



내부 희귀골격 약물 라이브러리를 활용한 연합학습 기반 신약개발 가속화 연구

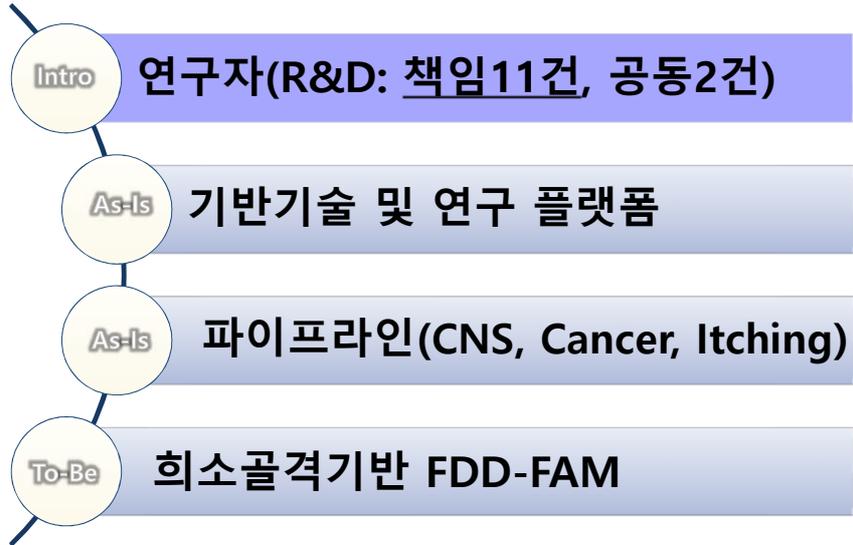
- 01 Background & Intro. **연구자 소개**
- 02 Platforms & Strategies **연구기반**
- 03 In-house IP Drug Pipelines **파이프라인**
- 04 Proposals & Plans **K-MELLODDY**

<https://sites.google.com/site/gacheondrlab/prof-mihyun-kim>
kmh0515@gachon.ac.kr
<https://orcid.org/0000-0002-2718-5637>



College of Pharmacy,
Gachon University

Mi-hyun Kim



Brief Intro.



학술 분야 다분야융합 연구 (chemistry, biology, data science & computer science for drug discovery)

- 1) 유기약품합성 화학, 의학화학 (Synthetic Organic Chemistry, Medicinal Chemistry)
- 2) 화학정보학 (Cheminformatics: CHEMISTRY+ INFORMATICS)

연구 주제 희소골격 약물 개발 (2D unprecedented 3D novel drug scaffolds with new methodologies)

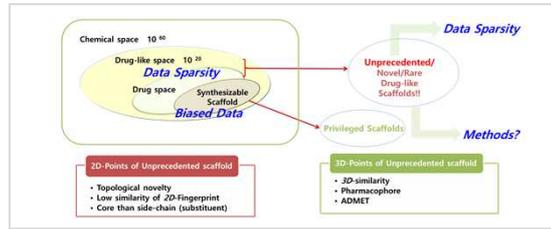
- 1) 타겟 스크리닝, 역설계 (target deconvolution & inverse design/optimization)
- 2) 중개 연구 organic chemistry (chemist's recognition), molecular featurization (machine's recognition), and molecular biology/pharmacology (biologist's recognition)

한국과학기술원원 차세대화학원 (의약학부 간사 역임)

대한약학회, 한국분자세포생물학회, 한국화학공학회, 한국 유기합성학회, 한국응용약물학회, 여성생명과학기술포럼
정회원/평생회원

한국계산과학공학회, 생화학분자생물학회, 한국생명정보 학회 발표 및 신약관련 국제 최정상 학술지 SCI JCR Top10% (컴퓨터공학, 화학, 정보학, 과학일반) 리뷰어/저자

Wiley Interdisciplinary Reviews: Computational Molecular Science	Nucleic Acids Research	PNAS
Journal of Medicinal Chemistry	Journal of Cheminformatics	Briefings in Bioinformatics



'α-to-w of chemistry'

