

메가스케일 ADMET-파운데이션 모델 기반 초경량 FAM 통합 솔루션 개발

주관연구책임: 광주과학기술원 전기전자컴퓨터/AI대학원 남호정

공동 연구기관 책임: HITS 황상연

메가스케일 ADMET-파운데이션 모델 기반 초경량 FAM 통합 솔루션 개발

주관  **광주과학기술원**
Gwangju Institute of Science and Technology

ADMET-파운데이션 모델 기반 초경량 FAM 모델 개발

- 주관 연구책임: 남호정 교수
- 참여연구진: 6 명
- 주관 연구 책임자 전문성
 - 최근 5년간 AI기반 약물개발 SCI 논문 출판 10회 이상
 - 화합물 구조 생성 알고리즘 개발 1건
 - 화합물-표적 예측 알고리즘 개발 3건
 - 화합물 독성 예측 모델 개발 2건
 - 화합물 부작용 예측 모델 개발 1건
 - 약물 분자 활성 예측 알고리즘 개발 3건



- ADMET 프로파일 개발
- ADMET-파운데이션 모델 개발
- 초경량 FAM 모델 개발

공동  **HITS**

범용성 FDD 데이터 전처리기 개발

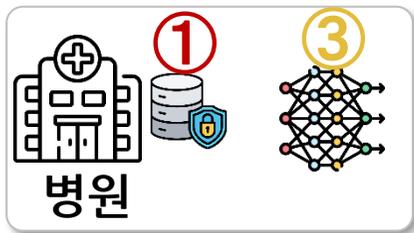
- 공동 연구책임: 황상연 박사
- 참여연구진: 4 명
- 공동 연구 책임자 전문성
 - AI 신약개발 플랫폼 Hyper Lab 개발
 - 클라우드 기반 SaaS 플랫폼
 - 단백질-리간드 복합체 구조 예측
 - 단백질-리간드 활성 예측
 - 리간드 디자인
 - ADMET 및 물리화학 물성 예측
 - 유효물질 가상탐색



