

연합학습 기반 신약개발 가속화 프로젝트 [RS-2024-00461734]

인공지능 신약 개발을 위한 연합학습 데이터 확보 및 생태계 구축

동 화 약 품 (주)

연구책임자 원대연

I 연구개발기관 및 연구자 소개

- 1897년 동화약품 설립
- 1972년 중앙연구소 설립
- 연구개발인력 100명 이상 (전체 대비 약 14%)
- 국내 신약 제3호 밀리칸주 & 제23호 자보란테정 개발
- 본 과제 참여 인력 구성 (연구책임자 원대연 박사 외 15명)



신약연구
신약 개발 과제 기획/총괄 (2명)

의약화학
물질 합성 및 최적화 연구 (2명)

ADME/PK
흡수/분포/대사/배설 및 약동학적 분석 및 평가 (4명)

In vivo
약물 효능 및 안전성 평가 (6명)

CADD
약물 설계 및 데이터 분석 (1명)

신약개발 전주기 과정 경험 및 노하우 보유

연구책임자 중심 합성신약 개발 분야 별 전문 인력 구성

II 연구개발 목표

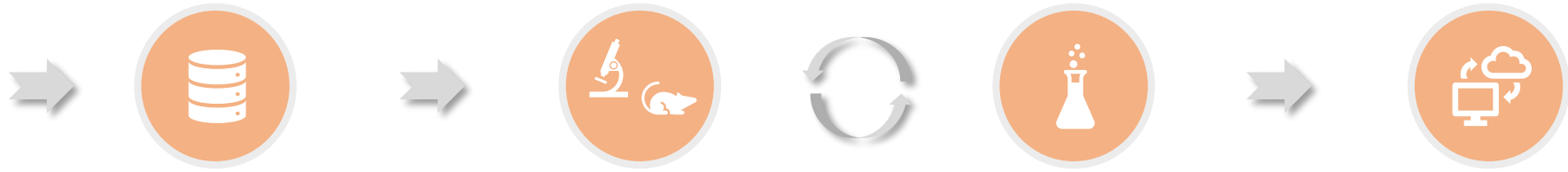
최종 목표

1. 연합학습의 이해 및 ADMET/PK 예측모델(FAM) 개발을 위한 데이터 확보
2. ADMET/PK 예측 모델 검증을 통한 산업적 활용 및 생태계 구축 기여

수행 계획

- 1단계 : 데이터 표준화 및 데이터 공급
- 2단계 : FAM 모델 검증 및 활용 검토

Ⅲ 핵심 연구 내용 및 기술 (1단계)



