



2022년 인공지능 학습용 데이터 구축 지원사업 (1차)

분야 30 만성질환 데이터

30-1 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

30-2 만성질환 관련 임상 및 생활습관 데이터

전남대학교병원 컨소시엄

목차 Contents

- Ⅰ 사업개요
- Ⅱ 만성콩팥병 및 합병증 추적관찰 임상의료 데이터
- Ⅲ 만성질환 관련 임상 및 생활습관 데이터
- Ⅳ 성과홍보 및 일자리 창출
- Ⅴ 성공적인 사업 수행을 위한 과제 조정



I 사업개요



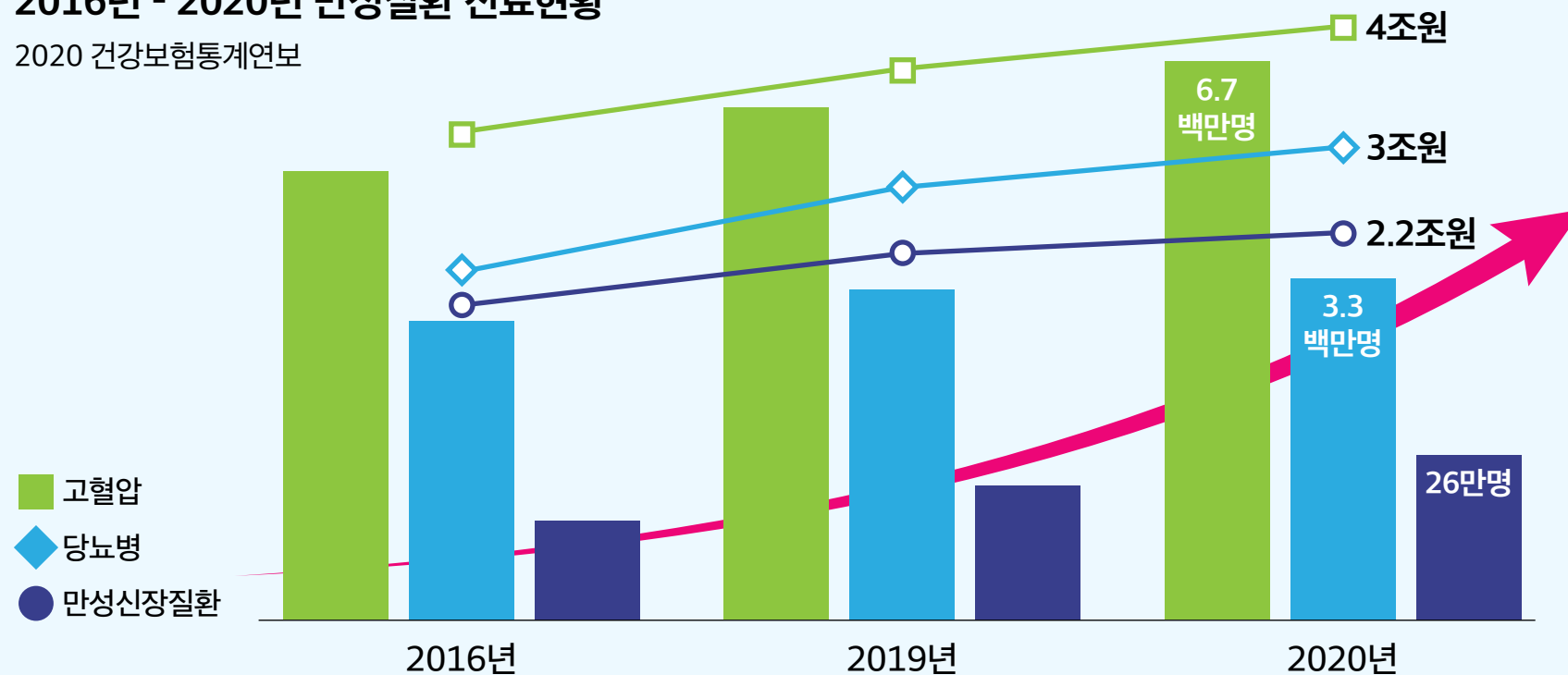
추진배경 및 필요성

급속한 고령화 및 만성질환 증가로 사회·경제적 부담 심각

만성질환자 1,891만명 / 만성질환진료비 36.2조원

2016년 - 2020년 만성질환 진료현황

2020 건강보험통계연보



I. 사업개요

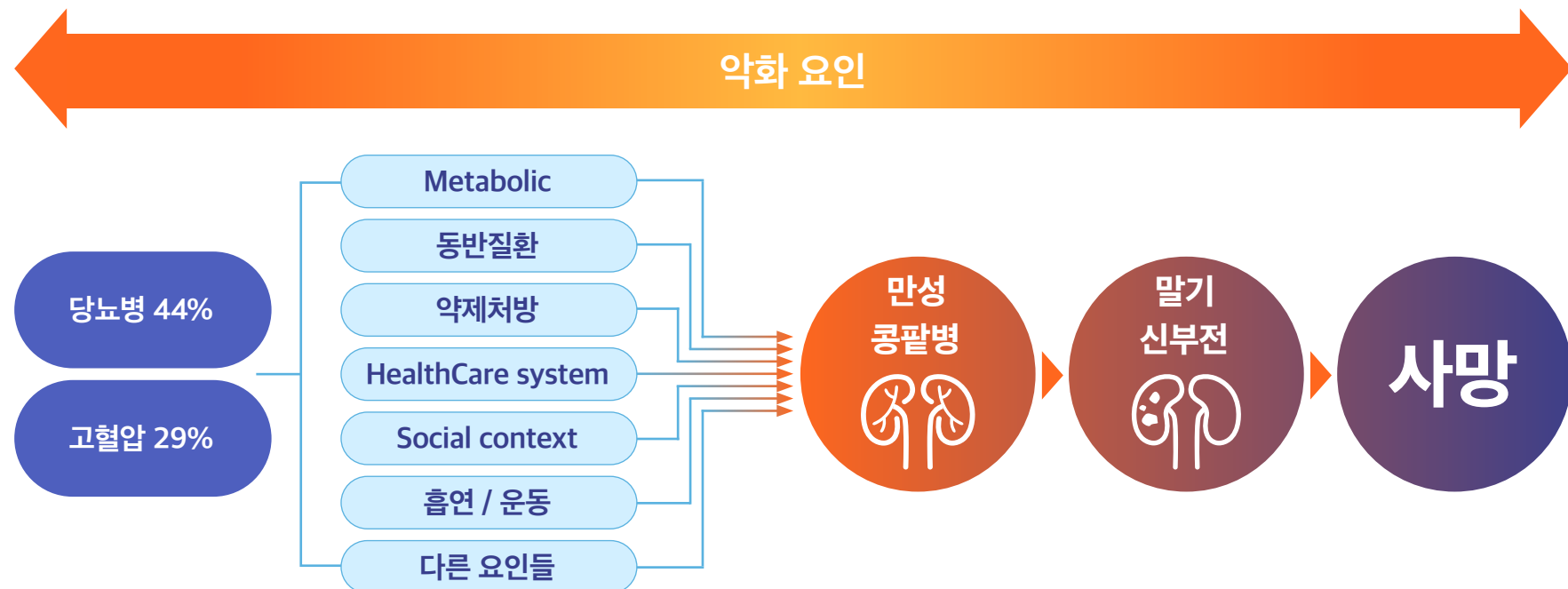
만성콩팥병과 당뇨/고혈압

당뇨병 및 고혈압에 의한 CKD 코호트 구축 및 위험인자분석

2009-2010: 건강검진데이터/국민건강보험, 심평원
고혈압 / 당뇨병 / 고혈압-당뇨 모두

CKD group (eGFR <60 ml/min/1.73 m²)

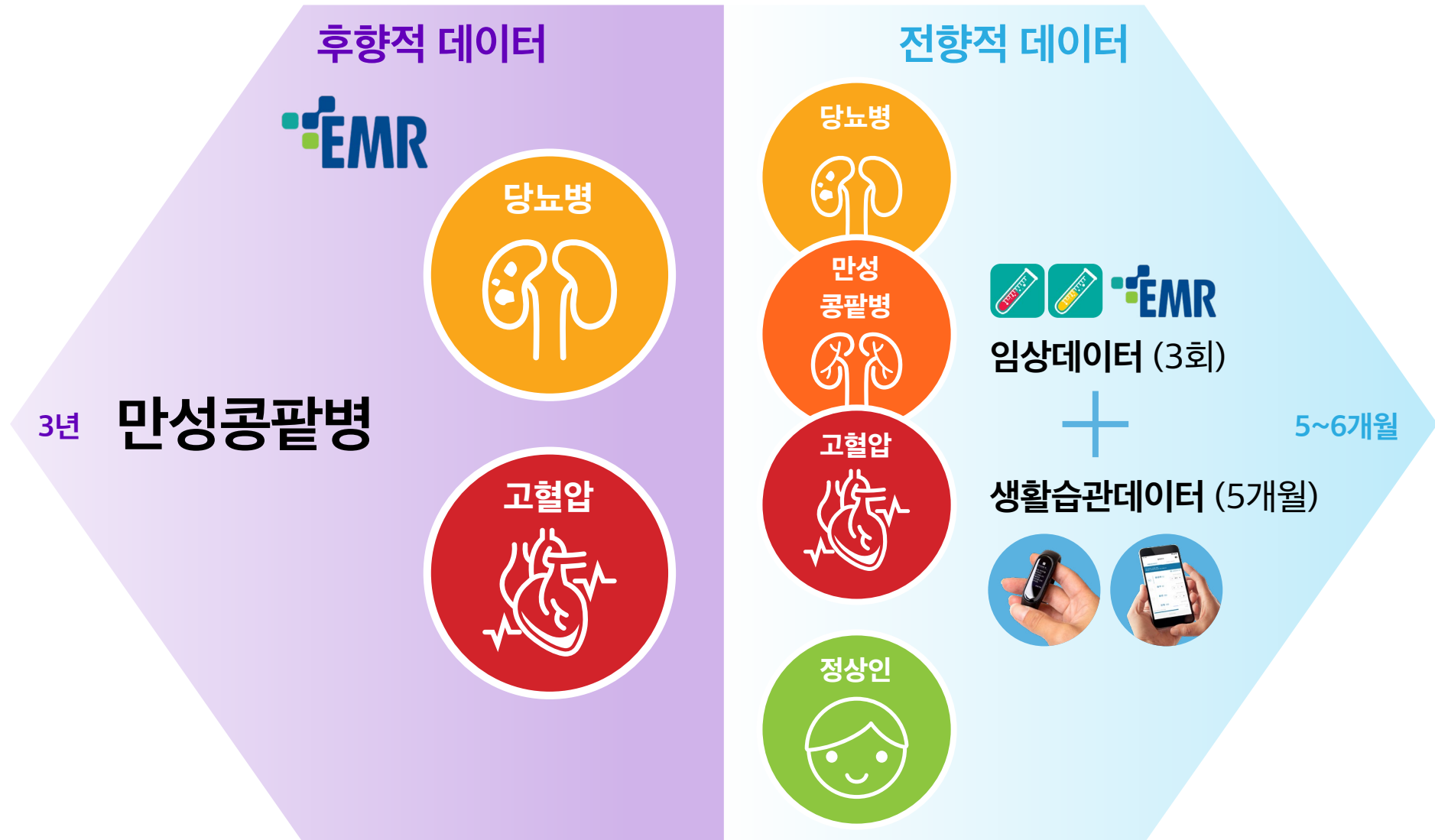
Control group (eGFR ≥60 ml/min/1.73 m²)



만성신장질환의 발생과 악화를 제어하기 위한 예측 및 예방 프로그램 개발 (전남대학교병원, 2018~20)

I. 사업개요

구축 데이터 특성



I. 사업개요

관련 동향

단일 요인이 만성질환에 미치는 영향 연구 다양한 요인들의 교호작용 반영 X

RESEARCH ARTICLE

Open Access

Effects of exercise training on proteinuria in adult patients with chronic kidney disease: a systematic review and meta-analysis

Lei Yang^{1†}, Xiaoxia Wu^{1†}, Ying Wang¹, Chunfeng Wang¹, Rong Hu^{1*} and Yong Wu^{2*}



EPIDEMIOLOGY

Association between Physical Activity and Kidney Function: National Health and Nutrition Examination Survey

INS¹, MARY ANN SEVICK², CAROLINE R. RICHARDSON^{3,4}, LINDA F. FRIED²,
K³, and ANDREA M. KRISKA¹

gh, Department of Epidemiology, Pittsburgh, PA; ²VA Pittsburgh Healthcare Systems, Pittsburgh, PA;
n Health Systems, Ann Arbor, MI; ⁴HSR&D VA Center for Clinical Management Research, Ann Arbor
d Center, Ann Arbor, MI; and ⁵University of Pittsburgh, Department of Biostatistics, Pittsburgh, PA

The effects of short sleep duration on proteinuria and chronic kidney disease: a systematic review and meta-analysis

Wisit Cheungpasitporn^{1,*}, Charat Thongprayoon^{1,*}, Maria L. Gonzalez-Suarez¹, Narat Srivali²,
Patompong Ungprasert³, Wonngarm Kittanamongkolchai¹, Sean M. Caples² and Stephen B. Granger¹

¹Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN 55905, USA, ²Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN, USA and ³Department of Rheumatology, Mayo Clinic, Rochester, MN, USA

Correspondence and offprint requests to: Wisit Cheungpasitporn; E-mail: wcheungpasitporn@gmail.com

*These authors contributed equally to this manuscript.

