 응용할 수 있는 기반이 마련되었다.

구분	Tele Health	E Health	U Health	Smart Health
시기	1990년대 중반	2000년대	2006년대	2012년대 이후
서비스 내용	원내 치료	치료 및 정보제공	치료/예방 관리	자가측정/치료/예방/복지/안전
주 Player	병원	병원	병원, ICT 기업	병원, 보험사, ICT기업, 서비스기업 등
주 이용자	의료인	의료인, 환자	의료인, 환자, 일반인	의료인, 환자, 일반인
주요 시스템	병원운영 (HIS, PACS)	의무기록(EMR) 웹사이트	건강기록(EHR) 모니터링	개인건강기록(PHR) 맞춤형 서비스 및 인공지능 연계 서비스

* 출처: 스마트 헬스케어산업 활성화 방안, 산업통상자원부, 2015

- 상기 표와 같이 일반 기업에게 데이터 수집 및 활용 권한을 순차적으로 개방한 이후부터 헬스케어 산업에서는 시장 및 산업 분류가 크게 변화되어 발전하고 있는 것을 다양한 조사기관의 시장 분석 보고서 등에서 확인 가능하다.



* 출처: 한국보건산업진흥원 디지털 헬스 산업 분석 및 전망 연구, 2020

- 전세계 모바일 헬스케어 산업은 2020년 기준 860억 달러에서 매년 16.6%의 성장 추세로 2027년에는 2530억 달러 수준으로 성장할 것으로 다양한 기관에서 전망하고 있고, 국내 시장은 기관 조사 내용에 따라 다양하지만 해외 시장의 약 3% 내외의 규모로 추정되고 있다.
- 본론에서는 반복 사용되는 단어인 데이터를 우선 정의한 후, 스마트폰은 물론, 스마트워치 등 다양한 환경에서의 데이터 수집 및 활용 방안에 대하여 전주기적 기획 예시를 들어 제시하고, 강원지역의 헬스케어 소프트웨어 개발 기업이 데이터를 수집하고 활용하는데 도움이 될 수 있는 방법을 제언하고자 한다.

II 본론

데이터의 정의

- 데이터란 이론을 세우는데 기초가 되는 사실 또는 바탕이 되는 자료라고 표준국어대사전에 정의되어 있으며, 국내 법령에서는 ‘개인정보 보호법’, ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률(약칭 : 정보통신망법)’, ‘신용정보의 이용 및 보호에 관한 법률(약칭 : 신용정보법)’과 같이 통칭 데이터 3법에서 데이터의 활용 방안을 규정하고 있지만, 데이터의 정의는 사용 용도에 따라 크게 달라지기 때문에, 본문에서는 데이터를 ‘의료 등 헬스케어 분야에 한정하여 사람이 생성하는 정보’로 정하고 논지를 전개하고자 한다.

실사용데이터(Real World Data, RWD)

- 사람이 생성하는 정보는 지금 이 순간에도 지속적으로 무한히 발생하고 있지만, 이를 구분하기 위하여 진료기반 데이터, 건강보험청구데이터, 라이프로그 등으로 나눌 수 있고 민간, 공공 데이터는 아래와 같이 구분한다.

구분	세부
민간 데이터	<ul style="list-style-type: none"> • 전자의무기록(EMR) : 진료과정에서 발생하는 환자의 진료 관련 정보를 기록, 저장, 조회 및 보고하는 진료정보시스템 • 전자건강기록(EHR) : 모든 의료기관의 전자의무기록(EMR)을 망으로 통합하여 공유하고 활용할 수 있는 의료정보시스템 • 개인 라이프로그(lifelog) : 유헤스케어 의료기기, 개인용 의료기기 등 환자의 일상생활에서 사용할 수 있는 의료기기로부터 발생하는 생체신호, 알람데이터 등 • 민간병원 레지스트리 데이터 : 의료기관 또는 의료기기 제조업체를 포함하여 특정 단체에서 구축한 레지스트리의 데이터로서, 의료기관의 특정 질병레지스트리 자료 등
공공 데이터	<ul style="list-style-type: none"> • 청구 데이터 : 의료기관 청구자료를 기반으로 건강보험심사평가원 또는 국민건강보험공단에서 제공하는 진료 정보, 의약품 정보, 치료재료 정보, 의료 자원 등에 대한 자료 • 공공 레지스트리 데이터 : 공공기관 또는 국가기관에서 구축한 레지스트리의 데이터로서, 국립암센터의 암과 관련하여 등록된 자료 등 • 행정데이터 : 행정안전부 또는 통계청에서 제공하는 사망원인자료 등 국가적 차원에서 조사된 자료