보유 데이터 표준화

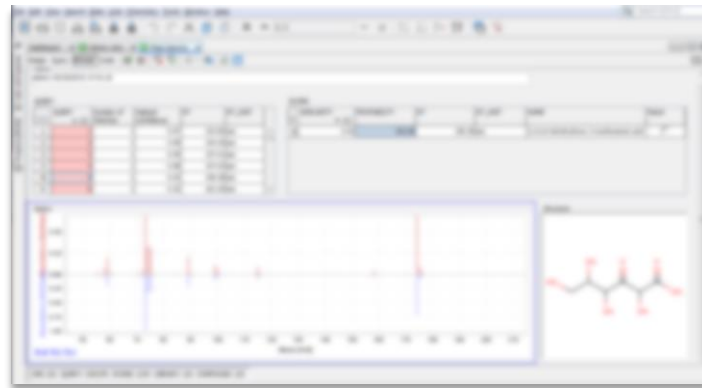
- 자사 합성신약 라이브러리 자체 운영
- 신약 전주기 실험데이터 구성

신규 데이터 생성

- 신규 데이터 생성 (ADMET/PK, 합성)
- 신약연구 전담부서 수행 및/또는 지정 CRO 의뢰
- 실험별 필요량에 따라 일부 합성 가능

데이터 공급

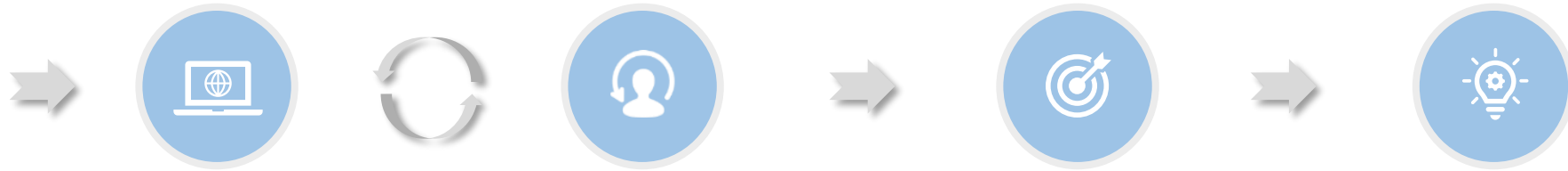
- 보유/신규데이터 표준화 작업
- 데이터 공급



ID	Structure	Molar Weight	Formula	SMILES
1		206.26	C ₁₄ H ₁₈ N ₂ O ₂	CCCCN(C)CC
2		242.31	C ₁₈ H ₂₂ N ₂ O ₂	CCCCN(C)CC
3		274.37	C ₂₂ H ₂₆ N ₂ O ₂	CCCCN(C)CC
4		306.42	C ₂₆ H ₃₀ N ₂ O ₂	CCCCN(C)CC
5		338.48	C ₃₀ H ₃₄ N ₂ O ₂	CCCCN(C)CC
6		370.53	C ₃₄ H ₃₈ N ₂ O ₂	CCCCN(C)CC
7		402.59	C ₃₈ H ₄₂ N ₂ O ₂	CCCCN(C)CC
8		434.64	C ₄₂ H ₄₆ N ₂ O ₂	CCCCN(C)CC
9		466.70	C ₄₆ H ₅₀ N ₂ O ₂	CCCCN(C)CC
10		498.75	C ₅₀ H ₅₄ N ₂ O ₂	CCCCN(C)CC
11		530.81	C ₅₄ H ₅₈ N ₂ O ₂	CCCCN(C)CC
12		562.86	C ₅₈ H ₆₂ N ₂ O ₂	CCCCN(C)CC
13		594.92	C ₆₂ H ₆₆ N ₂ O ₂	CCCCN(C)CC
14		626.97	C ₆₆ H ₇₀ N ₂ O ₂	CCCCN(C)CC
15		659.03	C ₇₀ H ₇₄ N ₂ O ₂	CCCCN(C)CC
16		691.08	C ₇₄ H ₇₈ N ₂ O ₂	CCCCN(C)CC
17		723.14	C ₇₈ H ₈₂ N ₂ O ₂	CCCCN(C)CC
18		755.19	C ₈₂ H ₈₆ N ₂ O ₂	CCCCN(C)CC
19		787.25	C ₈₆ H ₉₀ N ₂ O ₂	CCCCN(C)CC
20		819.30	C ₉₀ H ₉₄ N ₂ O ₂	CCCCN(C)CC
21		851.36	C ₉₄ H ₉₈ N ₂ O ₂	CCCCN(C)CC
22		883.41	C ₉₈ H ₁₀₂ N ₂ O ₂	CCCCN(C)CC
23		915.47	C ₁₀₂ H ₁₀₆ N ₂ O ₂	CCCCN(C)CC
24		947.52	C ₁₀₆ H ₁₁₀ N ₂ O ₂	CCCCN(C)CC
25		979.58	C ₁₁₀ H ₁₁₄ N ₂ O ₂	CCCCN(C)CC
26		1011.63	C ₁₁₄ H ₁₁₈ N ₂ O ₂	CCCCN(C)CC
27		1043.69	C ₁₁₈ H ₁₂₂ N ₂ O ₂	CCCCN(C)CC
28		1075.74	C ₁₂₂ H ₁₂₆ N ₂ O ₂	CCCCN(C)CC
29		1107.80	C ₁₂₆ H ₁₃₀ N ₂ O ₂	CCCCN(C)CC
30		1139.85	C ₁₃₀ H ₁₃₄ N ₂ O ₂	CCCCN(C)CC

자사 합성신약 라이브러리 발취 (샘플)

Ⅲ 핵심 연구 내용 및 기술 (2단계)



FAM 모델 검증

- 상용 예측 소프트웨어 2개 보유
- 실험값/상용 예측값 과의 비교 검증
- Validation set 이용한 예측력 평가
- 전향적/후향적 보완 데이터 공급