ORIGINAL INVESTIGATION

Open Access

Exercise-induced albuminuria increases over time in individuals with impaired glucose metabolism

Rafael Y. Brzezinski^{1,2,3,4}, Limor Friedensohn^{1,2}, Itzhak Shapira^{1,2}, David Zeltser^{1,2}, Ori Rogowski^{1,2},
Shlomo Berliner^{1,2}, Ayelet Grupper^{2,5} and Shani Shenhar-Tsarfaty^{1,2*}



I. 사업개요

참여병원 및 의료인력

주관기관



오태렴 교수

신장내과

- EHR 기반 의료 인공지능 연구
- 건강보험심사평가원
만성 콩팥병 전문심사위원(PRC)
- 질병특화 코호트 분석 연구
- 이공분야기초연구사업
우수신진연구

참여병원



이하린 교수

신장내과

- 만성콩팥병 환자
관리에 관한 연구
- 급성콩팥손상과
지속적신대체요법에 관한 연구
- 지속적신대체요법 이후
만성콩팥병 이행에 대한 연구

참여병원



강경표 교수

신장내과

- 전북대학교병원 의생명연구
심상연구사업단장
- 전북대학교병원
전북빅데이터센터 참여교수
- 염증성 신장질환에서
신장배액림프질의 역할 연구
수행중

참여병원



최홍상 교수

신장내과

- 임상과학자 연구역량 강화사업
- 빅데이터 분석을 통한
신장질환 위험인자 연구
- 급성 및 만성신장질환 분자기전
및 치료약물 연구

I. 사업개요

IRB 신청 및 승인 현황

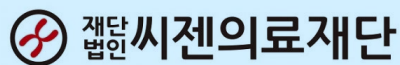
참여병원	만성콩팥병 및 합병증 추적관찰 임상의료 데이터		만성질환 관련 임상 및 생활습관 데이터	
	IRB 신청	IRB 승인	IRB 신청	IRB 승인
 전남대학교병원 CHONNAM NATIONAL UNIVERSITY HOSPITAL	○	○	○	4월 예정
 부산대학교병원 Pusan National University Hospital	○	○	○	○
 전북대학교병원 JEONBUK NATIONAL UNIVERSITY HOSPITAL	○	○	○	4월 예정
 화순전남대학교병원 CHONNAM NATIONAL UNIVERSITY HWASUN HOSPITAL	○	○	○	○

I. 사업개요

컨소시엄 구성방안

참여병원

지역거점 병원
만성콩팥병 데이터 수집
만성질환/정상인 대상자 모집
임상데이터 수집, 정제, 라벨링

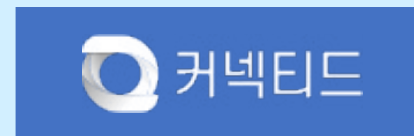


도구 개발

임상데이터 정제/라벨링 도구



생활습관데이터 수집 어플리케이션



품질 관리

ELUONINS

외부 자문 의료진

학습모델



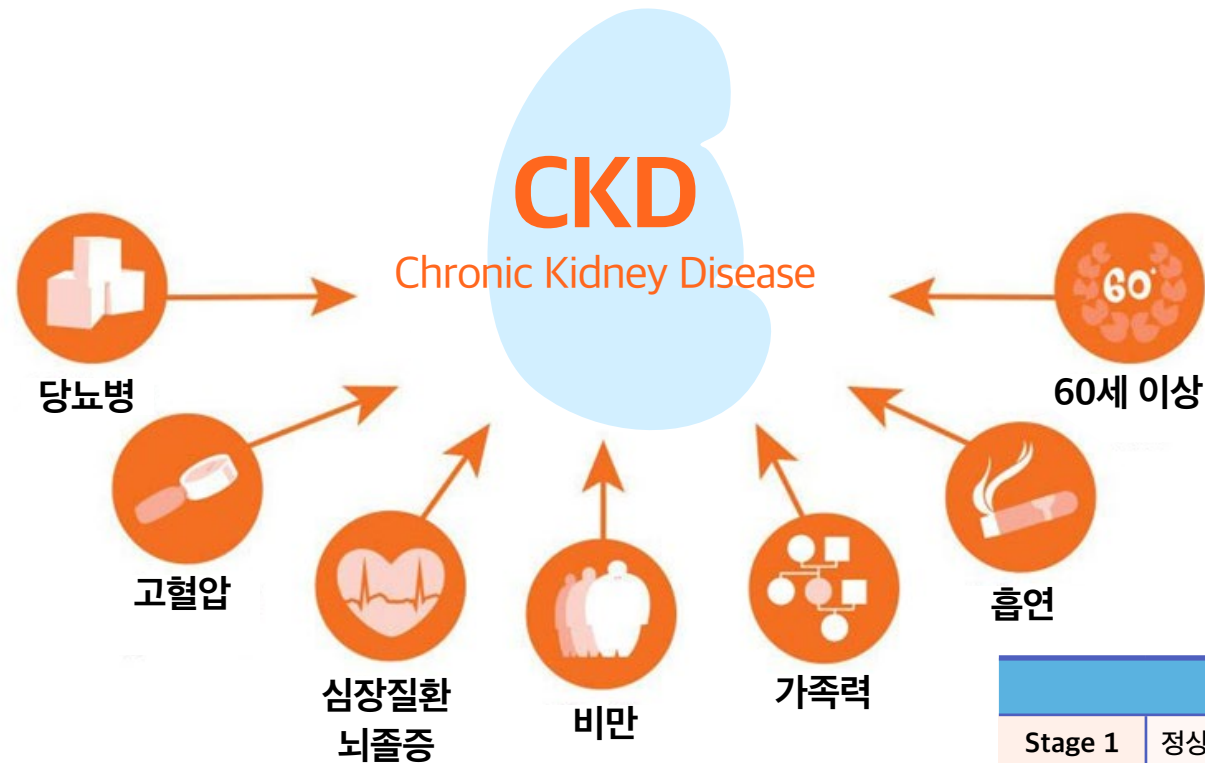
ELUONINS

Ⅲ 만성콩팥병 및 합병증 추적관찰 임상의료 데이터



II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

만성콩팥병의 복잡한 기전



만성콩팥병 진행 단계		GFR*
Stage 1	정상적인 신장 기능을 가진 신장 손상	≥ 90
Stage 2	경미한 신장 기능 상실을 동반한 신장 손상	89 - 60
Stage 3a	경증에서 중등도의 신장 기능 상실	59 - 45
Stage 3b	중등도에서 중증의 신장 기능 상실	44 - 30
Stage 4	신장 기능의 심각한 손실	29 - 15
Stage 5	콩팥 기능 상실	< 15

* GFR 수치는 신장 기능이 어느 정도인지 알려주는 수치임

II. 만성콩팥병 및 합병증 추적관찰 임상의로 데이터

만성질환 데이터

구축 목적

만성콩팥병 및 합병증 추적관찰 임상의로 데이터 구축

항목	원시데이터	원천데이터	라벨링데이터
대상자 수	10,000명	8천 명 이상	8천 명 이상
데이터셋	108,500건	100,000건 이상	100,000건 이상
Domain 개수	7개	7개	8개
Item 개수	264개	264개	287개 (원천264, 라벨링23)

질병 관련 합병증의 예측 / 조기 중재를 가능하게 하는 인공지능 모델 개발

만성콩팥병 환자 입원
(cause-specific
hospitalization)
예측 모델링

심혈관계 합병증
(cardiovascular event)
발생예측 모델링

신장기능
(사구체 여과율, eGFR)
변화 예측 모델 개발

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터 데이터 구축 프로세스

원시데이터 수집



II. 만성콩팥병 및 합병증 추적관찰 임상의로 데이터

원시데이터 수집

참여의료기관
의무기록 추출

익명화/비식별화

데이터 표준화

결측치/이상치
관리

데이터 라벨링

1년 이상 추적 관찰된 1만 명 이상 환자

참여병원	등록가능 대상자 수	예상 등록 대상	추정횟수	성별	연령대	병기
전남대학교병원	25,000명 이상	3,000명	3~4회	여성 3천 명 이상	20~39세 8백명 이상 40~59세 2천명 이상 60~75세 1.6천명 이상	4단계 이하 6천명 이하
화순전남대학교병원	13,000명 이상	2,000명				
부산대학교병원	15,000명 이상	3,000명				
전북대학교병원	20,000명 이상	2,000명				

Domain	설명	ITEM 개수	포맷
개인정보	생년월일, 성별, 거주지, 과거병력, 가족력	56개	CSV JSON
신체계측데이터	키, 몸무게, 허리둘레, 엉덩이둘레, 허리/엉덩이 비율, 체질량지수, 수축기/이완기 혈압, 맥박	12개	
혈액검사	해당 환자의 혈액검사 결과	33개	
소변검사	해당 환자의 소변검사 결과	16개	
약제처방	해당 환자의 처방된 약물정보	132개	
영상검사 결과	심전도, 엑스레이, 전산화단층검사, 자기공명영상 검사, 초음파 검사를 포함한 영상검사들의 검사결과 및 판독결과지	15개	
예후 데이터	예측하고자 하는 주요 질환관련 예후 사건의 발생 정보	11개	

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

만성질환 데이터

원천데이터 - 데이터 정제

참여의료기관
의무기록 추출

익명화/비식별화

데이터 표준화

결측치/이상치
관리

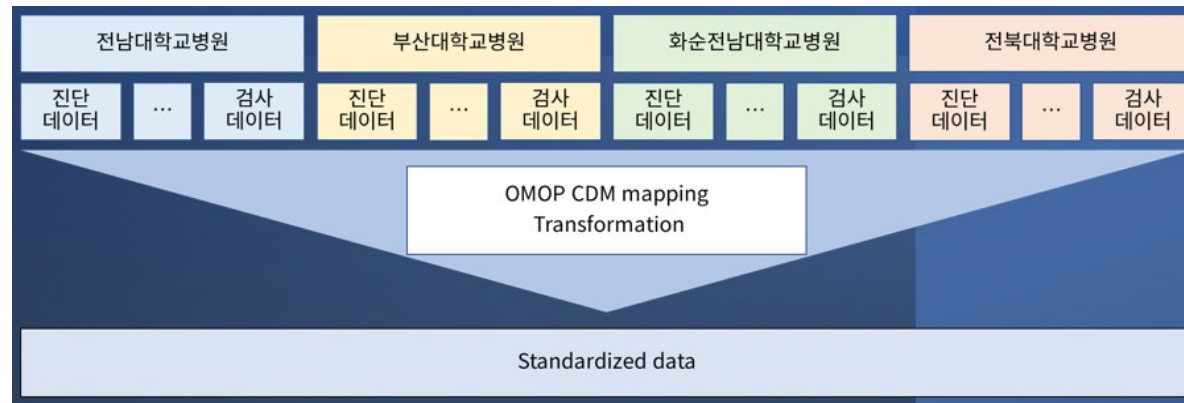
데이터 라벨링

원천데이터 규모 108,500 건 (RFP 100,000건 이상)