* 출처: 의료기기의 실사용증거(RWE) 적용에 대한 가이드라인 2023.7, 식약처

- 헬스케어 소프트웨어 개발 기업이 보다 쉽게 접근 가능한 것은 상기 표의 민간데이터 - 개인 라이프로그이며, 의료기기로부터 발생하는 신호 이외에도 일반 스마트기기에서도 확보 가능한 정보를 통칭하는 것으로 설정한다.
- 데이터를 활용하는 소프트웨어 개발 시, 아래와 같은 프로세스를 적용하여 제품 기획을 하는 경우가 일반적이며 비즈니스 문제를 제시하고 이를 해결할 수 있는 원천데이터를 확보하는 방법을 기획하여야 한다.



* 출처: Big data in business analytics: Talking about the analytics process model, 2019.05

- 전주기적 데이터 수집 및 활용을 위하여 상기 1~7개 프로세스 중, 7. 모델 해석 및 평가 배포를 제외한 나머지 1~6단계 항목에 있어 간단히 설명하고 예시를 들어 이해를 높이고자 하며, 피부암 진단을 목적으로 하는 헬스케어 앱을 개발한다고 가정할 때, 1~7단계의 구체적인 기획 사례를 들어보기로 하며, 논쟁을 막기 위하여 의료법이나 개인정보보호법 등의 법령이나 규정은 해소된 것으로 설정한다.

구분	설명	예시
1. 비즈니스 문제 식별	헬스케어 앱을 기획할 때는 해결할 문제가 사전 정의되어야 한다. 예를 들어 피부암을 진단하기 위한 앱을 만든다고 가정한다면, 피부암 범위를 확정해야 하며 피부암이 어떤 특징을 가지는 것인지, 진단을 어떤 방식으로 할 수 있는지를 정하여야 하는데, 이는 의학적 배경을 바탕으로 설정되어야만 데이터 수집의 방향을 정할 수 있다.	<p>피부암을 진단하는 앱을 개발한다. 피부암 중 발생빈도가 높은 악성흑색종을 대상으로 설정한다. 악성흑색종은 영상으로 판별 후, 조직검사를 통하여 확진 가능한 질병이므로 수집할 데이터는 피부 사진으로 정할 수 있다.</p> <ul style="list-style-type: none"> • 비즈니스 문제: 피부암 진단 필요 • 필요한 데이터: 피부 사진(이미지 등)
2. 데이터 소스 식별	여기서 말하는 데이터소스(DataSource)는 모든 방법을 활용하여 모든 곳에서 수집할 수 있는 원천데이터를 의미하며, 데이터를 많이 수집할수록 초기 목적을 달성할 수 있으므로 가능한 많은 데이터소스를 기획 단계에서 넣는 것이 좋다. 물론 가용할 수 있는 자원의 범위 내에서 많다는 것이 지, 무한정을 의미하는 것은 아니다.	<p>악성흑색종의 사진은 병원에서 돈을 주고 구입할 수도 있지만, 우리는 그 정도의 비용을 지급할 수 없다. 헬스케어 앱 사용자가 직접 찍는 사진을 데이터소스로 정하고, 사진을 촬영하고 저장하는 기능을 만들어서 사용자가 직접 사진을 올릴 수 있도록 한다.</p> <ul style="list-style-type: none"> • 데이터소스: 이미지파일(정상, 악성흑색종)