- ADMET 데이터베이스 구축
- FDD 데이터 전처리기 개발
- 사용자 매뉴얼 개발

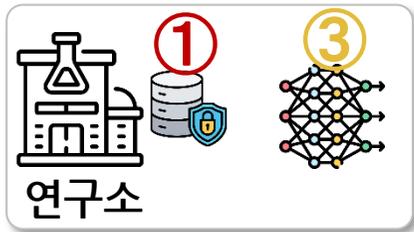


제약사

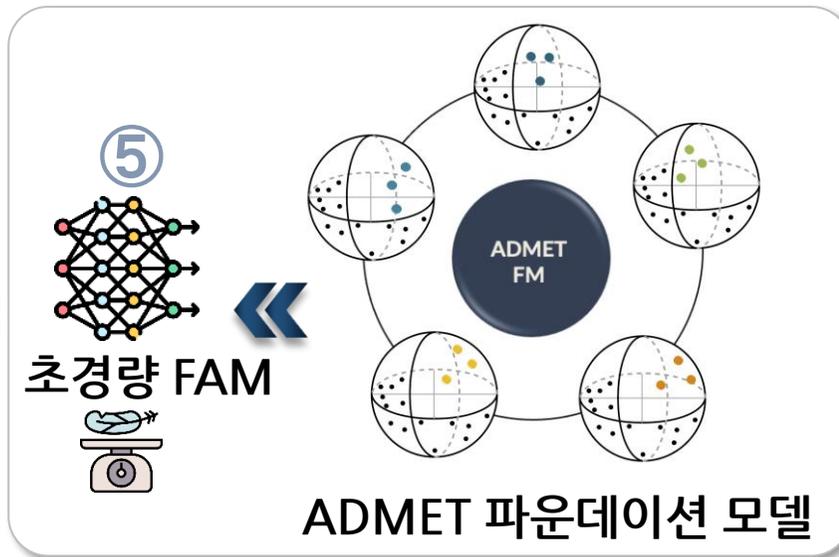


병원

⋮



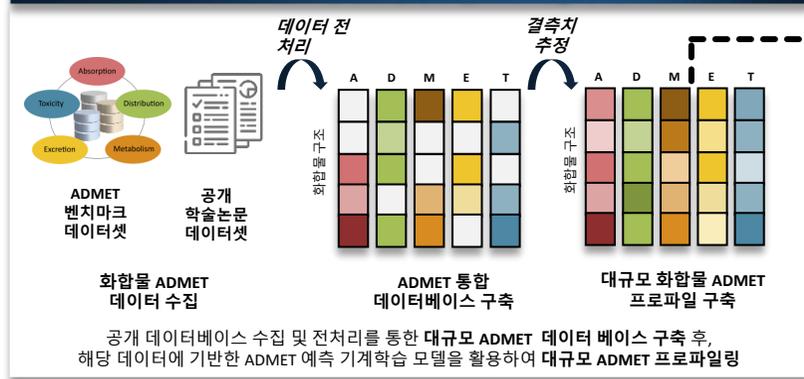
연구소



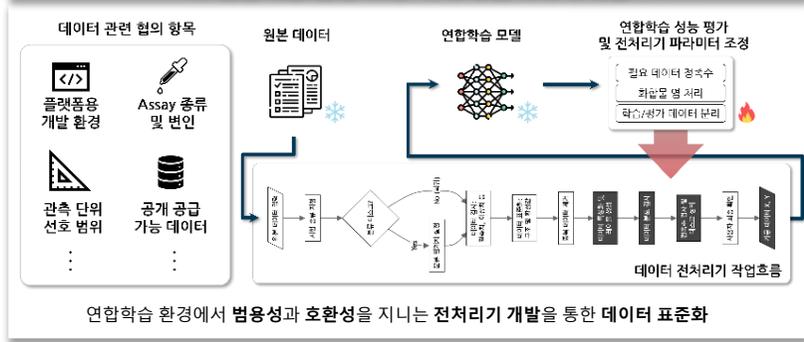
- ① : 데이터 전처리
- ② : FAM 모델 배포
- ③ : 로컬 학습
- ④ : 파라미터 취합
- ⑤ : 모델 업데이트

- ❖ 데이터 전처리 배포를 통한 데이터 보안 구현
 - ✓ 데이터 구조 정보 노출 위험 없음
- ❖ 메가스케일 ADMET-파운데이션 모델
 - ✓ +10M 화합물 구조 정보, >50 end-points, ~60K 화합물, ~200K 활성 정보
 - ✓ 임의의 end-point 태스크에 사용 가능한 범용적 파운데이션 모델
 - ✓ ADMET embedding을 기반한 multi-task learning 효과
- ❖ 경량 FAM
 - ✓ ~10K개의 학습 파라미터를 요구하는 초경량 모델
 - ✓ 대규모 태스크에 적용가능 (태스크당 A100 기준 1분 이내, RTX4090 기준 2분 이내)