익명화 / 비식별화 처리

환자등록번호 및 개인정보 / 생년월일 / 검사일자

로컬데이터 표준화



SNOMED 용어기반의 데이터 표준화 작업

전남대병원, 화순전남대병원, 부산대병원은 CDM 구축사업을 통해 용어 표준화 완료

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

만성질환 데이터

원천데이터 - 데이터 정제

참여의료기관
의무기록 추출

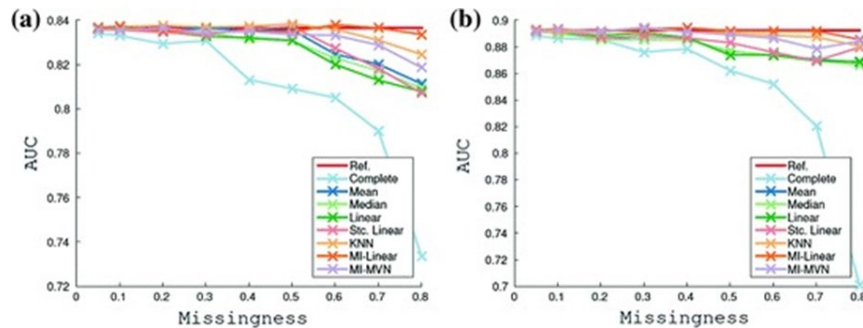
익명화/비식별화

데이터 표준화

결측치/이상치
관리

데이터 라벨링

결측값 처리



삭제법

평균대치법

중앙값대치법

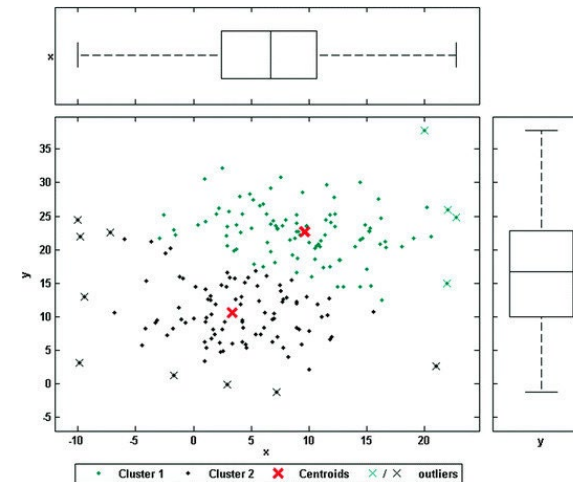
선형회귀대치법

다중대치법

단일값대치법

원본데이터 형태를 최대한 유지하는
결측값 처리 방식을 우선적으로 적용

이상치(outlier) 처리



Z-score

Modified
Z-scoreMahalanobis
distance

log-normal IQR

k-Means

k-Medoids

전문지식이 있는 의료진의 참여를 통하여 이상치
특성파악 및 최선의 이상치 처리방식을 선택

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

만성질환 데이터

원천데이터 - 데이터 정제

참여의료기관
의무기록 추출

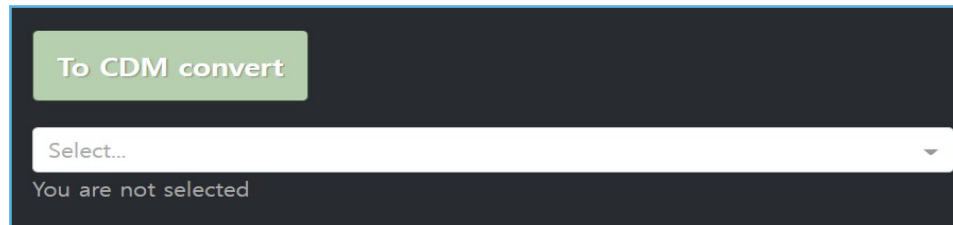
익명화/비식별화

데이터 표준화

결측치/이상치
관리

데이터 라벨링

데이터 정제 도구



To CDM convert [완료_] [재검토]

File : 전남대병원.xls

등록번호	외래일자	검사일자	검사코드	검사명	검사
1 030XXXX48	===Masking===	===Masking===	Local_XXXX		
2 030XXXX49	===Masking===	===Masking===	Local_XXXX	Creatinine	6
3 030XXXX50	===Masking===	===Masking===	Local_XXXX	Protein (Urine)	4
4 030XXXX51	===Masking===	===Masking===	Local_XXXX	Protein/Creatinine ratio	708
5 030XXXX52	===Masking===	===Masking===	Local_XXXX	T. Protein	
6 030XXXX53	===Masking===	===Masking===	Local_XXXX	Phosphorus	
7 030XXXX54	===Masking===	===Masking===	Local_XXXX	Calcium	
8 030XXXX55	===Masking===	===Masking===	Local_XXXX	Uric Acid	
9 030XXXX56	===Masking===	===Masking===	Local_XXXX	GFR(CKD-EPI)	
10 030XXXX57	===Masking===	===Masking===	Local_XXXX	GFR(MDRD)	

원시데이터 CDM 모형 변환

순서보존형 암호화 처리

로컬코드

File : CDM_전남대병원.xls

UID	외래일자	검사일자	OMOP-CDM	검사명	검사
2 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Protein/Creatinine ratio	80
3 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Creatinine	6
4 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Protein (Urine)	4
5 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Protein/Creatinine ratio	708
6 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	T. Protein	
7 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Phosphorus	
8 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Calcium	
9 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	Uric Acid	
10 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	GFR(CKD-EPI)	
11 Pat_XXXXXX	===Masking===	===Masking===	CDM_XXXXX	GFR(MDRD)	

공통코드

공통코드

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

만성질환 데이터

라벨링데이터 - 데이터 가공

참여의료기관
의무기록 추출

익명화/비식별화

데이터 표준화

결측치/이상치
관리

데이터 라벨링

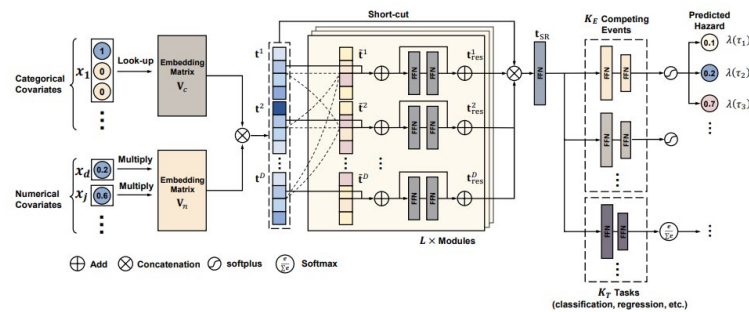


모델링에서 지도 학습을 활용하기 위해서는
 정확한 의료 사건 발생 및 관련 질환에 대한 검증이 필요함
**충분한 의료 도메인 (medical domain) 지식이 있는
 의료진의 참여가 필수**

II. 만성콩팥병 및 합병증 추적관찰 임상의로 데이터 인공지능 학습모델

합병증 예측 모형

revascularization, cause-specific hospitalization 예측모형



Transformer

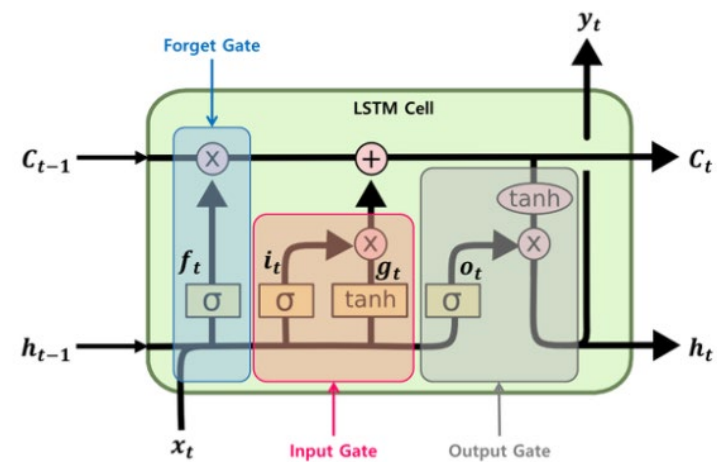
DeepSurv

DeepHit

Deep Survival
Machines

중도절단 자료의 분석 시 sensitivity 나
specificity 중 어느 한쪽에 치우친
평가지표 (evaluation metrics)를 사용 시
실질적으로 임상 현장의 활용도가 낮아짐

사구체 여과율 변화 예측 모형



LSTM (Long Short Term Memory)

사구체여과율 예측은
분류(classification) 문제로 볼 수 있는,
합병증 발생예측과는 전혀 다른 problem set으로,
연속형 결과변수에 대한 예측이 필요함

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터 인공지능 학습모델 품질지표

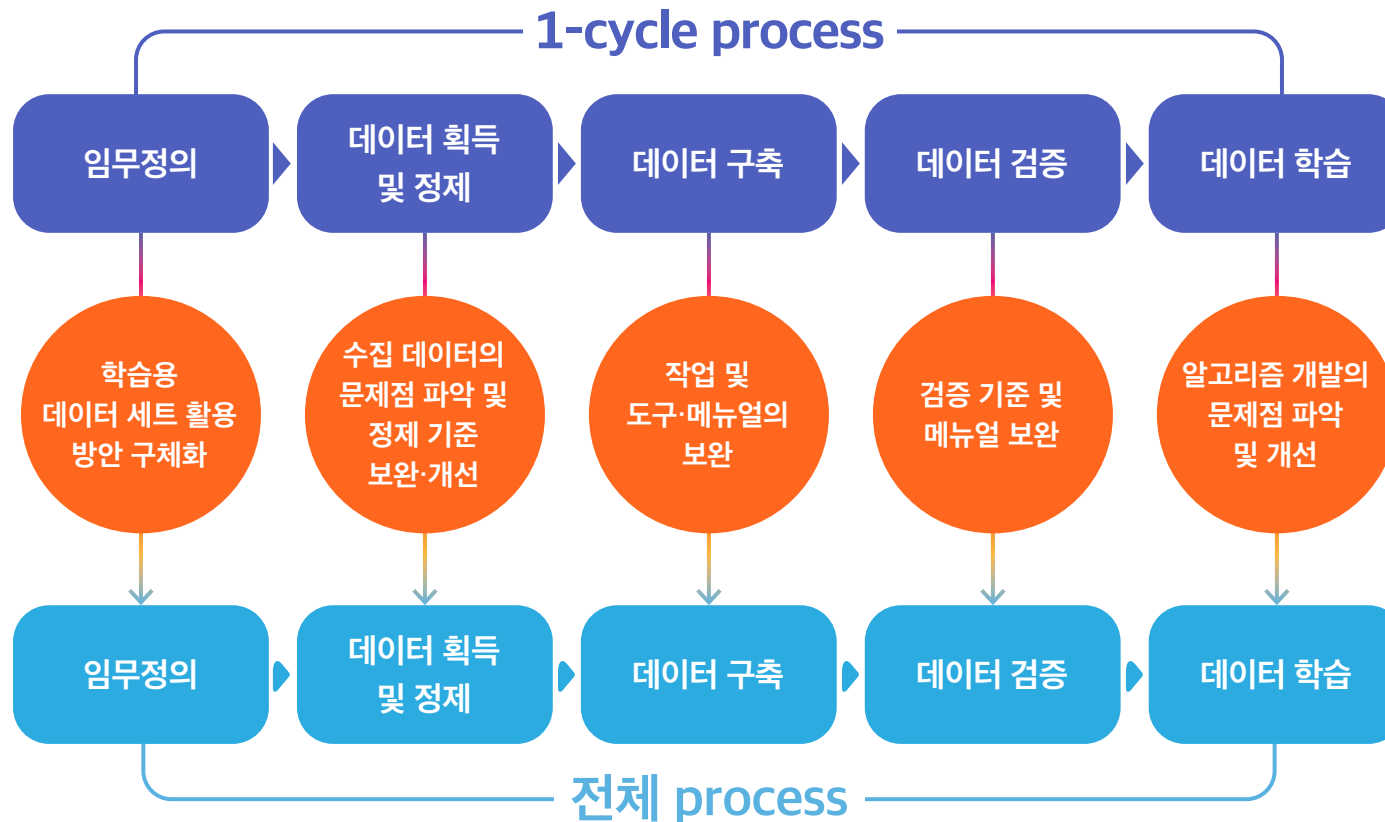
만성질환 데이터

합병증 예측 모형	
품질지표	C-index, Brier score, AUROC
선행연구	<p>DeepHit : A Deep Learning Approach to Survival Analysis with Competing Risks.</p> <p>DeepHit : C-index 0.68 기록</p> <ul style="list-style-type: none"> • 하지만 해당 선행연구와 개발하고자 하는 모델의 입력변수가 상이하여 모델간 직접적인 성능비교는 어려울 것으로 보임 • 제안된 모델을 적용 후 다른 모델에도 적용하여 결과를 비교분석 하여 성능을 측정함 • 선행연구에서 도출한 최적 모델의 성능지표는 c-index 기준 약 0.68정도 이므로 해당 수치를 고려하여 판단하도록 함
지표기준	<ul style="list-style-type: none"> • C-index 0.7 이상 혹은 Brier score 0.2 이하 혹은 AUROC 0.7 이상 (목적에 맞춰 평가 지표는 조정하여 적용)