Mi-hyun Kim

연구자 소개 (Brief CV): AI-DD R&D 과제 수행 이력

대표 연구 프로젝트	대표수행실적
분자 빅 데이터를 활용한 신규 신경 보호 활성 물질의 다약리 구조 활성 (MPSAR) 연구 (과기부- 5yr, 책임)	<ul style="list-style-type: none"> • 기존 CNS 약물과 차별적인 모핵 도출 • 3차원 약물 Big Data 제련, In-house DB 구축, API 작성 • AI/DT 기반 다중 타겟 활성 모델 (MPSAR model) 정립 • 총 21편 SCI논문(교신: 21편, 단독 교신: 13편, JCR Top10%: 5편) • 국내 특허 6건 출원, 인력양성 5명 (박사 4명, 석사 1명)
신중 및 유사 신경 화학작용제 관리를 위한 광범위 콜린성 작용 예측 모델 구축 (국방부- 3yr, 책임)	<ul style="list-style-type: none"> • AI기반 신경 화학작용제(Chemical warfare agent)의 콜린성 타겟 활성 예측 • CWA 판별 멀티 태스킹 (multi-tasking) 모델 설계 및 DB 구축 • 총 3편 SCI논문(단독 교신: 3편) • 국내 특허 1건 출원, 인력양성 3명 (연구교수 2명, 석사 1명)
의약품의 3차원 분자설계를 위한 멀티 컨포머 샘플링 알고리즘 (과기부- 1yr, 책임)	<ul style="list-style-type: none"> • 3D 디스크립터 개발을 위한 이형태체 (conformer) 샘플링 알고리즘 개발 • 총 2편 SCI논문(교신: 1편, 단독 교신: 1편) • 국내 특허 1건 출원, 석사 학위배출 2명
'합성법 지향'의 새로운 화합물 스케폴드(scaffold) 합성과 생체 내 타겟 발굴 (교육부- 5yr, 책임)	<ul style="list-style-type: none"> • 합성법 지향의 희귀 골격 화합물 도출 • 희귀 골격 화합물의 가능한 타겟을 도출하기 위한 기법 연구 • 희귀 골격 화합물의 합성법의 반응 메커니즘 연구 • 타겟별 인실리코 예측 모델 정립 • 총 12편 SCI논문(주저자: 9편, 교신: 6편, JCR Top10%: 1편) • 특히 국내 3건 등록, 해외 6개국 등록, 인력양성 4명 (석사 4명)

국가연구개발 사업 수행 현황 (책임 11건, 공동 2건; 63억)

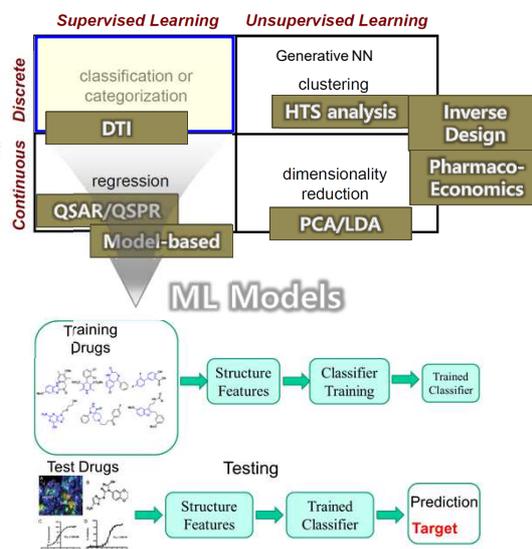
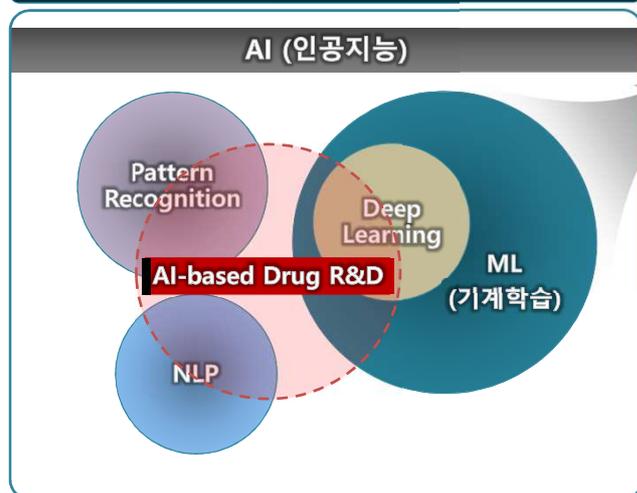
- K-MELLODDY Project(2024. 07 -2028. 12; 8.1억): 책임
- 단백질 응집 제어를 위한 펩타이드 프로브의 설계 및 위치 선택적 기능화 연구(2022. 09 -2027. 02; 9억)
- 중견연구자 지원 사업(NRF-기초): 책임
- 인공지능 기반 차세대 RET 표적항암제 개발 (2021. 06 -2023. 05; 2억), 인공지능 기반 RET 표적 항암제 개발 (2020. 06 -2021. 02; 0.5억)
- 중소벤처기업부(신약협력): 기관 책임
- 분자 빅 데이터를 활용한 신규 신경 보호 활성 물질의 다약리 구조 활성 연구 (2017. 11 -2022. 10; 12억)
- 4차 산업혁명 주제 전략과제(NRF-기초): 책임
- 알츠하이머 및 퇴행성 뇌질환을 타겟팅하는 방법으로서의 저산소증 유도인자(HIF) Prolyl Hydroxylase 억제제 개발 (2018. 06 -2022. 12; 3억)
- 해외우수신진연구자 유치(NRF-KRF): 책임
- 제 2유닛(3세부) 노인성 뇌질환 진단-치료를 위한 특화 방사성의약품과 분자영상시스템 개발 (2017. 01 -2019. 12; 11.6억)
- (가천대-김병원) 연구중심병원 사업 (복지부): 공동
- 신중 및 유사 신경 화학작용제 관리를 위한 광범위 콜린성 작용 예측 모델 구축 에 관한 연구 (2018.11 -2020. 12; 1.3억)
- 국방과학기술연구소 순수기초: 책임
- 효율적인 전분화능 줄기세포의 배양기술개발 (2017.05 -2018.12; 3억)
- 줄기세포연구 산업화 촉진지원(NRF-국책): 공동
- 신약개발 지식정보 DB 구축사업 (2016.03 -2016.10; 2억)
- (재)법부처신약개발사업단: 컨소시엄 총괄 책임
- 의약품의 3차원 분자설계를 위한 멀티 컨포머 샘플링 알고리즘 (2017.06 -2018.05; 0.5억) 개인연구자 지원 사업(NRF-기초): 책임
- '합성법 지향'의 새로운 화합물 스케폴드(scaffold) 합성과 생체 내 타겟 발굴(2012.09 -2017.08; 6.5억) 학문우수제대양성사업(NRF-기초): 책임

