

중증질환 위험도 예측 모델 개발 및 위험도 수준별 맞춤 의료 서비스 제공으로 헬스케어 시장 선점

B반 2조

공은비 김규리 김송일 박규동 박병수 이수진

INDEX

001 — 추진 배경

002 — 현황 및 개선기회

003 ----- 분석 계획

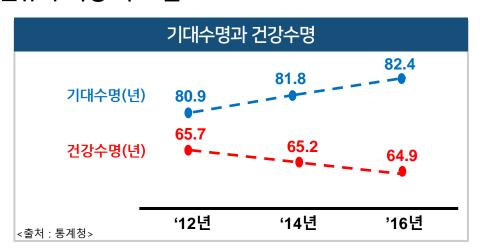
004 ----- 분석 결과

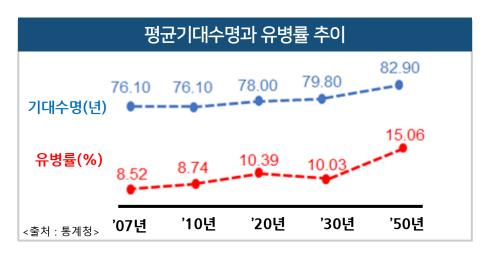
005 ---- 개선안

006 — Learned Lessons

01 추진 배경

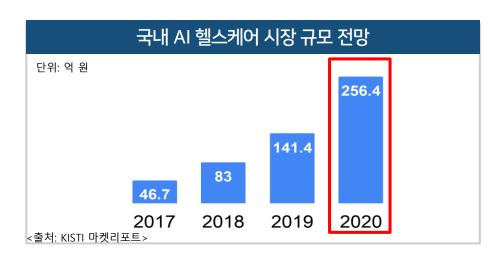
기대수명 증가와 더불어 유병기간이 증가하므로 건강한 삶을 유지하기 위한 의료 서비스 제공을 통해 신규 수익성 확보 필요





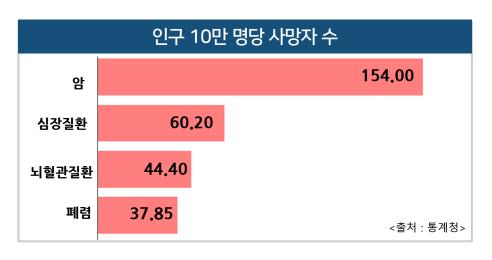
🛟 바이오ICT 융합기술의 발전으로 스마트 헬스케어 산업이 부상함에 따라 국내시장 진출을 통한 신 성장동력 확보 필요





02 현황 및 개선기회

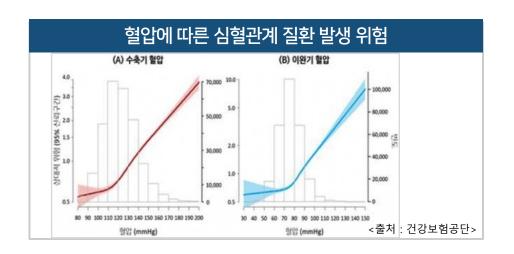
🛟 🛮 중증질환은 사망률이 높고 조기 발견이 중요하므로 중증질환 위험도를 사전에 진단할 수 있는 수단 필요



"<u>뇌혈관 질환은 예방 및 조기발견이 가장</u> 중요하다. 조기에 발견되면 뇌손상이 발생하기 전에 약물 치료를 실시해 빠른 회복 및 일상생활 복귀가 가능하다."

- 오세양 인하대병원 심뇌혈관센터 교수 -

🛟 심뇌혈관질환은 고혈압, 당뇨 등 다양한 기초질환을 원인으로 발생하는 만큼 예방을 위한 기초질환 관리 프로그램 필요



"당뇨병 환자의 32.3%가 심뇌혈관 질환 환자이다. 또한 <u>심뇌혈관 질환이 당뇨병 환자</u> 사망원인의 50.3%를 차지하다."