사구체 여과율 변화 예측 모형	
품질지표	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$ <p>y_i: 실제 데이터 \hat{y}_i: 예측 데이터</p>
선행연구	<ul style="list-style-type: none"> • LSTM을 활용하여 만성 콩팥병 환자에서 사구체 여과율의 예측 모형을 구축 • 최적의 Layer Unit, Dens Unit를 설정하고 패널 데이터를 적용 할 수 있는 모델들과 비교 • 해당 선행연구와 개발하고자 하는 모델의 입력변수가 상이하여 모델간 직접적인 성능비교는 어려울 것으로 보임 • 제안된 모델을 적용 후 다른 모델에도 적용하여 결과를 비교분석 하여 성능을 측정함 • 선행연구에서 도출한 모델의 성능지표는 RMSE 기준 약 20정도 이므로 해당 수치를 고려하여 판단하도록 함
지표기준	<ul style="list-style-type: none"> • RMSE 기준 20이하 (목적에 맞춰 평가 지표는 조정하여 적용)

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

1-Cycle 수행 (만성콩팥병, 만성질환 임상&생활습관)



1-Cycle 수행 시기 및 산출물

시기	수행내용	산출물
IRB 승인 이후	원시데이터를 수집	수집 가이드라인
4월	수집된 원시데이터 정제	정제 가이드라인
4월	정제된 원천데이터 가공	가공 가이드라인
5월	가공된 라벨링데이터 학습모델 적용 및 결과 도출	학습모델 유효성 테스트 결과서

1-Cycle 대상 데이터 수량 : 1,500건

II. 만성콩팥병 및 합병증 추적관찰 임상의로 데이터

데이터 품질관리

구분	지표	품질목표	품질목표 달성기준
구축 공정	준비성	95% 이상	준비성 체크리스트 목록의 95% 이상 '적정'
	완전성		완전성 체크리스트 목록의 95% 이상 '적정'
	유용성		준비성 체크리스트 목록의 95% 이상 '적정'
데이터 적합성	기준 적합성	95% 이상	구축데이터 품질 기준 적합성 체크리스트 목록의 95% 이상 '적정'
	기술 적합성	99%	파일 포맷 준수율 99% 이상
	통계적 다양성	여자 : 37.5% 이상	성별 다양성을 확보하기 위하여, 구축 목표 데이터 8,000건 기준 여성 임상 데이터 3000건 (37.5%) 이상 확보
		20~39세 : 10% 이상 40~59세 : 25% 이상 60~75세 : 20% 이상	연령 다양성을 확보하기 위하여, 구축 목표 데이터 8,000건 기준 20~39세 환자 800명(10%) 이상, 40~59세 2000명(25%) 이상, 60~75세 1600명 (20%) 이상 확보
		병기 1~4 단계 : 75% 이상	만성콩팥 병기 다양성을 확보하기 위하여, 구축 목표 데이터 8,000건 기준 1~4단계 환자 6000명 (75%) 이상 확보
데이터 정확성	의미 정확성	99% 이상	사업 참여/비 참여 의료진 교차검증 합의도 99%
	구문 정확성	99% 이상	라벨링 데이터 구조, 형식 오류율
학습모델	알고리즘 적정성	Pass	알고리즘 적정성 과제심의 통과
	유효성	-	심혈관계합병증 발생 예측 모형 C-index 0.7 이상 혹은 Brier score 0.2 이하 혹은 AUROC 0.7 이상
		20 이하	사구체 여과율 변화 예측 모형 RMSE 20 이하

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터

품질관리 도구 (공통)

품질검사영역	구문 정확성 검사	의미 정확성 검사	학습모델 유효성 검사
도구명	file format 검사 도구	교차검증 분배 도구	학습모델 검사 도구
설명	수집 데이터 file format 자동 전수 검사 도구	라벨링 데이터 교차검증을 위한 분배 도구	인공지능 모델을 훈련용 데이터에 대해 학습하고, 시험용 데이터의 성능을 기반으로 학습 모델 유효성 검사
도구유형	오픈소스 기반 자체 개발		
주요기능	<ul style="list-style-type: none"> 데이터 구조, 속성 확인 데이터 구조, 속성 등의 변경으로 수정이 필요할 경우 전문 개발자를 통해 해당하는 부분만 작업이 진행될 수 있는 도구를 개발하거나 수정하여 구축된 대량의 데이터에 대한 검수 진행 	<ul style="list-style-type: none"> 수집된 원시, 원천, 라벨링 데이터의 교차검증을 위한 분배 기능 	<ul style="list-style-type: none"> 구축된 데이터를 학습 및 평가할 모델의 품질관리 수행 학습용 데이터셋으로 학습한 모델을 검증용 데이터셋에 대해 테스트한 결과를 평가지표를 통해 정량적으로 평가
사용환경	<ul style="list-style-type: none"> OS: Win10 H/W: x86 PC S/W: Python 	<ul style="list-style-type: none"> OS: Win10 H/W: x86 PC S/W: Python 	<ul style="list-style-type: none"> OS: Ubuntu H/W: A100 S/W: Python

II. 만성콩팥병 및 합병증 추적관찰 임상의료 데이터 사업추진일정

세부과제명	수행내용	추진일정 (월)								비중 (%)
		4	5	6	7	8	9	10	11	
만성 콩팥병 및 합병증 추적관찰 임상의료 데이터	사전작업 - IRB 승인 - 정제, 라벨링 저작도구 개발 설계 - 검증도구 개발 설계									10
	착수보고									5
	참여병원 담당자(수집~가공) 교육									5
	원시데이터 샘플 수집									10
	알고리즘 개발 및 검증									5
	데이터 수집									35
	데이터 정제									
	데이터 라벨링									
	라벨링 데이터 검수									10
	인공지능 모델 개발									10
	AI 경진대회 개최									5
	학습데이터 활용방안 연구 및 개발									5
	완료보고									

III 만성질환 관련 임상 및 생활습관 데이터



III. 만성질환 관련 임상 및 생활습관 데이터

만성질환 관리

대표적인 만성질환 예방과 관리를 위한 9대 생활수칙 질병관리청

2020.3.12. 질병관리본부 KCDC

만성공팔병 예방과 관리를 위한 9대 생활수칙

- 음식은 싱겁게 먹고 단백질 섭취는 가급적 줄입니다.
- 칼륨이 많은 과일과 채소의 지나친 섭취를 피합니다.
- 공팔의 상태에 따라 수분을 적절히 섭취합니다.
- 담배는 반드시 끊고 술은 하루에 한두 잔 이하로 줄입니다.
- 적정 체중을 유지합니다.
- 주 3일 이상 30분에서 1시간 정도 적절한 운동을 합니다.
- 고혈압과 당뇨병을 꾸준히 치료합니다.
- 정기적으로 소변 단백질과 혈액 크레아티닌 검사를 합니다.
- 꼭 필요한 약을 공팔 기능에 맞게 복용합니다.

7/8

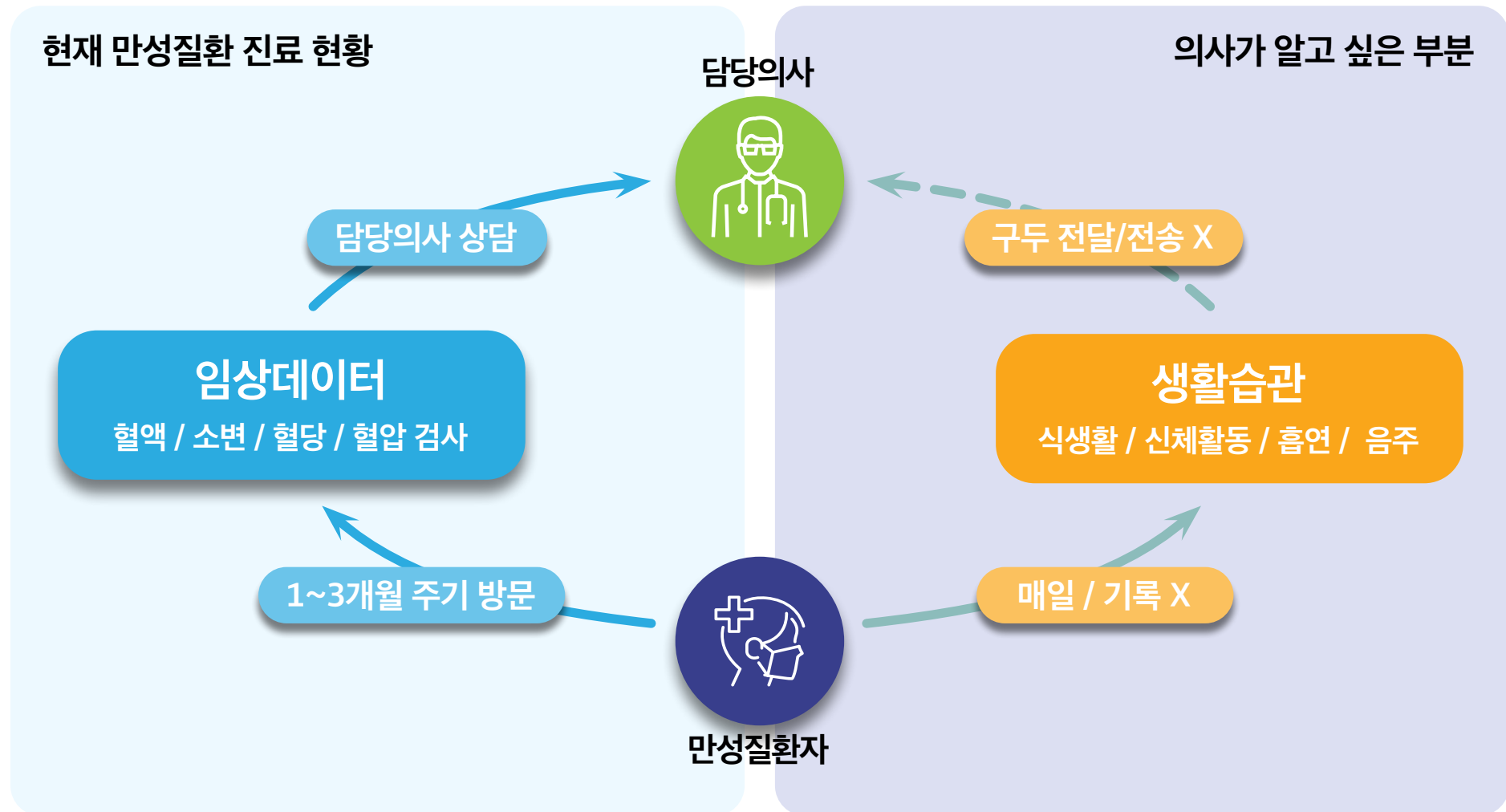


III. 만성질환 관련 임상 및 생활습관 데이터

만성질환 진료에 필요한 데이터

병원

일상생활



III. 만성질환 관련 임상 및 생활습관 데이터

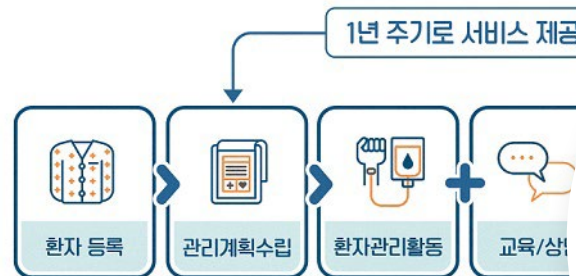
만성질환 데이터

만성질환 관리 동향

인공지능 데이터 수집 X 인공지능 활용 X

만성질환 관리 (당뇨병 / 고혈압)

일차의료 만성질환관리 사업이란?