'α-to-w of chemistry'

Mi-hyun Kim

Intro.: Cheminformatics as (ML + DT) Methods

AI (Artificial Intelligence) > ML (Machine Learning)



'α-to-ω of chemistry'

Mi-hyun Kim

Intro.: Cheminformatics as (ML + DT) Methods

시니피에(signified, 記意)

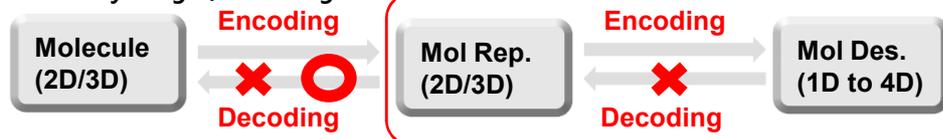
시니피양(signifier, 記表)



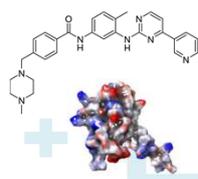
Apple

Cheminformatics as a part of Data Science : **Unique Featurization** of Molecular Data

Chemistry insight/knowledge

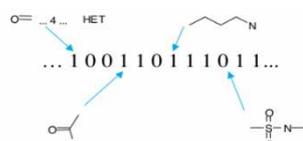


Features for ML Model



PLIP

e.g.) SMILES, InChI, SMARTS, Molecular Fingerprint (FP)



Featurization = Process to make & optimize features

Variables for Molecules
≈ Molecular Descriptors (Mol. Des.)
≈ Molecular Features

'α-to-ω of chemistry'

Mi-hyun Kim

Intro

연구자 소개 및 수행 사업

As-Is

기반기술 및 연구 플랫폼

As-Is

파이프라인(CNS, Cancer, Itching)

To-Be

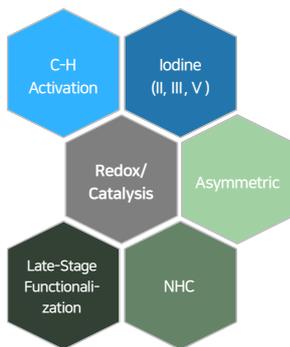
희소골격기반 FDD-FAM

' α -to- ω of chemical'