3. 데이터 선택	수집된 데이터를 구분하여 어떤 데이터가 헬스케어 앱에 적합한지 자체 검증을 하는 작업이 이 단계이며, 많은 데이터 중 정형화가 가능한 데이터를 추출하는 것도 본 단계에서 필요한 작업 중 하나이다. 이 때 선택된 데이터는 객관화되어 사용하기 쉬운 정형화 데이터를 우선으로 하지만, 앱의 목적에 따라서 비정형화된 데이터를 그대로 사용하는 방법을 택할 수도 있다.	앱을 개발했다고 가정하고, 데이터베이스에 쌓인 이미지가 있으니 선택한다. 이미지 중에서 해상도가 낮거나, 떨림이 있는 영상 등을 제거한다. • 데이터 선택: 1. exif를 확인하여 얼마 이상의 해상도를 가진 이미지 2. exif를 확인하여 1M 이상의 용량을 가진 이미지
4. 데이터 정제	정형화 또는 비정형화된 어떤 데이터를 사용하기로 3단계에서 정하였다면 이 단계에서는 최소 범위를 설정하여 정제하는 과정을 진행하여야 한다. 예를 들어 피부암을 진단하기 위한 앱에서 피부 사진을 촬영하여 사용하고자 한다면, 암의 특성 상 장기 추적이 필요하므로 1초 단위의 사진을 촬영할 것이 요구되지는 않을 것이다. 그렇기 때문에 특정 부위의 촬영된 사진을 1개월 단위로 정제할 것인지, 1주일 단위로 정제할 것인지를 결정하여야 한다.	사용자는 피부 사진을 매일 찍을 수도 있지만, 이전에 찍은 사진을 업로드할 수도 있으므로, 매일 찍은 사진을 부위별로 구분하고 일자별로 구분하여 리스트로 변경한다. 이미지 파일의 exif를 식별할 수 있다면, 한 번에 업로드한 사진도 일자에 따라 리스트로 변경하여 저장한다. • 데이터 정제: 3단계의 데이터 중, 정기적 시간에 따라 1일~1개월 정도의 이미지 리스트를 저장
5. 데이터 변환	4단계에서 정제된 데이터는 이미지, 텍스트 등 다양한 파일 확장자로 수집되어 있으므로, 이를 보다 사용하기 쉽고 신뢰 가능한 리소스로 변환하는 것을 의미한다. 예를 들어 피부 사진 이미지 파일이 정제되었다면, 이를 표준화하여 헬스케어 앱 데이터베이스에 넣어놓고 활용할 수 있도록 사진의 가로세로 사이즈를 정하여 그만큼만 잘라내서 저장하는 등 전반적인 활동이 필요하다.	원본 데이터베이스에 저장된 이미지는 그대로 두고, 4단계에서 리스트로 구분된 데이터를 대상으로 잘라내기 작업을 진행한다. • 데이터 변환: 이미지가 가로, 세로, 파노라마 등 다양한 사이즈가 있으므로 환부를 판별할 수 있도록 가로세로 사이즈를 확정하고 그만큼 잘라내서 새로 저장한다.
6. 데이터 분석	변환된 데이터집단을 활용하여 초기 목적을 달성할 수 있는지를 검토하는 단계이며, 이 과정에서는 전문가를 활용하여 정답과 오답을 걸러내는 작업을 한번 해보는 것이 앱의 신뢰성을 높이는 데 필요하다. 헬스케어 앱에서는 피부암이라고 진단하였는데, 피부과전문의도 피부암이라고 진단할 가능성도 서로 비교하여야 정확도를 보다 높일 수 있다.	5단계에서 새로 구성된 데이터 리스트를 활용하여 2단계에서 확보한 정상, 악성흑색종과 비교하여 판별하고, 판별한 값을 피부과 전문의에게 진단을 받아 비교한다. • 데이터 분석: 1. 자체 기술로 정상, 악성흑색종으로 구분한 결과값 2. 피부과 전문의가 동일한 이미지를 보고, 정상, 악성흑색종으로 구분한 결과값 3. 1과 2를 비교한 결과값으로 산출한 앱 진단의 정확도
7. 모델 해석 및 평가 배포	이 단계는 데이터 수집 이후 활용 과정에서 필요하며, 기 확보된 데이터리스트와 새로 수집되는 데이터를 비교하면서 개발된 분석모델(헬스케어 앱)의 정확도와 신뢰도를 지속적으로 높여가는 과정을 반복한다.	6단계의 3번 항목 정확도를 높이기 위하여 지속적으로 사진을 수집하고, 모델을 주기적으로 업그레이드한다.