- 심혈관 당뇨학회지 2017 09 -

02 현황 및 개선기회

의료인력과 병원이 수도권에 집중되고 농어촌과 대도시간 의료자원 불균형이 갈수록 심각해지므로 이를 해소할 의료 서비스 제공 필요



"대형병원을 찾아 몇시간씩 가야하거나, 고가의 검사부터 받아야하거나… 우리 의료의 현실이다. 일상의 건강관리 체계를 만들면, 질병이 악화하기 전 평소 관리에 힘쓸 수 있을 뿐 아니라 <u>지방 주민과</u> 같은 의료 소외 계층을 끌어안을 수 있다." - [국민일보] 김양중 의료전문기자 -

🗘 과제수행목표

측정지표 (KPI)	운용 정의	가중치 (%)	현수준		목표수준		
			연구판	2019년	2020년	2021년	
신규 가입자 수	서비스 신규 가입자 수(명)	70	-	1,000	10,000	20,000	
서비스 이용자 만족지수	고객 만족 요소 정량화 지수 (질병관리 적합정도, 건강증진 도움 정도 등) 조사 결과(점)	30	-	90	93	95	

03 분석 계획

목적	분석 계획				
⊐`⊣	분석방법	분석 내용			
	막대 그래프	· 성별, 연령, 흡연상태 등 범주 변수의 각 도수 확인			
각 변수의 분포 및 단일 변수 특성 파악	히스토그램	· 혈압, 콜레스테롤, 공복혈당 변수 등의 분포 및 이상치 확인			
	Box Plot				
	누적 막대 그래프	· 연령, 성별로 그룹화 하여 중증 질환 발병 비율 차이 확인			
각 변수간의 주요 특성 및	구국 국내 그대트	· 기초 질환 유무로 그룹화 하여 중증 질환 발병 비율 차이 확인			
주요 질환간 관계 파악	통계적 검정	· 연령, 성별에 따른 중증 질환 발병위험도 차이 검정			
		· 기초 질환 유무에 따른 중증 질환 발병위험도 차이 검정			
	로지스틱 회귀분석				
	의사결정나무	· 개인 건강검진 데이터로 중증 질환의 발병 위험도 예측 모델링			
중증 질환유형별 발병위험도 예측 모델 개발	랜덤 포레스트	· 중증 질환에 영향을 미치는 인자 확인			
	Gradient Boosting	· 평가 지표를 종합적으로 고려하여 가장 높은 모형 선정			
	XGBoost				

04 분석 결과

- 지역간 의료격차 심화로 중증질환 진료가 가능한 상급종합병원 수가 수도권에 집중되어 있어 타 지역의 중증질환에 대한 자가 진단 및 사전에 관리할 수 있는 수단 필요
 - 중증질환 진료가 가능한 상급종합병원이 수도권과 광역시에 집중되어 있고, 특정지역의 사람에게만 제한적으로 이용됨
 - 의료 접근성에 구애 받지 않고 중증질환을 자가 진단 및 사전에 관리할 수 있는 수단 필요

전국 상급종합병원 분포 상급종합 ■ 종합병원

인구수 대비 의료 기간 수 차이 검정						
지역	인구대비 상급종합병원					
서울	3.2					
부산	1.6					
대구	0.7					
인천	1.1					
광주	1.2					
대전	0.6					
울산	0.4					
세종	0					
경기	3.3					
강원	0.8					
충청도	1.4					
전라도	2					
경상도	2.5					
제주	0.3					

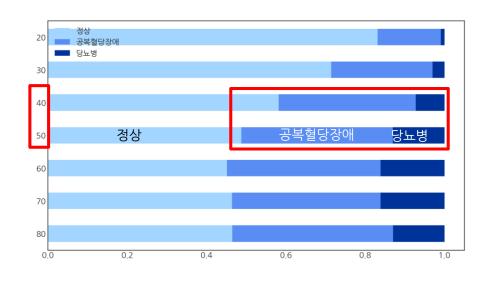
p_value: 0.002

※ 인구 100만 명 당 병원 수

※ 병원 등급: 상급병원 > 종합병원 > 일반 병원 > 의원

04 분석 결과

- 🛟 연령별 당뇨 위험도 수준 분석결과 40, 50대 남성이 위험도가 높다고 도출됨
- 🛟 연령별 심뇌혈관질환 위험도 수준 분석결과 40, 50대 남성이 위험도가 높다고 도출됨
 - ➡ 주요 질병의 위험도가 높은 중년남성 대상 집중서비스 필요







