

연합학습 기반 ADMET-PK/PD 모델 개발

Federated ADMET-PK/PD model (FAM-PK/PD) development

주관: 전북대학교 의과대학 김민걸

공동: 충남대학교 컴퓨터융합학부 정상근



전북대학교
JEONBUK NATIONAL UNIVERSITY



충남대학교

| 선행연구경험을 바탕으로 한 실력있는 AI-Domain의 경계없는 융합연구팀

주관연구 책임자

- 임상시험을 총 200건 이상 수행 및 약동력학을 주제로 한 논문을 다수 저술하고 있음
- 연합학습을 다기관 EHR 데이터에 실제 적용하여 간암/대장암 환자의 예후를 예측한 경험이 있음

공동연구 책임자

- ADMET prediction 관련하여 AI-ADMET 관련 기술플랫폼을 개발 및 사업화
(최근 3년간 등록 1건, 출원 8건 (PCT 1건 포함))

주관연구책임자 김민걸

- 전북대학교병원 임상시험센터장
- 대한임상시험센터협의회 총무이사
- 대한임상약리학회 정보이사
- 대한의료정보학회 임상시험이사
- 식품의약품안전처 중앙약사심의위원

유사 수행 사업

- ✓ 신종·풍토성 감염병 치료제 개발을 위한 전향적 실사용 데이터 레지스트리 구축과 연구 지원체계 마련 (2022~2024)
- ✓ 감염병 임상시험 표준데이터 활용연구 (2021)
- ✓ 차세대 신약 초기 임상시험 기술 개발 (2019-2022)
- ✓ 블록체인 기반 차세대 임상시험 플랫폼 구축 (2019-2021)

참여연구진

- 전북대학교 약학대학 김재현 교수
- 전북대학교병원 임상약리센터
유주은, 류수향, 양금주, 손보라 연구원

공동연구책임자 정상근

- 충남대학교 컴퓨터융합학부 부교수
- 2024 한글 및 한국어 정보처리 학술대회 학술위원장
- 부스트캠프 NLP 마스터
- LiLac CTO

유사 수행 사업

- ✓ AI 기반 약물-단백질 상호작용 예측 서비스 (2023)
- ✓ AI 기반 ADME/T 예측 서비스 (2022)
- ✓ 인공지능 기반 제약 및 약료 기술 연구 스타트업 (2022 ~)
- ✓ 충남대학교 BIO-AI 융합 연구 (2021 ~)

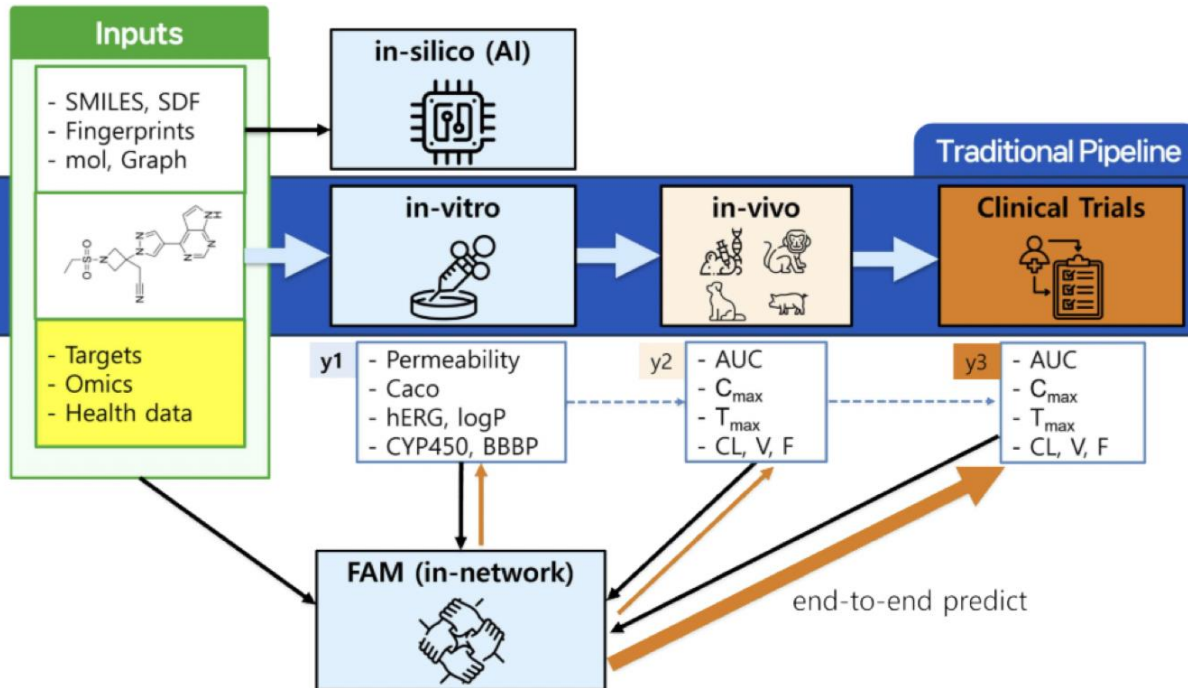
참여연구진

- 충남대학교 약학대학 채정우 교수, 박상민 교수, 이소영 교수
- 충남대학교 공과대학 황태욱, 서혜인, 정지수, 강현석, 남궁혁 연구원
- 충남대학교 약학대학 구성우, 이현정, 김도영 연구원