- 위와 같이 수집 기획이 되었다면, 데이터를 어떻게 수집할 것인지 결정하여야 하는데, 가장 쉬운 방법은 병원 등에서 만들어 놓은 데이터를 구매하는 것이지만 이 방법은 금액이 많이 들기 때문에 중소기업에서는 접근하기 어렵다.
- 두 번째 방법으로는 국가 사이트를 이용하는 것인데, 'AI HUB <https://www.aihub.or.kr/>'에서 필요로 하는 데이터가 있는지 검색해보고 다운받아서 사용하는 방법이 있다.

* 출처: <https://www.aihub.or.kr/>

- 세 번째 방법으로는 사용자의 동의를 얻어서 필요한 데이터를 수집하는 방법이 있는데, 헬스케어앱이 출시된 후에 가능하기에 수집에 시간이 오래 걸리는 단점이 있지만, 신뢰성이 높은 데이터를 확보할 수 있는 장점이 있다.

III Diab Flow의 활용

데이터의 수집 및 활용 방법

- 헬스케어앱은 기본적으로 스마트기기를 활용할 수 있으므로 사용자 동의를 받은 후, 수집할 수 있는 데이터가 많은데, 그 중 위치를 추적할 수 있는 GPS 데이터가 대표적이다. GPS 데이터는 실내가 아닌 곳에서 수집이 가능한데, 이를 활용한 것으로 사용자의 이동거리, 이동시간 등을 확인할 수 있으므로 만보기와 같이 운동량을 계산할 수 있는 것은 물론, 위치 정보가 포함되어 있기에 네이버, 구글, 카카오 지도와 연계한 서비스 기획이 가능하다. 이와 같이 헬스케어앱 기획 과정에 있어 가장 중요한 것은 사업 환경이 변경되더라도 변화되지 않는 데이터 기준을 설정하는 것이 필요하며, GPS데이터는 그 중 하나이므로 필수 수집 데이터로 설정하여 사용자 동의를 구할 때 반드시 추가하여야 한다.
- 수집된 데이터는 사용자의 동의에 따라 재가공과 활용이 가능하므로, 불법적 사용만 피한다면 어떤 방향으로든 사용할 수 있다. 불법 활용의 대표적인 사례는 보건의료데이터를 비식별화하지 않은 상태에서 유통하는 것이므로, 헬스케어 앱 내에서 재가공하여 사용자를 위하여 활용하는 것이 아니라면 앱 사용자의 동의를 받았다고 하더라도 법적 허용 범위를 사전에 반드시 확인한 후에 외부로 반출하여야 한다.

〈비정형데이터 가명처리 방안 개정 전·후 비교〉

구분	현 행	개 선
유전체 데이터	널리 알려진 질병의 유전자 변이유무 확인 등 예외적인 경우를 제외하고 가명처리 유보	NGS 기반 유전자 검사를 통해 생성된 표준화된 SAM 및 VCF 등은 메타데이터 및 염기서열에 대한 가명처리
자유입력 데이터	가명처리 유보	자연어 처리기술 등 정형데이터로 변환 후 식별정보 삭제·대체 등 방법으로 가명처리
음성 데이터	가명처리 유보	음성인식 기술 이용하여 텍스트 처리한 뒤 자유입력 데이터 가명처리 방법으로 처리