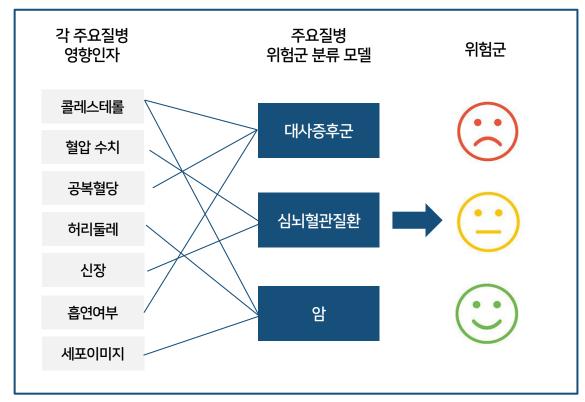
결당상애	당뇨병
2044	851
886	382
209	67
	2044 886

심혈관질환	저위험	중등도위험	고위험	최고위험	심혈관질환	저위험	중등도위험	고위험	최고위험
나이					나이				
20	2440	322	368	3	60	2619	1870	742	48
30	5100	1576	1039	8	70	1162	801	368	38
40	5918	2354	1279	39	80	251	166	89	10
50	4080	3527	880	66					

04 분석 결과

🛟 건강정보를 활용한 중증질환별 발병위험도 사전 제공을 위한 예측 모델 개발 결과

- 질병 예측 모델 평가에서는 재현율에 더 높은 가중치
- 평가 지표를 종합적으로 고려하여 중증질환별 가장 우수한 모델 선정



▲ 즈요	거간 지표르	HIEHO로	저이하	위험군 지표
	116 AIII	'비등'프도	(A) = Y	기위한 시표

* 대사증후군은 정상, 대사증후군 위험단계 심뇌혈관질환은 저위험, 중등도위험, 고위험, 최고위험군 암은 양성 종양, 악성 종양으로 구분

		Models						
구분	평가 지표	Logistic Regression	의사결정나무	Random Forest	Gradient Boosting	XGBoost		
	정확도	0.844	0.919	0.917	0.922	0.921		
대사증후군	재현율	0.671	0.822	0.775	0.827	0.829		
	F1 Score	0.715	0.855	0.845	0.861	0.859		
심뇌혈관질환	정확도	0.769	0.817	0.896	0.989	0.988		
	재현율	0.461	0.506	0.72	0.976	0.965		
	F1 Score	0.478	0.524	0.89	0.978	0.974		
암	정확도	0.89	0.932	0.945	0.945	0.959		
	재현율	0.853	0.898	0.889	0.908	0.917		
	F1 Score	0.853	0.906	0.92	0.923	0.941		



대사증후군 XGBoost 심뇌혈관 Gradient Boosting 암 XGBoost

05 개선안

중증질환을 진료할 상급종합병원 수가 수도권에 편향된 점을 고려하여, 의료기관 접근성이 낮은 사람도 질병을 자가 진단하고 관리할 수 있는 플랫폼 구축