연구개발과제의 목표 및 내용

연합학습 AI 신약 개발의 새로운 길을 열 수 있을까?

- Federated ADMET-PK/PD Model (FAM-PK/PD)



FAM 모델 → 기관에 분산되어 있는 데이터를 직접 공유하지 않고, 예측 모델을 구성하는 파라미터에 대한 가중치를 전달받는 형태로 각 기관의 데이터 프라이버시를 보호함과 동시에 모델의 예측력 향상

ADMET-PK/PD → ADMET feature 의 (전)임상 PK/PD 연계 및 전파를 통한 예측력 향상

연구개발과제의 목표 및 내용

목표: 연합학습 기반 ADMET-PK/PD 모델 (*Federated ADMET-PK/PD Model, FAM-PK/PD) 개발

*FAM-PK/PD: 본 연구팀 제안 연합학습 기반 *in-vitro/in-silico* ADMET 특성 및 전임상-임상 PK/PD 파라미터가 연계된 전주기 AI 솔루션

1차년도

- FAM 모듈 구축을 위한 *in-vitro* ADMET feature 및 임상 데이터 확보
- FedAvg 를 활용한 FAM 모듈 구축
- FAM-PK/PD 모듈 구축을 위한 데이터 확보

2차년도

- FAM-PK/PD 모듈 구축
- FAM-PK/PD 플랫폼 활용 ADMET 결과와 전임상-임상 PK/PD 파라미터 연계
- (전)임상 PK/PD 예측을 위한 연합학습 기반 모델 개발 및 검증

3차년도

- 플랫폼 전개용 패키지 제작
- 표준화를 위한 SOP 마련
- FAM-PK/PD 플랫폼 고도화
- FAM-PK/PD 운영을 통한 성능 개선 방안 제시

주요 핵심 정량 성과

제약사 및 연구소 등 FAM-PK/PD 활용 연구기관 데이터 표준화를 위한 전처리 도구 1건 이상 개발

데이터 전처리 도구 및 FAM-PK/PD 활용을 위한 표준작업지침 수준의 사용 가이드라인 작성

FAM-PK/PD 관련 개발 및 활용 관련 논문 발표 (2건 이상) 및 특허 등록

FAM-PK/PD 모듈의 Good software 인증 취득

사업화 지원을 위한 FAM-PK/PD 활용 관심 연구소 및 기업의 기술 지원 (목표 24건) 및 플랫폼 이식 계획 수립 및 실행 (목표 4건)