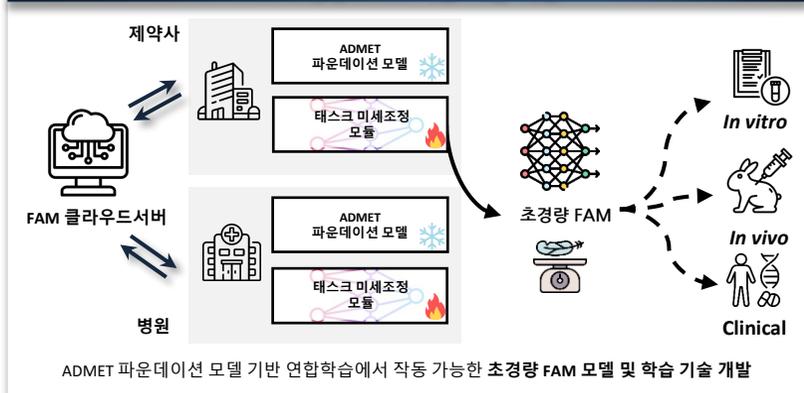
1. ADMET 데이터베이스 및 프로파일 구축



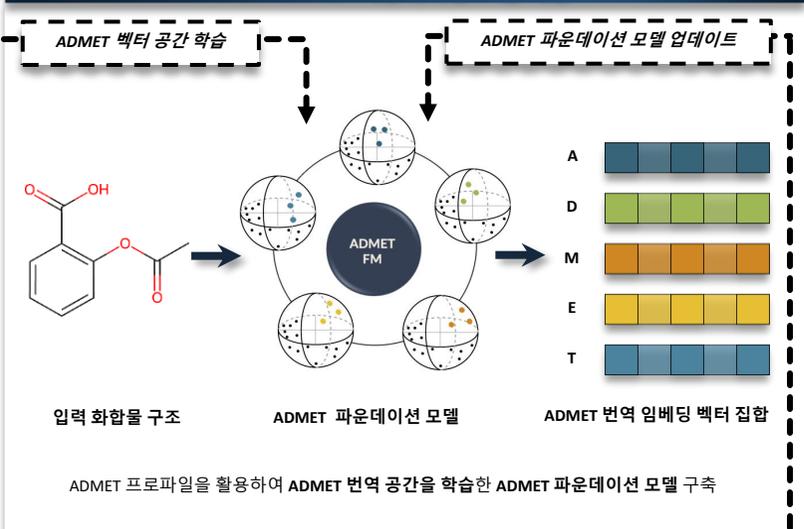
2. 연합학습 환경 호환 데이터 전처리기 개발



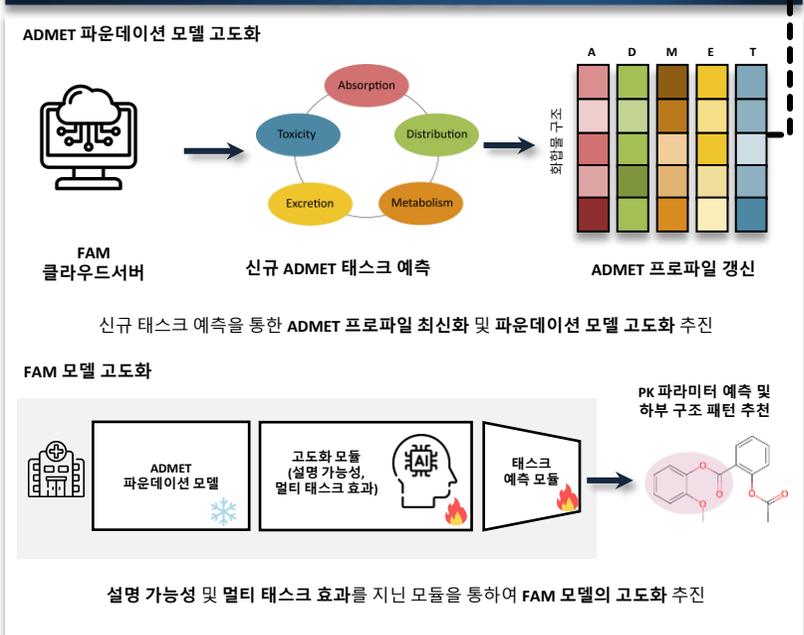
4. 초경량 FAM 학습 기술 개발