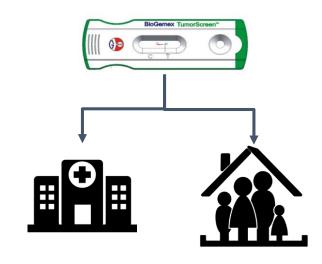
중증질환(암, 심뇌혈관질환) 자가진단 서비스 개발



의료 정보 공유 플랫폼 개발



웹 서비스와 연동가능한 건강검진키트 제공

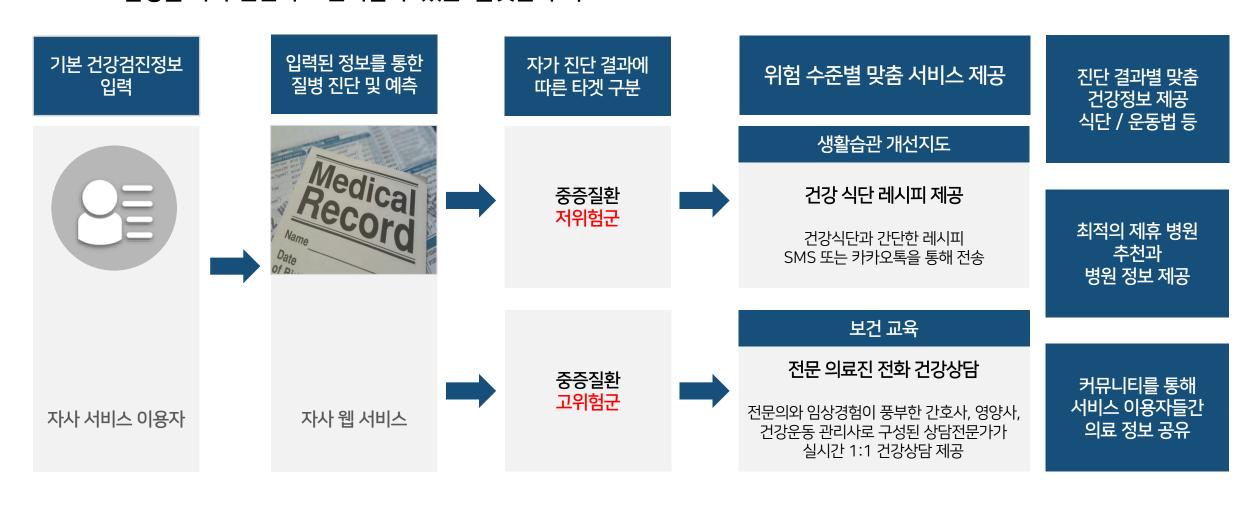


- 자가진단 결과 해석 전화상담 제공
- 일상 데이터 기록에 따라 위험수준 변화여부 등 정보제공
- 최적의 제휴 병원 위치 및 정보 제공
- 중증질환 관련 대형병원의 명의 안내 및 의료전문진 / 사용자 커뮤니티 제공

- 제휴 의료기관 또는 보호자에게 데이터 제공
- 헬스케어 서비스 콜센터 통화 시 ARS 단계 없이 상담원과 바로 연결
- 모든 안내와 상담에 느린 말 서비스 제공

05 개선안

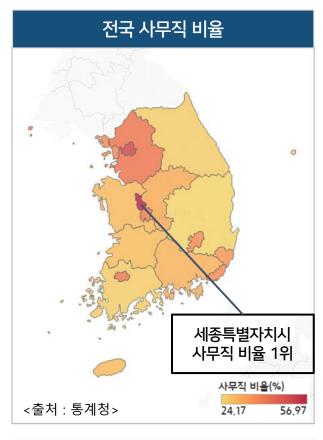
중증질환을 진료할 상급종합병원 수가 수도권에 편향된 점을 고려하여, 의료기관 접근성이 낮은 사람도 질병을 자가 진단하고 관리할 수 있는 플랫폼 구축



05 개선안

병원·기업과 연계하여 중증질환 위험도가 높은 40-50대 남성 대상 건강 관리 프로그램 제공





"<u>사무직의 중증질환으로 인한 사망률</u>이 전체 대비 약 17% 높은 것으로 나타났다." - 메디컬옵저버 안경진 기자 -



STEP 1: 파일럿 타겟 선정 기준

- 중증질환 위험도가 높은 <mark>중년 남성 비율</mark>이 높은 지역
- 전체 취업자 수 대비 <mark>사무직 인구 비율</mark>이 높은 지역