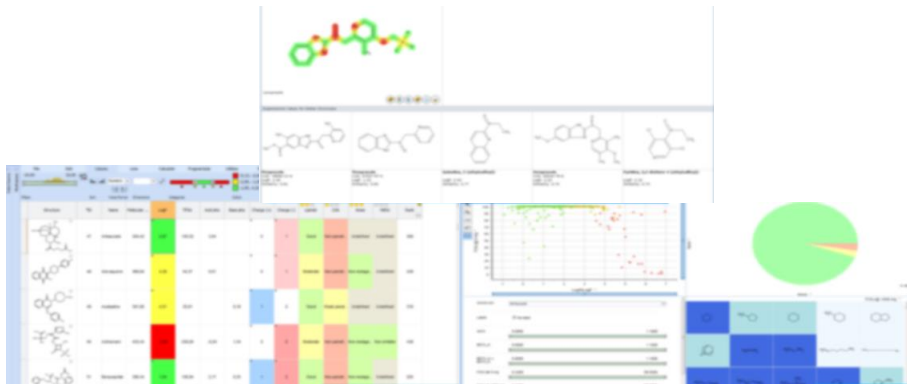
보완 데이터 공급

모델 검증 고도화

- 모델 검증 고도화
*예측력은 얼마나 상승되었는지,
상용 제품보다 예측력이 우수한지*

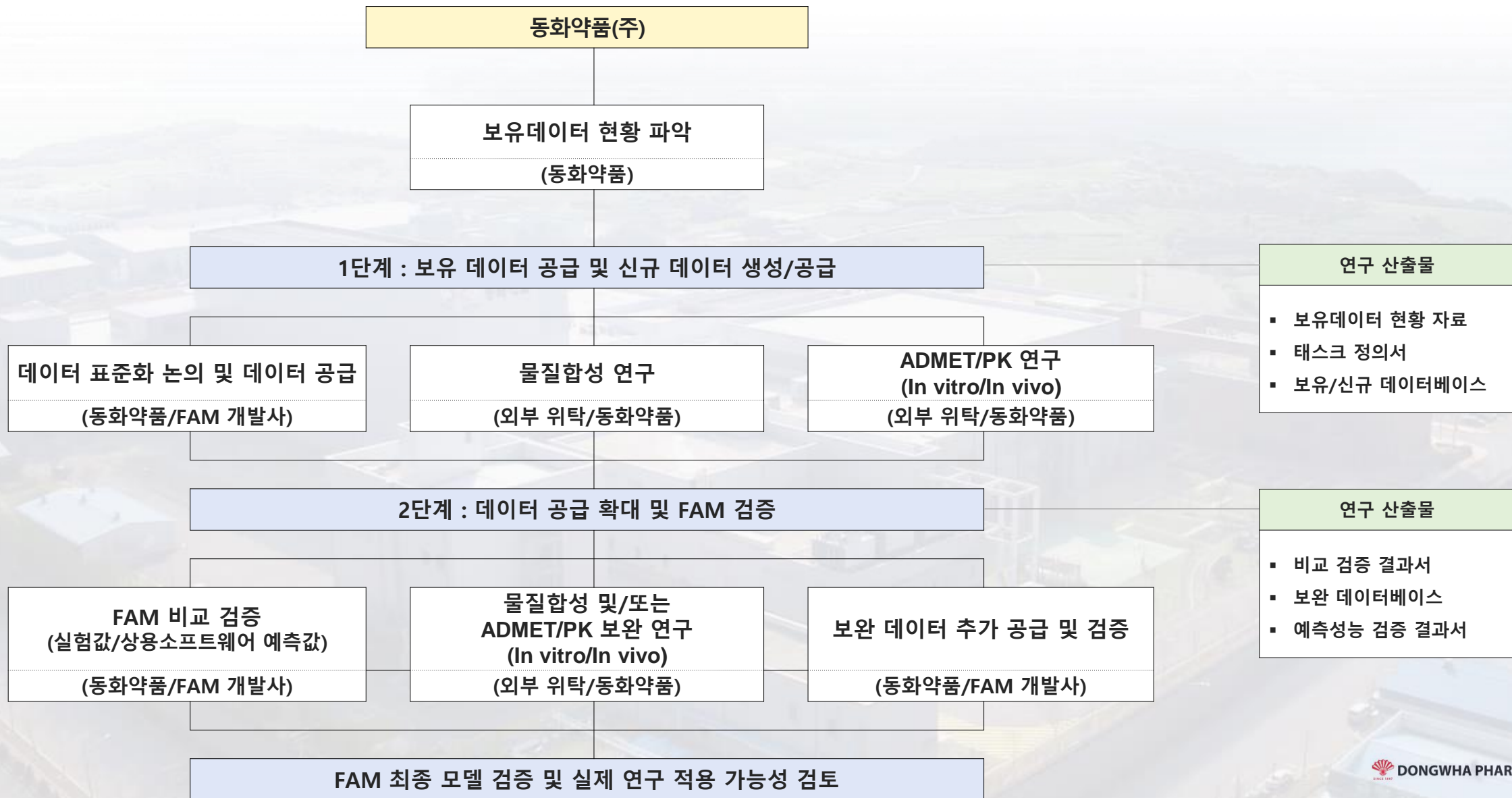
산업적 활용

- External set(신물질) 예측력 평가
- 실제 연구 적용 가능성 검토



자사 보유 소프트웨어 GUI 및
실제 실험값/예측값 비교 분석

IV 연구개발 추진 체계 및 연구산출물



V 연구 가능성 검토 및 활용 계획

파일럿 프로젝트 검증

- 목표 : FAM 모델 예측값과 자사 내부 데이터를 비교하여 정확도 평가
- 결과 : FAM 모델의 신약 개발 과제 적용에 대한 최종 검증



산업적 활용

- FDD 플랫폼을 활용하여 자체 신약 후보물질 발굴에 실제 적용 예정
- 상용 소프트웨어 대비 산업적/경제적 측면의 효율성/생산성 개선효과 확인 예정