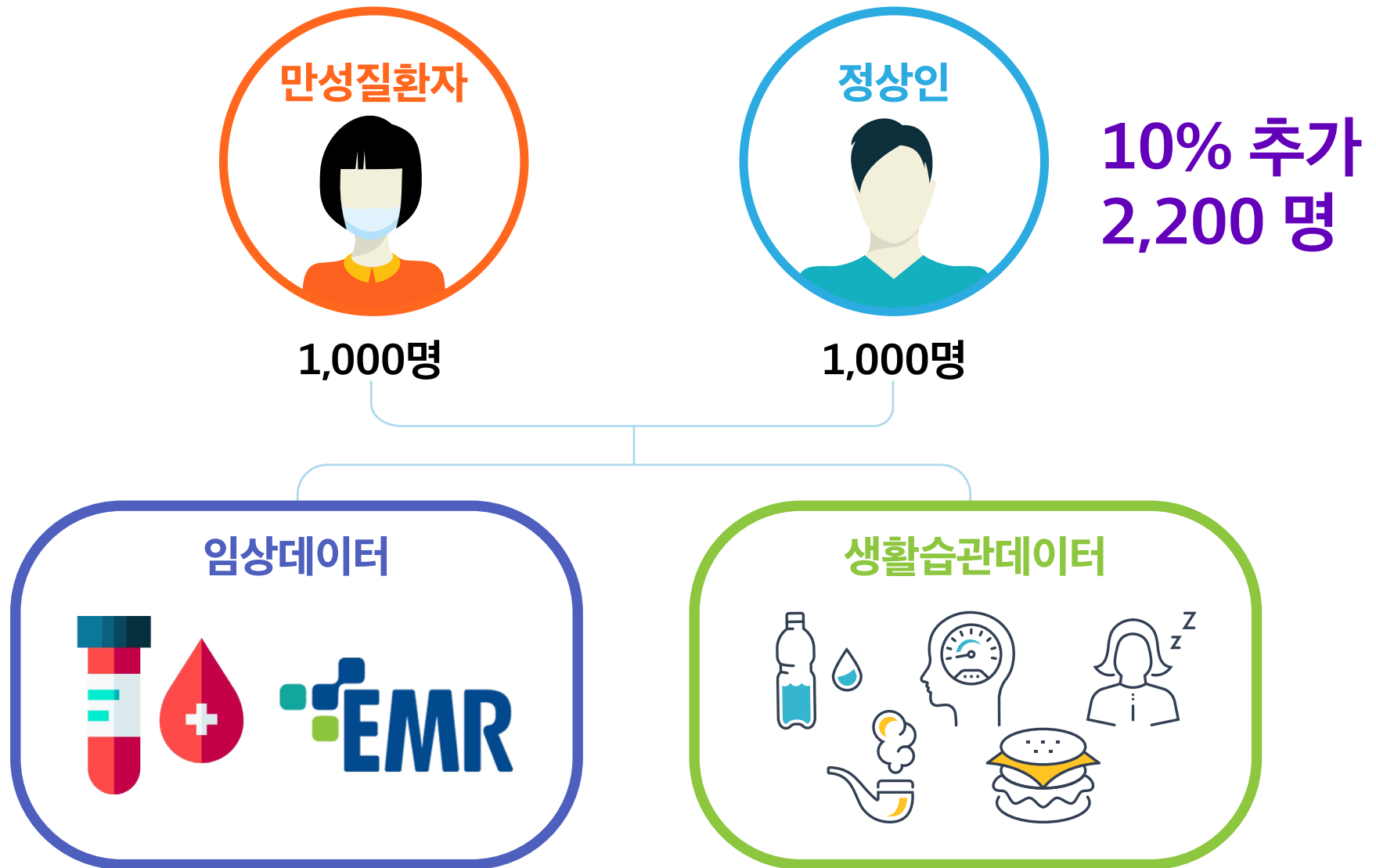
동네의원이 고혈압·당뇨병 환자에게
합병증 예방을 위해 **건강관리 서비스를**
제공하는 사업입니다.

일차의료만성질환관리
시범사업 개선을 위해
인공지능을 활용한 모니터링과
환자 피드백 제공,
케어코디네이터 공동 활용 등이
제안됐다.

예방적 건강관리를 위한 사업



III. 만성질환 관련 임상 및 생활습관 데이터 데이터 구축 내용



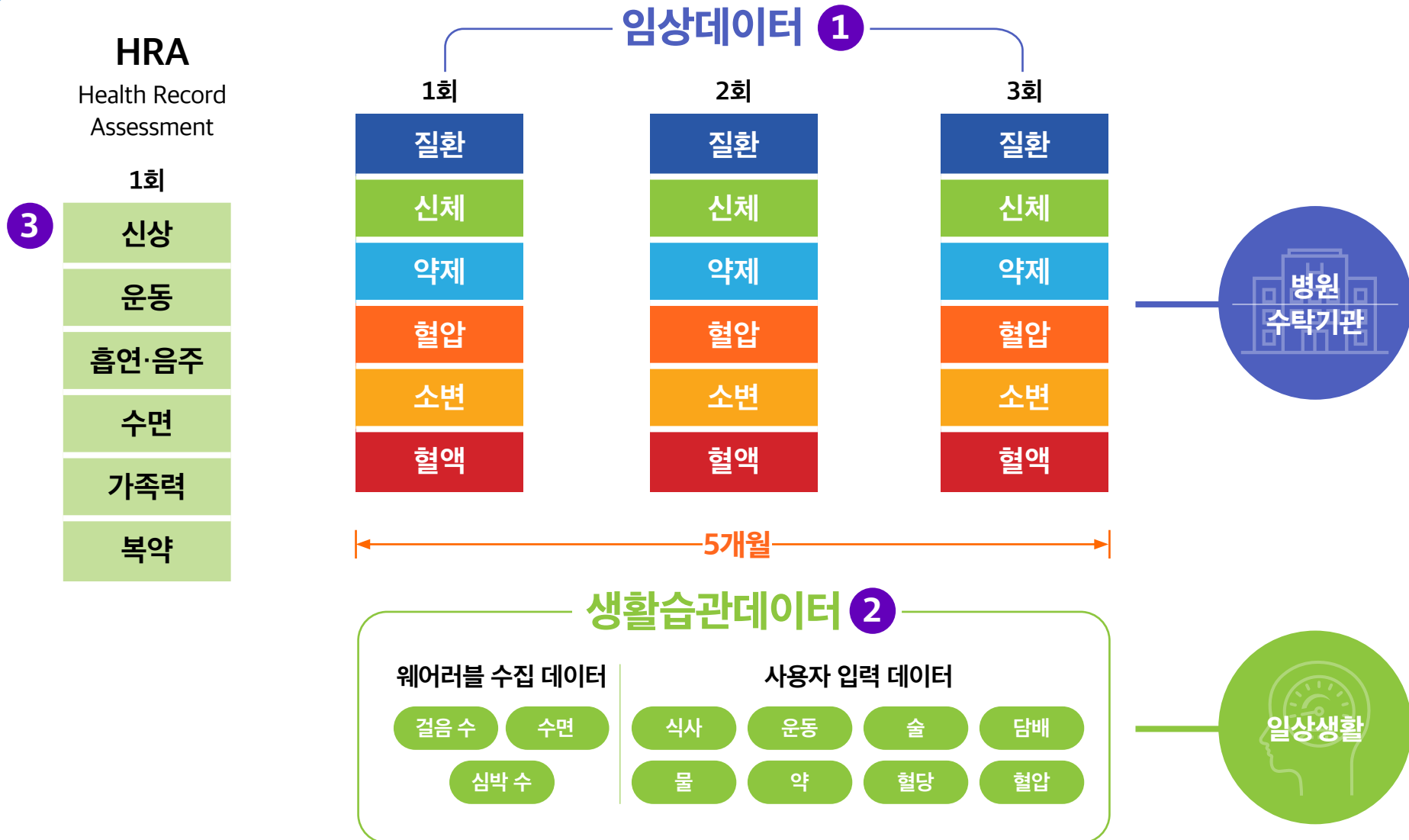
III. 만성질환 관련 임상 및 생활습관 데이터

데이터 수집 대상자 기준

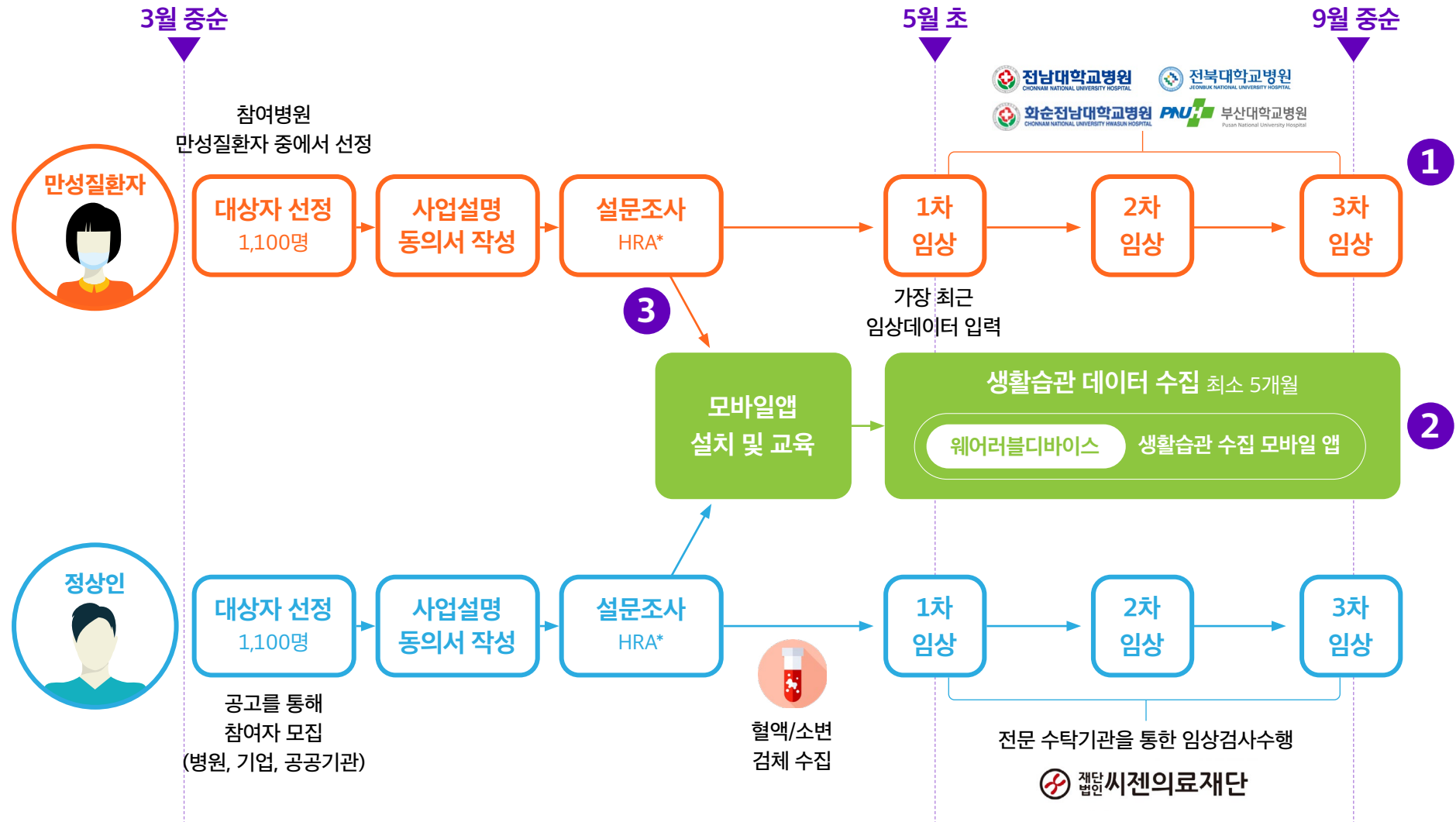
구분	선발기준		비고
만성질환자 1,100명	당뇨병	1. 당화혈색소 > 6.5 이상 2. 공복혈당>126mg/dl 3. 무작위 시점의 혈당>200mg/dl 4. 포도당 75g을 섭취후 2시간후 혈당이 200mg/dL 이상 5. 당뇨약을 복용중인 경우 위 5가지 기준 중 하나 이상에 해당되는 경우	300명 이상
	고혈압	1. 수축기 혈압 > 130 mmHg 혹은 이완기 혈압 > 80 mmHg 2. 혈압약을 복용중인 경우 위의 2가지 기준 중 하나 이상에 해당되는 경우	200명 이상
	만성콩팥병	1. 3개월이상의 신장기능 구조적/기능적 이상이 확인된 경우 2. 3개월 이상 사구체여과율의 저하(eGFR<60ml/min/1.73m2)가 확인된 경우 위의 2가지 기준 중 하나 이상에 해당되는 경우	200명 이상
정상인 1,100명	당뇨병, 고혈압, 만성콩팥병에 해당되지 않는 대상자		1,100명