가명정보 처리·활용 시	가명정보 제3자 제공 시
<ul style="list-style-type: none"> • 가명정보 처리·활용 과정에서 의도치 않게 특정 개인을 알아볼 수 있는 정보가 생성되었다는 사실만으로 가명 정보를 처리한 자 또는 해당 가명정보를 제공한 자를 처벌하지 않음 ※ 단, 해당 정보의 처리를 즉시 중지하고, 지체없이 회수·파기하여야 함 	<ul style="list-style-type: none"> • 가명정보를 제공받은 자의 안전조치 미이행 등으로 가명정보 유출 등의 문제가 발생하였거나 고의로 재식별 행위를 하는 등 그 행위 주체의 위법으로 발생한 사안에 대해서는 해당 행위자만 제재함

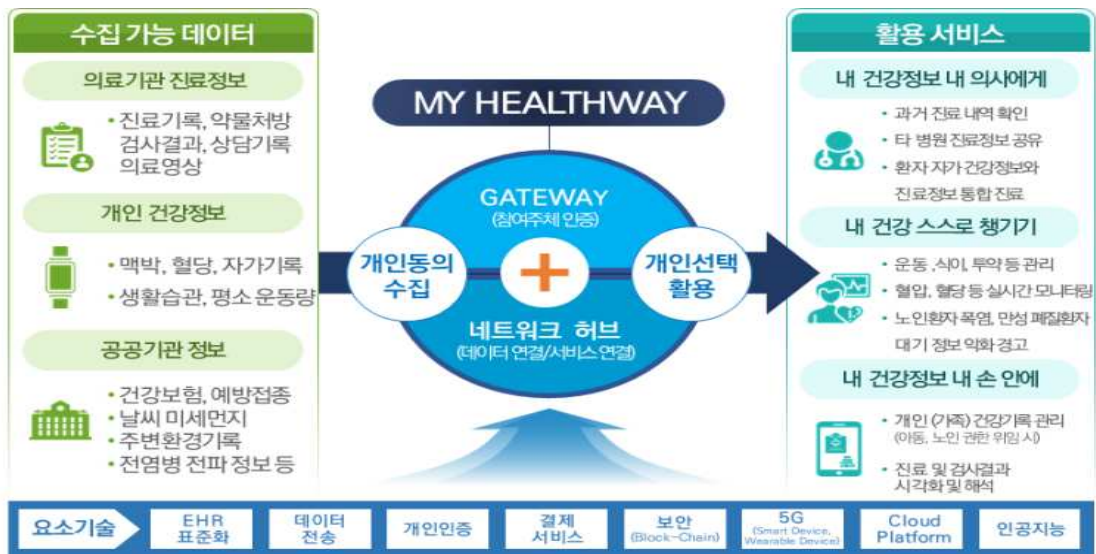
* 출처: 보건의료데이터 가명처리 방법 개선, 보건복지부 2024

IV 결론 및 제언

- 지금까지 헬스케어 앱을 기획할 때 사전에 검토하여야 하는 데이터 수집과 활용 방법을 예시로 알아보았다. 2012년 마이헬스웨이 구축 보도와 2023년 국가 마이데이터 혁신 추진 전략에 따르면, 2025년 이후에는 의료마이데이터를 보다 쉽게 활용할 수 있는 기반 플랫폼을 마련할 계획이므로 헬스케어 앱을 사업화하는 기업이라면, 공공데이터에서 사용자의 선택적 동의를 통해 확보 가능한 개인정보와 융합할 수 있도록 민간데이터를 수집하도록 기획을 할 필요가 있다.



* 출처: 국가 마이데이터 혁신 추진 전략, 관계 부처합동 2023



* 출처: 마이 헬스웨이((가칭)건강정보 고속도로) 구축 시작, 부처합동 보도자료 2021

- 강원지역에서 헬스케어 앱을 개발하는 기업이라면, 강원지역산업진흥원(<https://gw.riia.or.kr/>)의 사업공고 및 자료정보를 수시로 열람하면서 데이터 수집이나 R&D를 지원해주는 사업에 관심을 갖고 신청하는 것이 필요하다.

2024년도 지역혁신클러스터육성(BIR&D) 혁신셀 이슈페이퍼

RIIA (재)강원지역산업진흥원

강원특별자치도 춘천시 강원대학길 1 60주년기념관

<https://gw.riia.or.kr/>