STEP 2: 지역 선정 및 제휴 병원과 연계 서비스



거점 지역의 제휴 병원·기업과 연계하여 4-50대 남성 대상 생활 습관 형성을 위한 당뇨, 고지혈증, 비만 관리 프로그램 실시



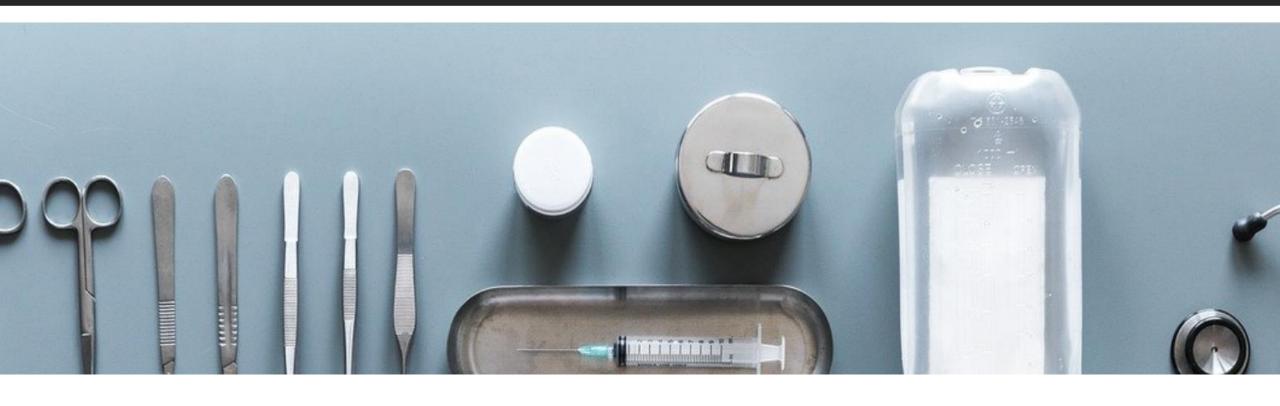
건강검진 예약 대행 및 우대 전국 제휴 종합병원, 전문 검진센터 네트워크를 활용한 건강검진 예약 대행 및 우대 혜택을 가족에게도 동일하게 적용

[서비스 구현 모습]

127.0.0.1:5000

.....

메인화면 진단하기 최적의 병원 찾기 서비스 정보 회사소개



06 Learned Lesson



[조장] 박규동: 5주 동안 빅데이터를 배운 것으로 조원분들과 같이 의료 데이터를 수집하고, 분석한 경험이 값지다고 생각합니다. 하지만 돌이켜보면 저의 부족한 역량 때문에 많은 아쉬움이 남습니다. 또한 저를 따라준 조원분들에게 감사하다는 말을 남깁니다.



[조원] 공은비 : 프로젝트 경험이 없어 처음에는 방향 잡는 것이 어려웠습니다. 하지만 조원들의 도움으로 잘 해결해 나갈 수 있었습니다. 분석을 배우는 뜻 깊은 시간이었습니다.



[조원] 김규리 : 짧은 시간이었지만 도메인 지식을 쌓는 것부터 데이터를 분석하여 결과를 도출하는 것까지 데이터 분석의 전체적인 과정을 경험해볼 수 있어서 의미 있는 시간이었습니다.



[조원] 김송일 : 분석 경험이 많지는 않지만 할 때마다 항상 아쉬움과 후회가 남는 것 같습니다. 시간이 촉박하여 빅데이터 기간 배웠던 기법들을 충분히 활용하지 못했고 여러 시행착오가 많았습니다. 하지만 고생했던 만큼 기억에 오래 남을만한 프로젝트였습니다.



[조원] 박병수 : 의료 데이터를 다루는 프로젝트라서 힘든 점이 많았습니다. 도메인 지식이 많이 필요하다 보니 정작 데이터 분석할 시간이 부족한게 정말 아쉬웠습니다. 기회가 된다면 다양한 분석 방법을 통해 더 많은 인사이트를 도출해내고 싶습니다.



[조원] 이수진 : 익숙하지 않은 분야의 도메인 지식을 쌓고 그에 대한 인사이트를 도출하는 데에 어려움이 많았습니다. 또한, 보고서를 쓰는 방법에 대해 많은 고민을 할 수 있는 기회가 되었습니다.