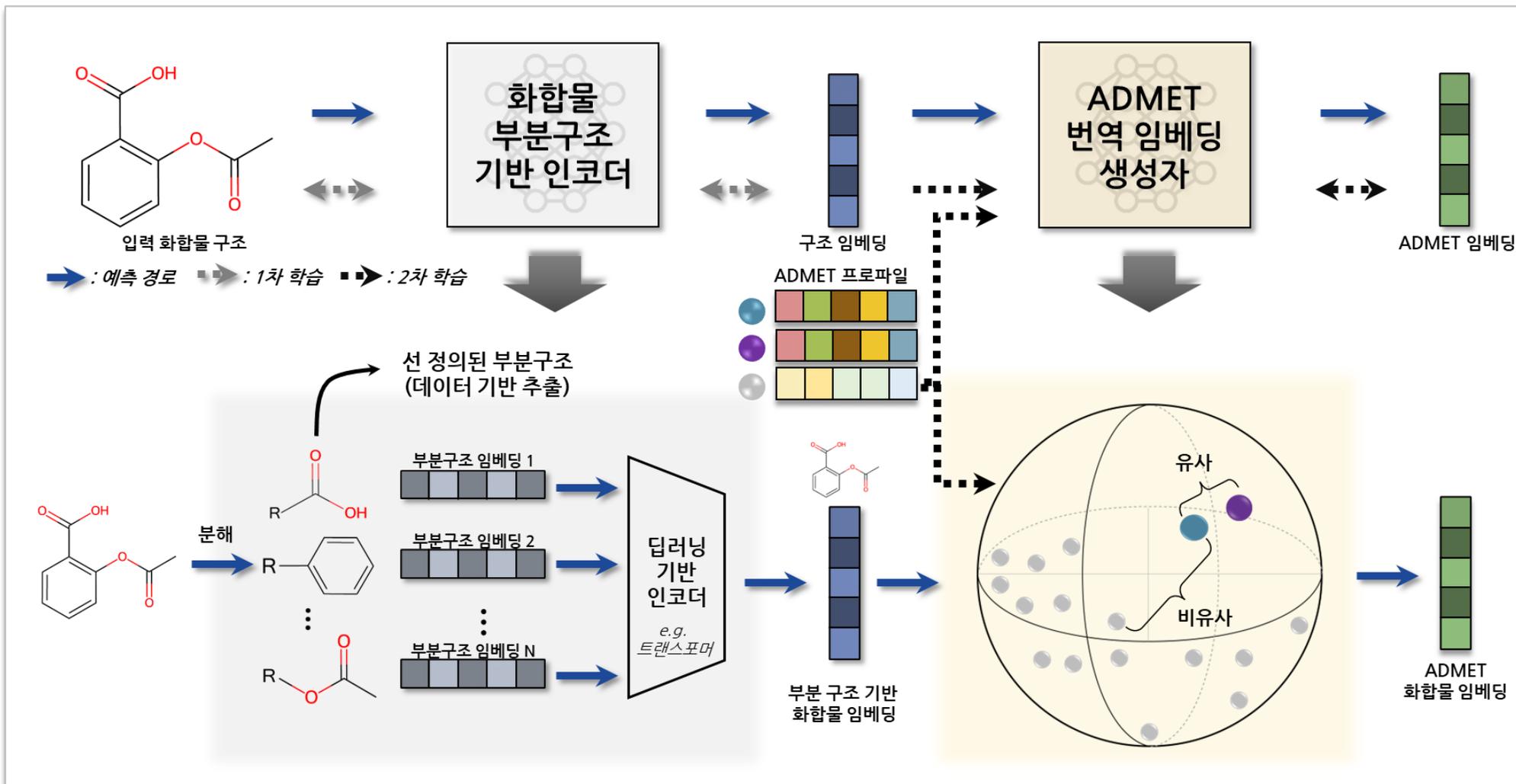


3. ADMET 파운데이션 모델 개발



5. ADMET 파운데이션 모델 및 FAM 고도화

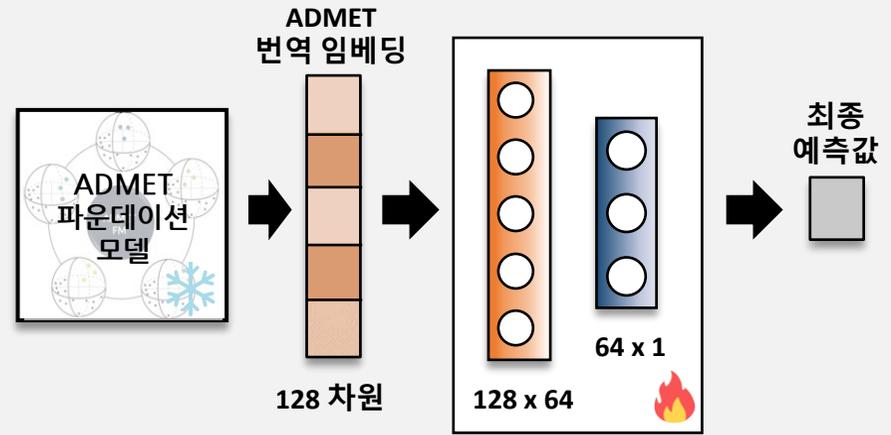




- ✓ 화합물 부분구조 정보와 ADMET 프로파일 정보 활용
- ✓ 입력 화합물 구조로부터 ADMET 임베딩 벡터를 생성하는 파운데이션 모델 개발