III. 만성질환 관련 임상 및 생활습관 데이터

데이터 구조



III. 만성질환 관련 임상 및 생활습관 데이터 데이터 구축 프로세스



* HRA : Health Risk Assessment

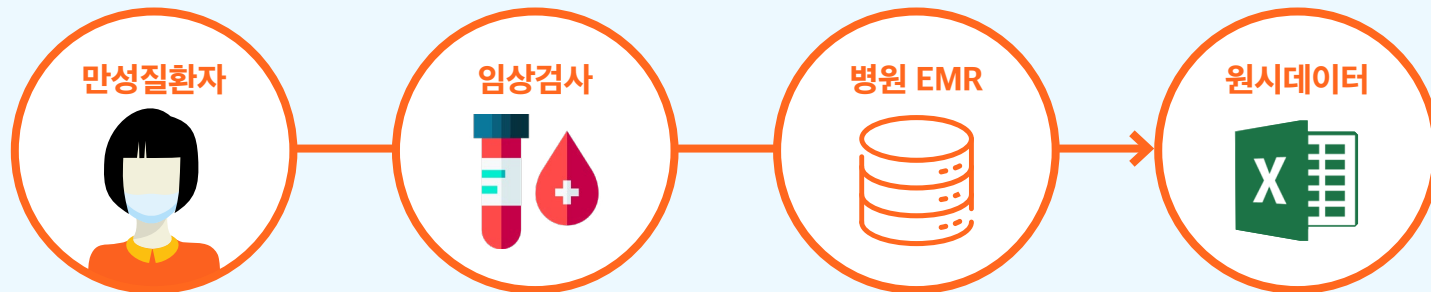
III. 만성질환 관련 임상 및 생활습관 데이터

임상데이터 수집

의료임상데이터 정제, 가공 절차와 방법은
만성콩팥병과 동일

만성질환자 임상데이터 수집

4개 참여병원 내부



정상인 임상데이터 수집



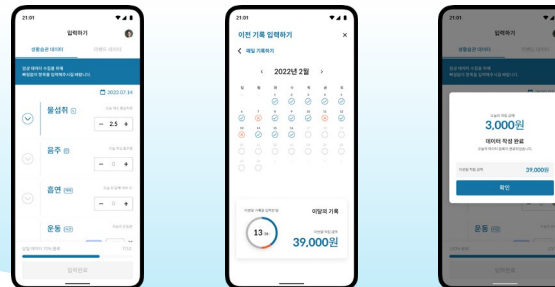
III. 만성질환 관련 임상 및 생활습관 데이터

생활습관 데이터 획득

웨어러블 기기를 사용한 **자동수집 데이터**



모바일 앱을 이용한 **사용자입력 데이터**



클라우드

III. 만성질환 관련 임상 및 생활습관 데이터 사용자 순응도 Compliance



한 개인의 행동, 즉 약물 복용, 생활 습관, 병원 방문 등의 행위가
얼마나 의사 또는 의료진의 권고와 일치하는가 하는 정도를 의미



III. 만성질환 관련 임상 및 생활습관 데이터

웨어러블 기기 선정

○ 좋음 ◐ 보통 ● 나쁨

웨어러블 기기	타입	주요기능	편의성	가격 (만)	밴더 종속성	데이터 연계	대상자 사례비
	반지	심전도 맥동 맥박 산소포화도	○	50~100 (11~22억)	○	◐	●
	밴드	걸음수 수면 심박수	○	5~20 (1~4억)	◐	○	○
	시계	걸음수 체성분 심전도 수면	◐	30~50 (7~11억)	●	◐	●

III. 만성질환 관련 임상 및 생활습관 데이터 사용자 입력 데이터 - 모바일 앱

일일 데이터 입력창

21:01

입력하기

생활습관 데이터 마벤드 데이터

임상 데이터 수집을 위해
빠짐없이 항목을 입력해주시길 바랍니다.

2022.07.14

물섭취 오늘 마신 물섭취량

음주 오늘 마신 음주량

흡연 오늘 핀 당에 개비 수

운동 오늘의 운동량

당일 데이터 70% 완료 7/10

입력완료

월간 데이터 입력창 및
월간 적립금액

21:01

이전 기록 입력하기

< 매일 기록하기

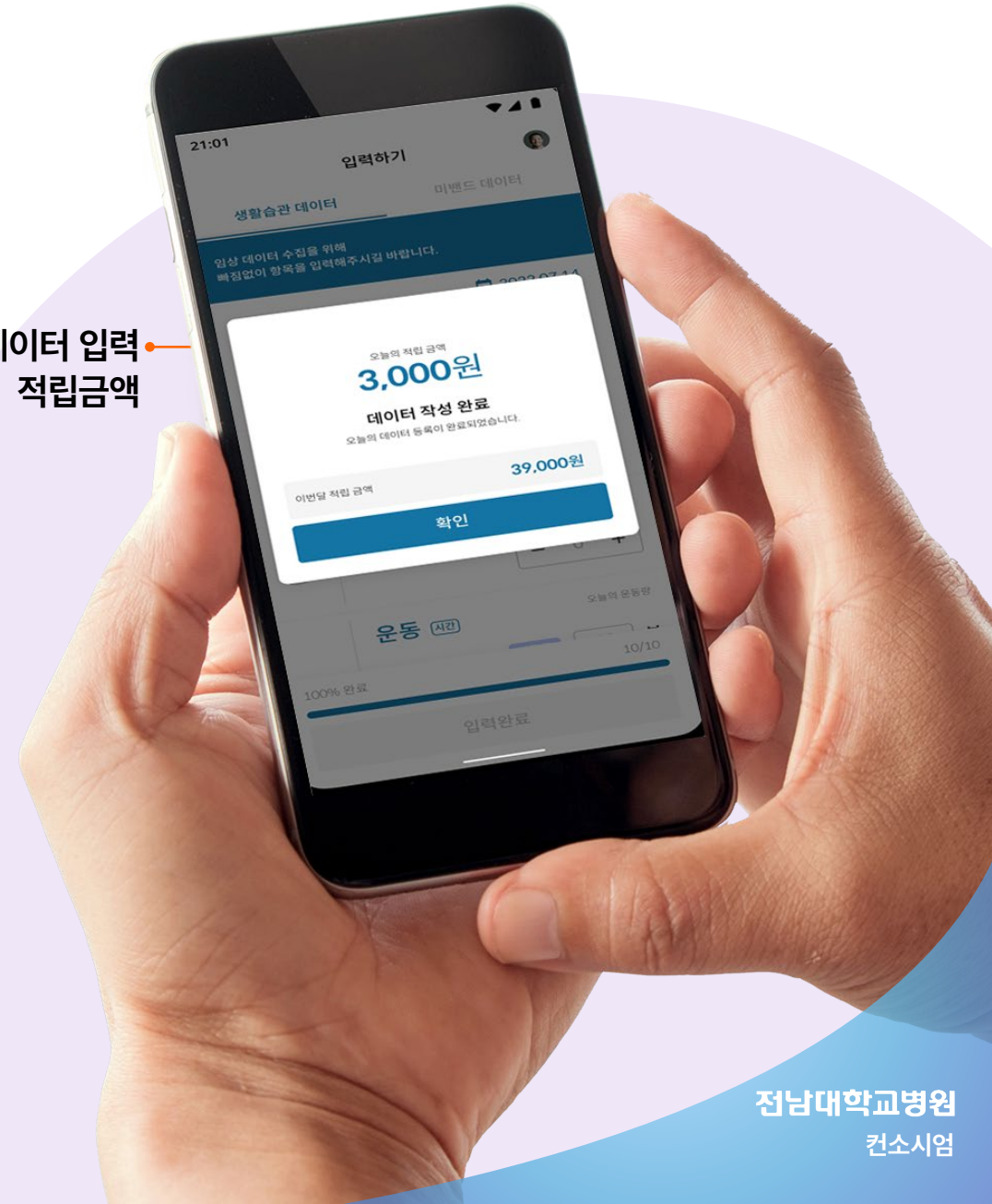
< 2022년 7월 >

일	월	화	수	목	금	토
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

이번달 기록을 입력한 날 이달의 기록

13/16 이번달 적립 금액

39,000원

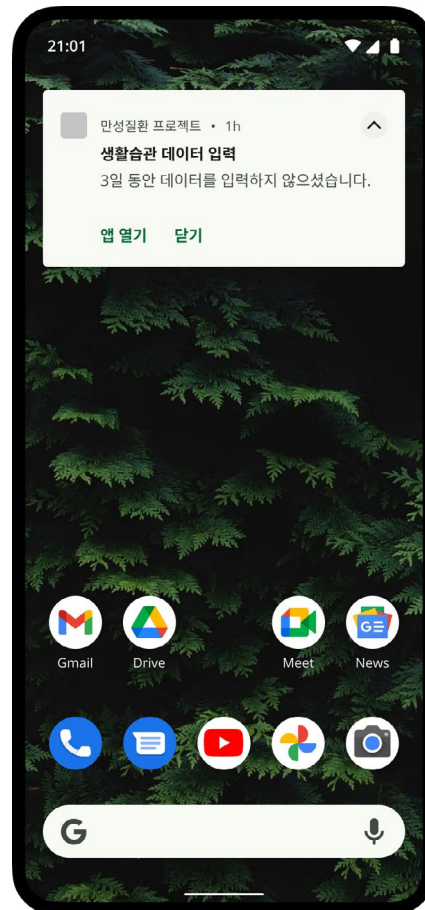
일일 데이터 입력
적립금액

III. 만성질환 관련 임상 및 생활습관 데이터

사용자 입력 데이터 - 데이터 출력 및 알람



사용자 가치(user value)의 일환으로
내가 입력한 데이터를 시각화하여
사용자에게 전달함으로써 건강에 대한
인식/지각 향상 및 습관 형성 유도