FAM 모듈 구축: *in-vitro* ADMET feature 예측을 위한 연합학습 기반 모델 개발 및 검증

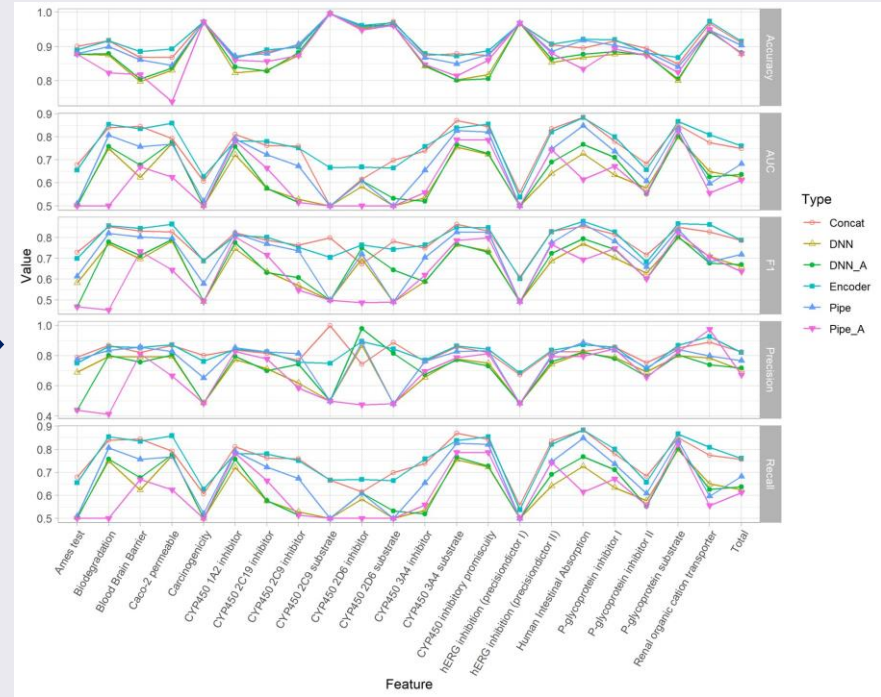
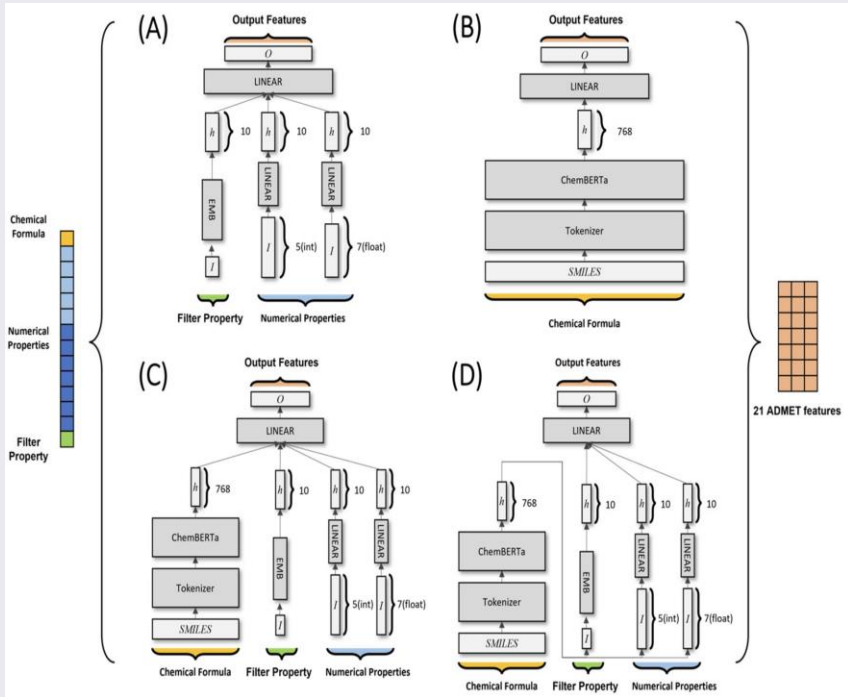
- FAM 모듈 구성 요약

In-vitro ADMET feature 예측을 위한 연합학습 기반 모델 개발 및 검증			
활용DB	활용모델	학습입력값	예상출력값
Drugbank, ChEMBL, KEGG, HMDB, SMPDB 등	Federated 모델, DNN, CNN, GNN (Transfer learning, Python, R)	SMILES, metabolic pathway, action pathway, Solubility, LogP, LogS, pKa, Charge, Hydrogen acceptor/donor count, Polar surface area, Rotatable bond count, Refractivity, Polarizability, Number of Rings, Bioavailability	Intestinal absorption, Blood brain barrier, p-gp substrate/inhibitor, CYP substrate/inhibitor, CYP inhibitory promiscuity, Carcinogenicity, Acute toxicity, hERG inhibition

- 연합학습의 방법 중 하나인 FedAvg를 사용한 FAM 모듈 구축
 - 글로벌 모델 구축: 공개된 DB 활용 SMILES 정보에 따른 ADMET 특성 예측 모델
 - 로컬 모델 학습과정: 각 참여기업 및 기관은 글로벌 모델 활용하여 독립적 학습 수행
 - model merge: 로컬 모델 가중치 및 글로벌 모델 통합을 위한 알고리즘으로 Federated Averaging (FedAvg) 적용 및 업데이트

선행연구 기술 활용

선행 연구 예시: SMILES 정보에 따른 대표 ADMET 특성 예측



- Drugbank 등 공개 DB 활용
- Investigator Initiated Trial (IIT) 데이터 및 Vmi 포털 활용을 통한 임상데이터 확보

연차별 연구과제 산출물

- 1차년도
 - 임상데이터 포맷 정의서: 자체 활용가능한 임상데이터 및 타 세부과제 임상데이터 구성 확인
 - 전임상데이터 포맷 정의서: 공개 DB의 in vitro feature 및 타 세부과제 in vitro 데이터 구성 확인
 - FAM 모듈 및 FAM을 활용한 in vitro ADMET feature 예측
- 2차년도
 - FAM-PK/PD 모듈 및 활용 매뉴얼
 - 전임상/임상 데이터 전처리 도구
 - 관련 특허 출원
- 3차년도
 - FAM 활용 고도화 방안
 - Good Software 인증 및 software 등록
 - 논문 게재 및 특허 등록
 - 추가 데이터 활용 방안 및 플랫폼 환류 시스템 구축