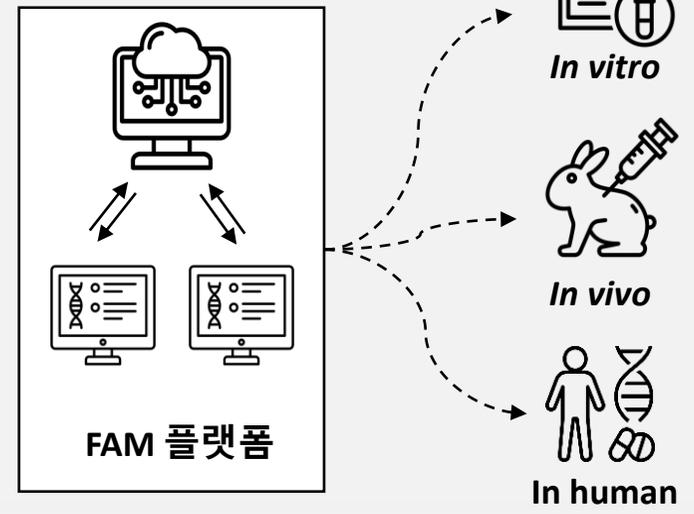


태스크별 미세조정 경량 네트워크 세부 내용 예시

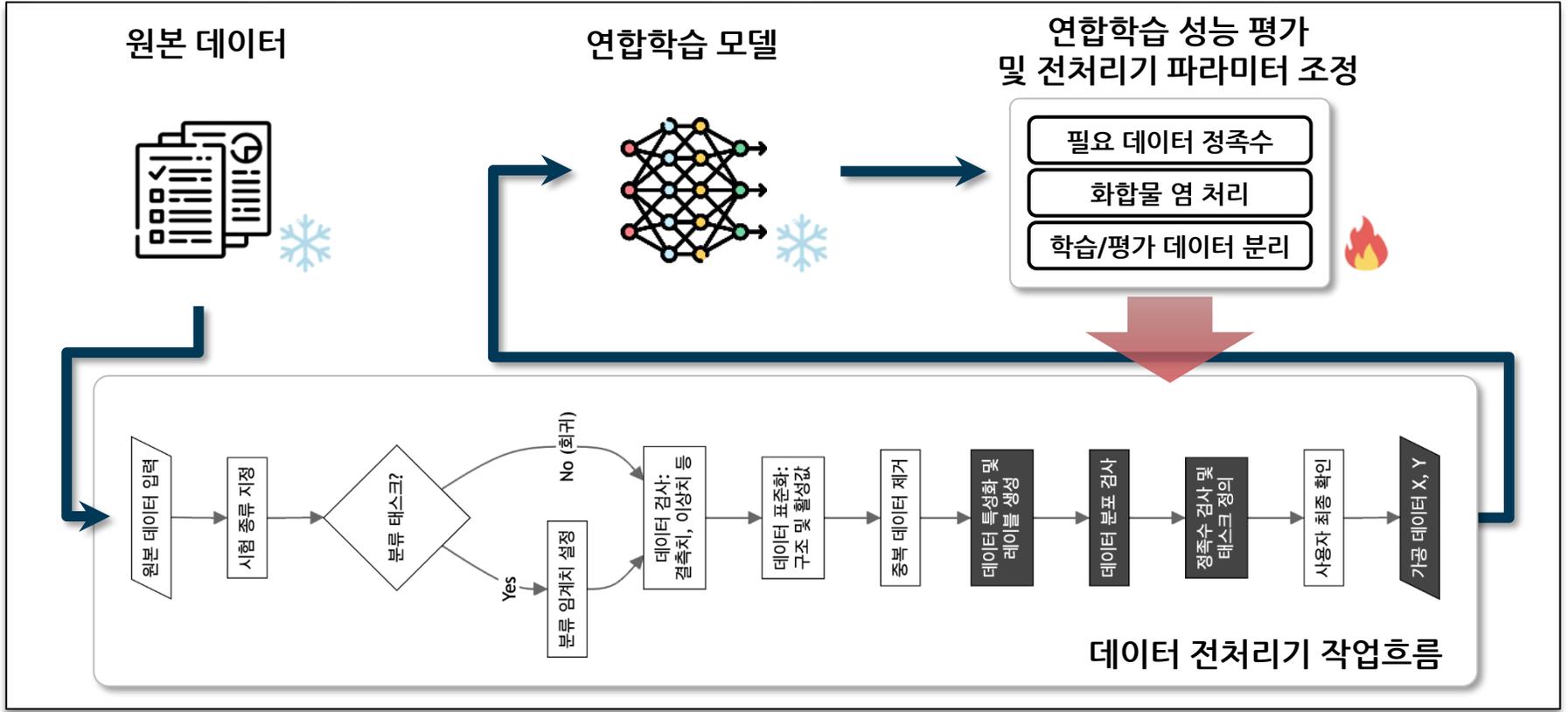


예상 학습 파라미터 수 = $(128 \times 64) + (64 \times 1) = 8,256$ 개
 중간 차원을 변경함에 따라 더욱 경량화 가능