데이터를 입력하지 않았을 경우
알림(notification)을 통하여
지속적으로 피험자와 소통하며
이탈률 관리

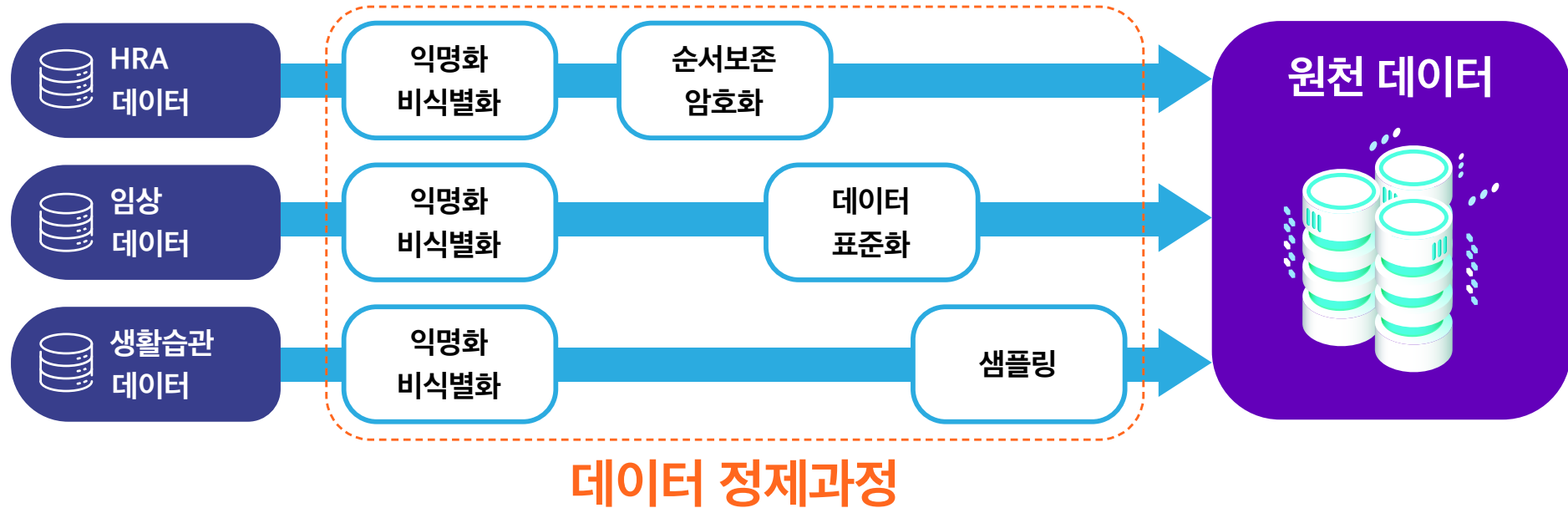
미밴드를 통한 알람도 설계하여
실험 이탈률을 줄이고
데이터의 양적/질적 확보



III. 만성질환 관련 임상 및 생활습관 데이터

만성질환 데이터

원천데이터 - 데이터 정제



생활습관 데이터 중 연속데이터 처리 예시
(심박수 데이터)

원시데이터

시간	BPM
15:02:10	65
15:04:10	70
15:06:10	65
15:08:10	120
...	...
23:56:10	65
23:58:20	59

빈도수 측정

bpm41_50	bpm51_60	bpm61_70	...	bpm210_
0	20	432	...	0

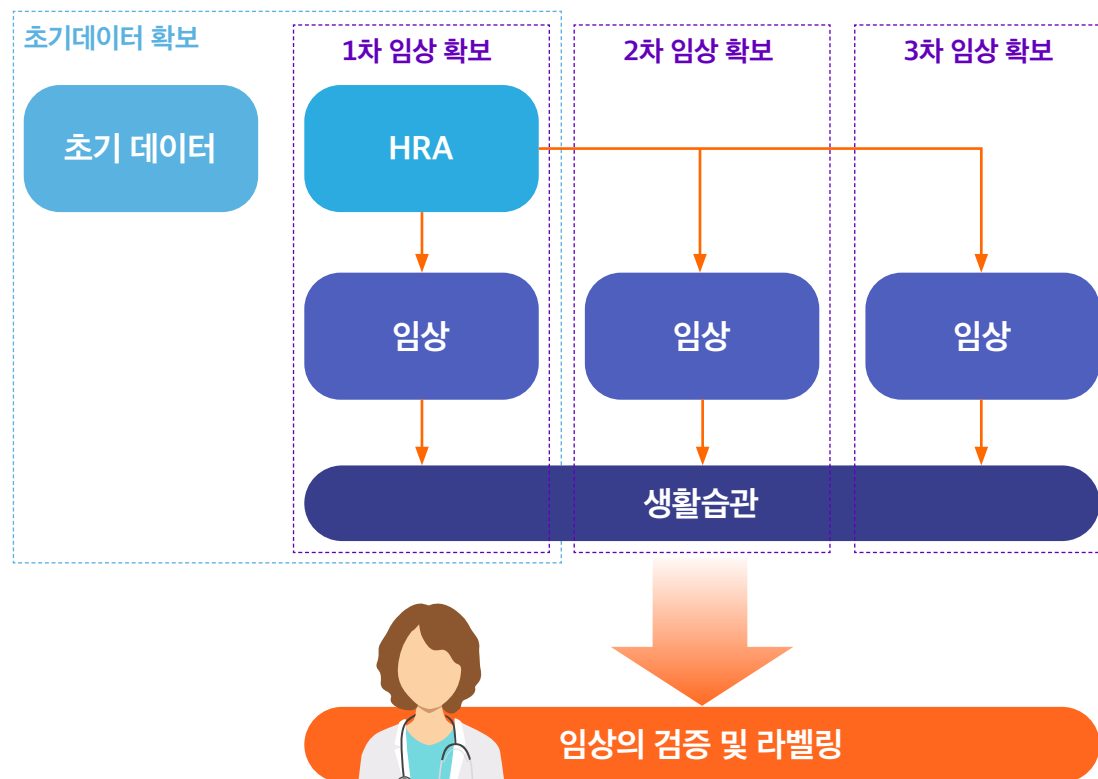
정규화 (빈도 수/총 빈도 수)

원천데이터 if 총 빈도 수 720회

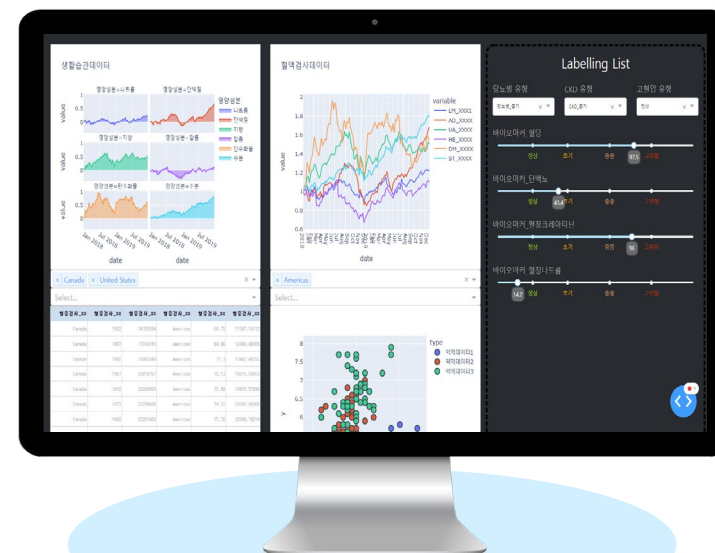
bpm41_50	bpm51_60	bpm61_70	...	bpm210_
0	0.0277	0.6	...	0

III. 만성질환 관련 임상 및 생활습관 데이터 라벨링데이터 - 데이터 가공

라벨링 프로세스



라벨링 도구

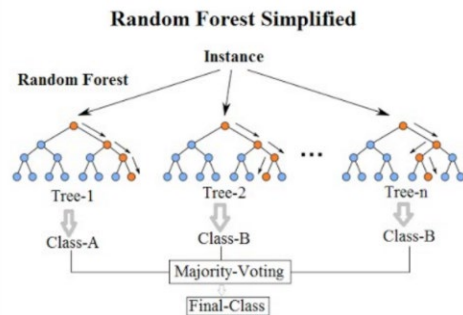


III. 만성질환 관련 임상 및 생활습관 데이터

인공지능 학습모델 : 환자의 당뇨병 (알부민뇨) 검출 유무 예측 모델

당화혈색소 정상 여부 예측 모델

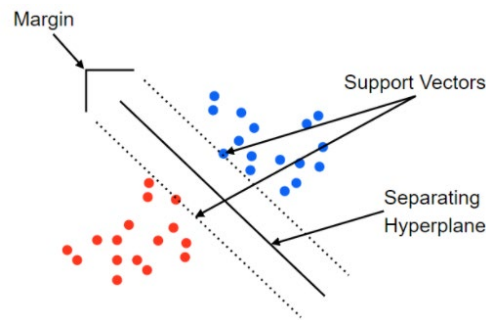
Random Forest



Random Forest는 분류모델로써
사용하며 입력하는 데이터에 대해
다양한 Decision Tree를 생성하여
최적의 결과를 출력하도록 함

SVM

Support Vector Machine

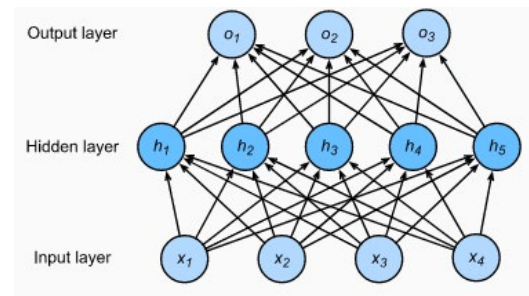


데이터에서 결정되는
Support Vector와
Hyperplane으로 이진분류를
수행하는 알고리즘이며
비확률적 선형 분류모델

이진분류에 특화되어있는 모델로
다중분류의 경우 일대다 분류모델을
다양하게 만들어야함

MLP

Multi Layer Perceptron



MLP는 정방향 인공신경망(feed
forward neural network)으로
데이터는 입력층으로 입력되어
은닉층을 지나
출력층에서 결과가 출력됨

역전파 과정에서 optimizer를
적용하여 가중치를 갱신하여 오차를
최소화 하는 방향으로 학습

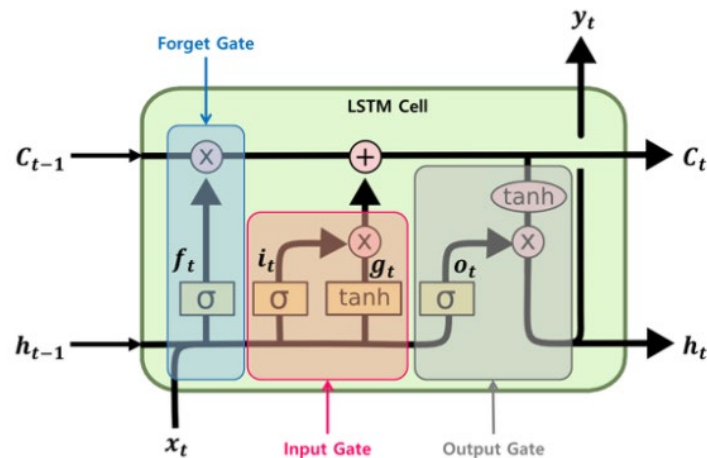
III. 만성질환 관련 임상 및 생활습관 데이터

만성질환 데이터

인공지능 학습모델 : 혈당 수치 예측 모델

LSTM

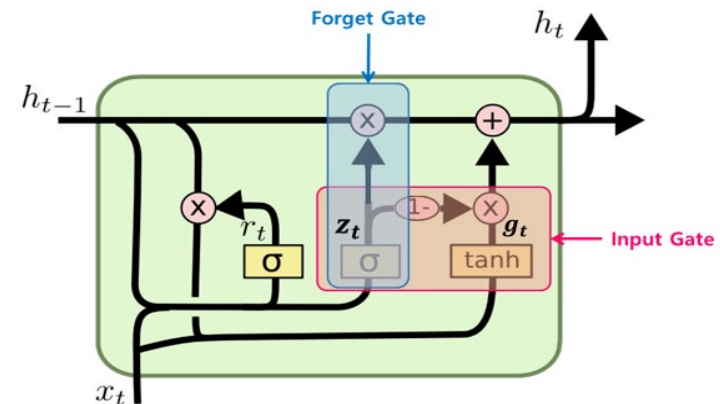
Long Short Term Memory



LSTM은 장기 의존성 문제를 가진 RNN에서 발전된 모델로 각 LSTM cell은 단기상태, 장기상태로 나누어짐

GRU

Gate Recurrent Unit



GRU는 LSTM cell의 간소화된 버전이며 기존 장기상태와 단기상태로 나누어졌던 구조를 하나의 벡터로 합치게 됨

III. 만성질환 관련 임상 및 생활습관 데이터

인공지능 학습모델 품질지표

환자의 단백뇨 (알부민뇨) 검출 유무 예측 모델 당화혈색소 정상 여부 예측 모델	
품질지표	AUROC:ROC curve를 False Positive Rate 에 따라서 적분한 값
선행연구	<ul style="list-style-type: none"> 단백뇨(알부민뇨) 여부 분류모형에 관한 기존 연구에서는 테스트 데이터셋에서의 모델들의 auroc가 0.6~0.65 으로 보고됨 당화혈색소 정상여부 분류 모형에 관한 기존 연구에서는 테스트 데이터셋에서의 모델들의 auroc가 대략 0.8 으로 보고됨
지표기준	<ul style="list-style-type: none"> 단백뇨(알부민뇨) 여부 분류모형 : AUROC 0.6 이상 당화혈색소 정상여부 분류 모형 : AUROC 0.8 이상

혈당 수치 예측 모델	
품질지표	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$ <p>y_i : 실제 데이터 \hat{y}_i : 예측 데이터</p>
선행연구	<ul style="list-style-type: none"> 혈당 수치 예측 모형에 관한 기존연구에서는 테스트 데이터 셋에서의 모델들의 RMSE 가 15~20 사이로 보고됨
지표기준	<ul style="list-style-type: none"> RMSE 기준 20이하

III. 만성질환 관련 임상 및 생활습관 데이터

데이터 품질 관리

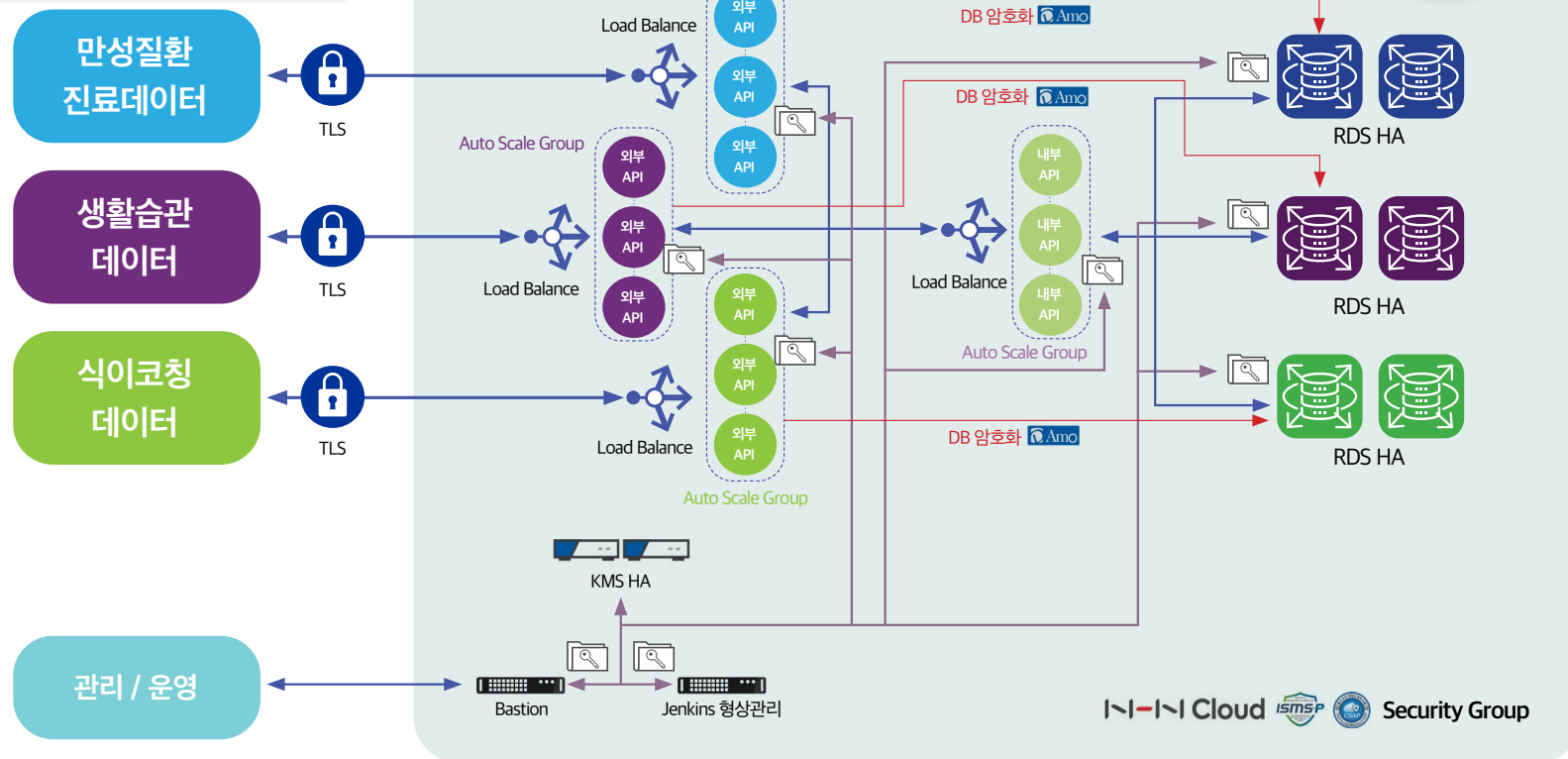
구분	지표	품질목표	품질목표 달성기준
구축 공정	준비성	95% 이상	준비성 체크리스트 목록의 95% 이상 '적정'
	완전성		완전성 체크리스트 목록의 95% 이상 '적정'
	유용성		준비성 체크리스트 목록의 95% 이상 '적정'
데이터 적합성	기준 적합성	95% 이상	구축데이터 품질 기준 적합성 체크리스트 목록의 95% 이상 '적정'
	기술 적합성	99%	파일 포맷 준수율 99% 이상
	통계적 다양성	여성: 30% 이상	성별 다양성을 확보하기 위하여, 구축 목표 데이터 2,000건 기준 여성 임상 데이터 600건 (30%) 이상 확보
		만성질환자: 50% 정상인: 50%	질병 다양성을 확보하기 위하여, 구축 목표 데이터 2,000건 기준 만성질환자 1,000건 (50%) 이상, 정상인 1,000건 (50%) 이상 확보
데이터 정확성	의미 정확성	99% 이상	사업 참여/비 참여 의료진 교차검증 합의도 99%
	구문 정확성	99% 이상	라벨링 데이터 구조, 형식 오류율
학습모델	알고리즘 적정성	Pass	알고리즘 적정성 과제심의 통과
	유효성	0.6 이상	단백뇨(알부민뇨) 여부 분류모형: AUROC 0.6 이상
		0.8 이상	당화혈색소 정상여부 분류 모형: AUROC 0.8 이상
		20 이하	혈당 수치 예측모형 : RMSE 기준 20이하

III. 만성질환 관련 임상 및 생활습관 데이터 사업추진일정

세부과제명	수행내용	추진일정 (월)								비중 (%)
		4	5	6	7	8	9	10	11	
만성질환 관련 임상 및 생활습관 데이터	착수보고									10
	생활습관 기기 피검사자 교육									10
	알고리즘 개발 및 검증 (1)									
	라벨링 인력 모집 및 교육, 훈련									10
	라벨링용 시스템 환경 구축									
	원시 데이터(검진데이터) 20% 획득									20
	원천 데이터 정제 및 검수									
	라벨링 작업 및 검수									
	AI 학습 및 알고리즘 성능검증 (2)									
	원시 데이터(검진데이터) 획득									40
	원천 데이터 정제 및 검수									
	라벨링 작업 및 검수									
	AI 학습 및 알고리즘 성능검증 (3)									
	완료보고									10

III. 만성질환 관련 임상 및 생활습관 데이터

활용방안 - 만성질환 관리 마이데이터 플랫폼 구축

인공지능 기반 만성질환관리
마이데이터 플랫폼

IV 성과홍보 및 일자리 창출



IV. 성과홍보 및 일자리 창출

성과홍보



2022년 인공지능 학습용 데이터 구축 지원사업 (1차)

2022 만성질환 관리 인공지능 경진대회

일시 : 2020년 10월 참가대상 : 제한 없음 총 상금 : 1,000만원

전남대학교병원 컨소시엄

관련학회 및 세미나 참석



언론 홍보

Google

전남대학교병원 인공지능학습데이터

[All](#) [News](#) [Images](#) [Videos](#) [Maps](#) [More](#) [Tools](#)

About 231,000 results (0.54 seconds)

<https://www.cnuh.com/board>

전남대병원 AI데이터구축사업 선정

Aug 2, 2021 — 이에 따라 국내외 최대 규모의 청력검사 학습데이터 구축과 청력진단분야의 인공지능 연구개발이 활성화 될 것으로 전망된다.

<https://www.cnuh.com/board>

전남대병원 핵의학과 팀, AI학습용 데이터톤 최우수상

Jan 20, 2022 — 가천대 산학협력단 주최 '2021의료데이터 AI학습용 데이터톤' 대회서 ... 파킨슨병 환자의 데이터를 이용하여 AI 학습모델을 개발하기 위해 진행됐다.

<http://www.bosa.co.kr/news/articleView>

전남대병원 'AI데이터 구축사업' 선정 - 의학신문

Aug 2, 2021 — [의학신문·일간보사=차원준 기자] 전남대학교병원이 최근 과학기술정보통신부의 2021년도 인공지능(AI) 학습용 데이터 구축사업의 2개 분야에서 ...

<http://www.bosa.co.kr/news/articleView>

일자리 창출

“총 69명의 신규 인력 채용 예정”

전문 의료데이터 라벨링을 위한 경력단절 연구간호사 우선 채용

만성질환자/정상인의 생활습관데이터 입력을 설명하고
독려하는 코디네이터 채용

신규 채용인력에 헬스케어 인공지능 교육기회 제공 및 정규직 전환 지원

V 성공적인 사업 수행을 위한 과제 조정 건의사항



V. 성공적인 사업 수행을 위한 과제 조정

만성질환 데이터 활용 목표 → 만성질환 관련 디지털 치료제 개발

1. 만성질환 데이터 수량 조정 가능 여부

사전공고 내용	본공고 내용
<ul style="list-style-type: none"> 만성질환자와 정상인을 포함한 최소 1천명 이상의 임상데이터와 생활습관 데이터 	<ul style="list-style-type: none"> 만성질환자 최소 1천명 이상의 임상데이터와 생활습관 데이터 정상인 1천명 이상의 임상데이터와 생활습관 데이터

- 의료분야에서 전향적 데이터 수집에 1천 명 이상의 코호트 구성은 매우 큰 규모임
- 정상인 데이터의 경우 만성질환자와의 대조군 역할로 1천명의 대조군은 너무 큰 규모임
- 만성질환 데이터의 경우 1회가 아니라 참가자들이 4~5개월 동안
매일 데이터 수집에 참여하여야 하므로 데이터 수집 난이도가 매우 높음
- 만성질환 데이터는 4~5개월 연속적으로 데이터를 수집하여야 하며, 중간에 이탈 시 해당 데이터의 의미가 없어짐
- 정상인 참가자는 관리가 가능한 그룹을 대상으로 최소화 하고,
정상인 참가자를 줄인만큼 만성+정상인 참가자에 대한 보상을 높여 참가자 이탈을 최소화 와 의미 있는 데이터 수집에 집중
- 만성질환자 데이터 목표 1,000명(모집 1,200명), 정상인 데이터 목표 200명 (모집 300명)**

2. 임무정의 변경

- 혈당수치 예측 모델 삭제 또는 '만성질환자의 수면 시간 예측'으로 변경
- 짧은 데이터 수집 기간으로 인한 LSTM 모델 적용이 어려움

Q &

A

질의응답



전남대학교병원
컨소시엄