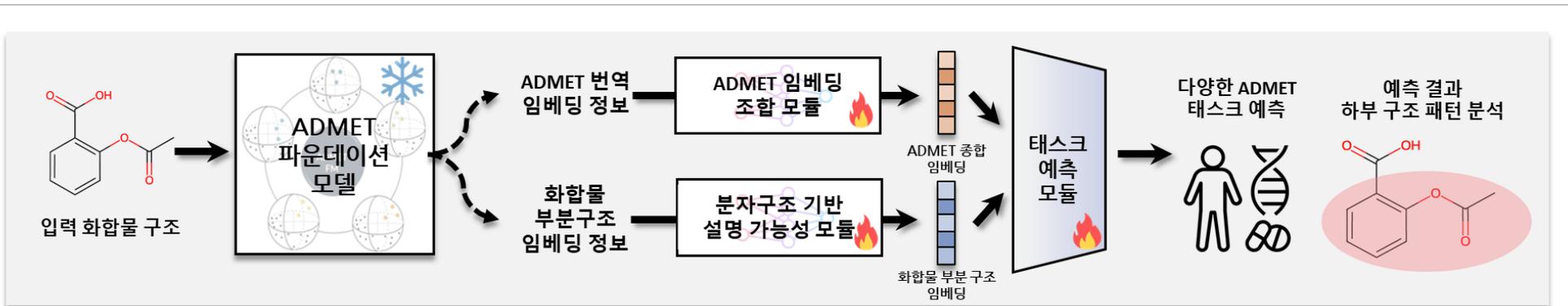
범용적 FAM 미세조정 프레임워크



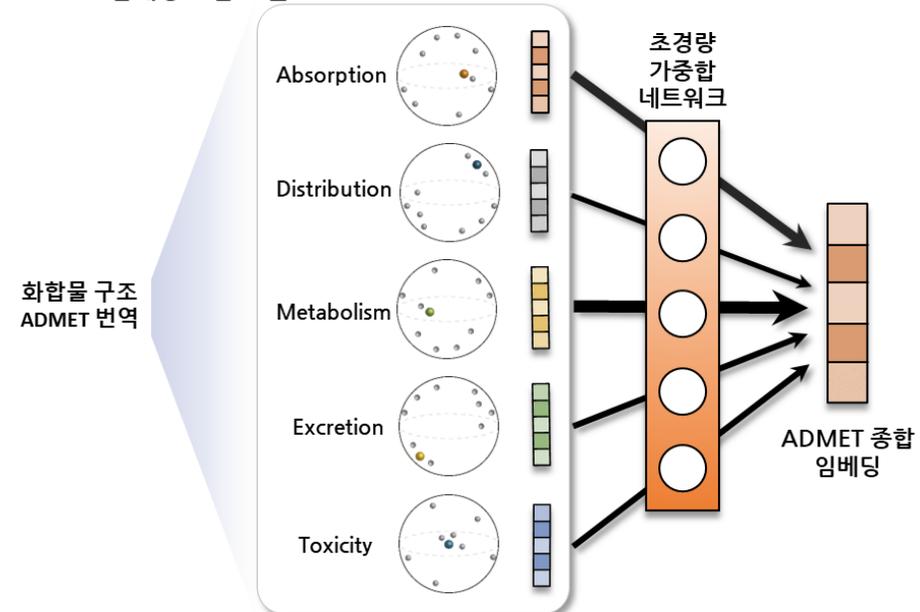
✓ 다운스트림 태스크에 제약이 없는 범용적 FAM 개발



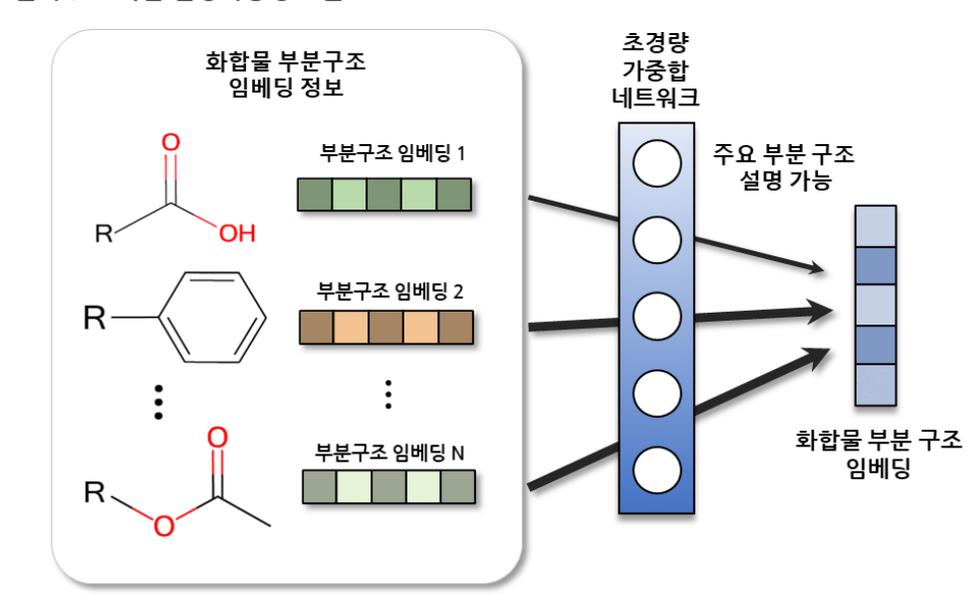
✓ 가상 연합학습 실험을 통한 전처리기 최적화 및 로컬라이징



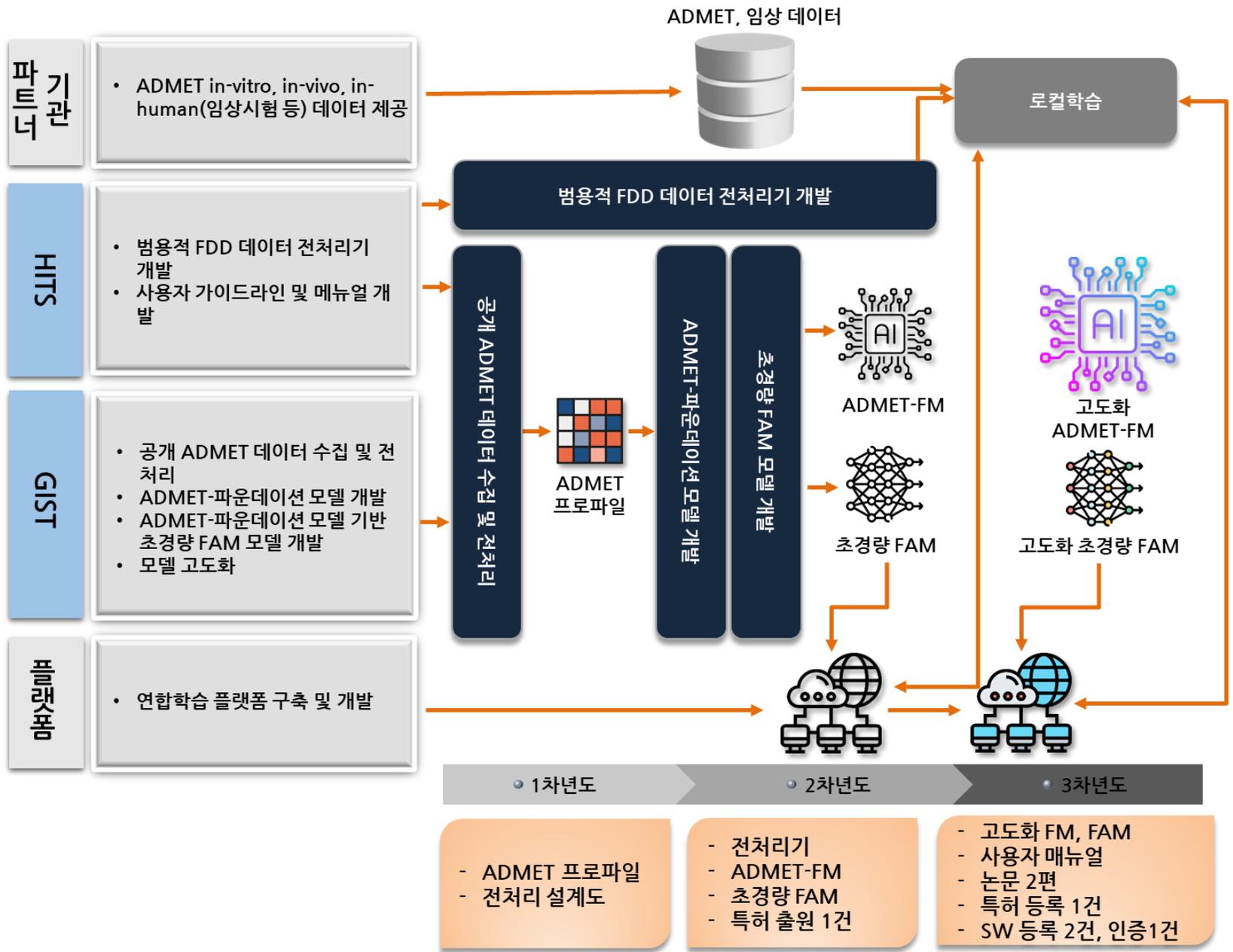
ADMET 임베딩 조합 모듈



분자구조 기반 설명가능성 모듈



- ✓ ADMET 임베딩 조합을 통한 멀티태스크 러닝 효과 구현
- ✓ 다운스트림 태스크 연관 분자 하부구조 패턴 분석 기능 고도화



파이프라인

- ADMET in-vitro, in-vivo, in-human(임상시험 등) 데이터 제공

HITS

- 범용적 FDD 데이터 전처리기 개발
- 사용자 가이드라인 및 매뉴얼 개발

GIST

- 공개 ADMET 데이터 수집 및 전처리
- ADMET-파운데이션 모델 개발
- ADMET-파운데이션 모델 기반 초경량 FAM 모델 개발
- 모델 고도화

플랫폼

- 연합학습 플랫폼 구축 및 개발

1차년도 2차년도 3차년도

- ADMET 프로파일
- 전처리 설계도

- 전처리기
- ADMET-FM
- 초경량 FAM
- 특허 출원 1건

- 고도화 FM, FAM
- 사용자 매뉴얼
- 논문 2편
- 특허 등록 1건
- SW 등록 2건, 인증 1건