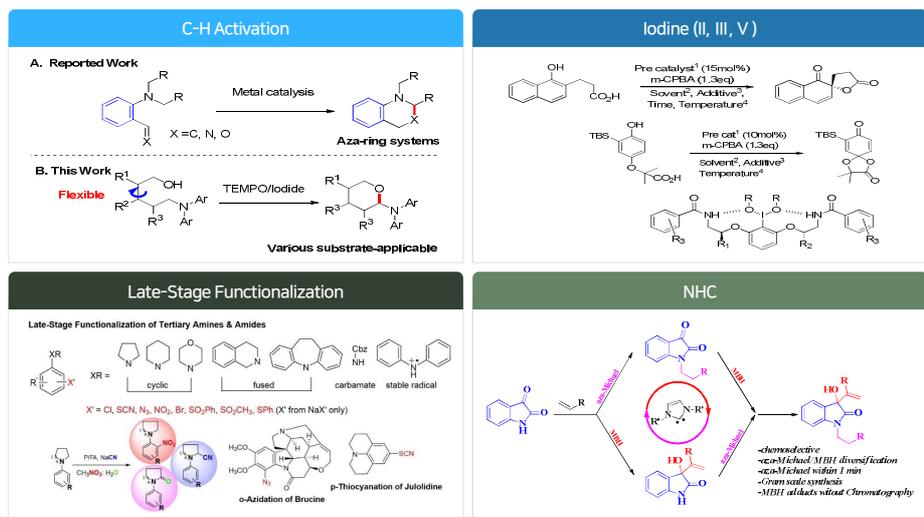
Mi-hyun Kim

플랫폼1: 약물성 희소골격 합성 (SDG/ESG 적합 방법)

Chemistry-oriented SYNTHESIS & unit synthetic methods for drug scaffolds



Asian J. Org. Chem. 2016,
ACS Omega 2018 & 2024
Org. Lett. 2019, Front. Chem. 2021,
Org. Biomol. Chem. 2020 & 2023
Sci. Rep., 2017, 2018, 2020, 2022, 2024
Eur. J. Med. Chem., 2019 & 2022

' α -to- ω of chemistry'

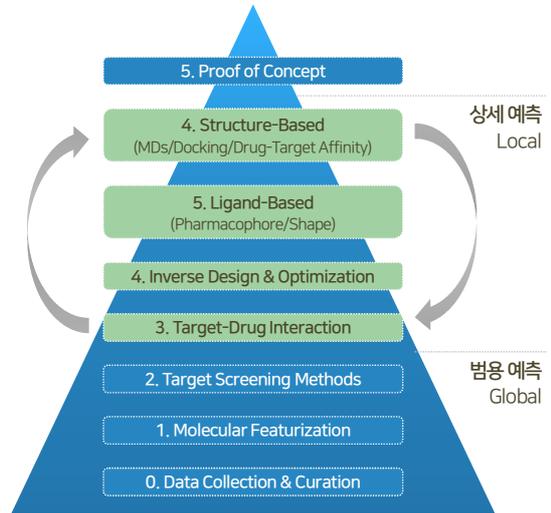
Mi-hyun Kim

플랫폼2: 머신러닝 기반 신약설계 플랫폼 I-V (예측/역설계)

- I 다질한 표적 맞춤형 활성 예측 모델
- II 약물과 표적 간 결합 친화도 예측 범용 예측 (DTA Model)
- III 표적 구조 없는 약물의 타겟/오프-타겟 프로파일 (DTI Model)
- IV 머신러닝 기반 역설계 플랫폼 (Inverse Design Model)
- V 다중 피쳐 기반 GPCR 약물 기능 및 활성 경로 예측 모델

출간 논문 및 연구자 정보: <http://orcid.org/0000-0002-2718-5637>

Conceptual Stages



'α-to-ω of chemistry'

Mi-hyun Kim

In-house AI 플랫폼 세부 대표 모델 분석

	SMPLIP-Score	ChargeNET	GPCR-IPL	KLD-DTI
경쟁사	다수 글로벌 시회사 발표, 아론티어		회사 경쟁사 없음	Transformer(챗GPT방식) 모델 출간-신약연구, 아론티어
데이터	약물-타겟 복합체 구조		약물-GPCR 복합체 구조	약물의 2차원/3차원 구조
피쳐화 방법 (변수 생성법)	잔기별 약물과의 3D 결합여부, 포개진 약물 구조	결합자리 내 Grid 생성 및 물리화학적 에너지 계산	3중 피쳐 (INT, POCK, LIG)	2D/3D-약물 유사도 분포값의 상대 엔트로피 (KLD)
학습범위	>8종류 Benchmark Dataset 검증	Benchmark Dataset 검증	GPCR-Exp/GPCR-DB/ChEMBL	ChEMBL DB
예측범위	결합친화도 예측		결합 후 기능 예측	직/간접적 관계 유무 예측
RWD 검증 및 활용	다중돌연변이 RET 저해제 구조최적화에 활용		소양증 치료제 설계에 활용	뇌염증 억제약물 분석에 활용
공개	J. Cheminformatics (2021), Top10%	비공개	Briefings in Bioinformatics (2024), Top1%	J. Cheminformatics (2022), Top10%

'α-to-ω of chemistry'

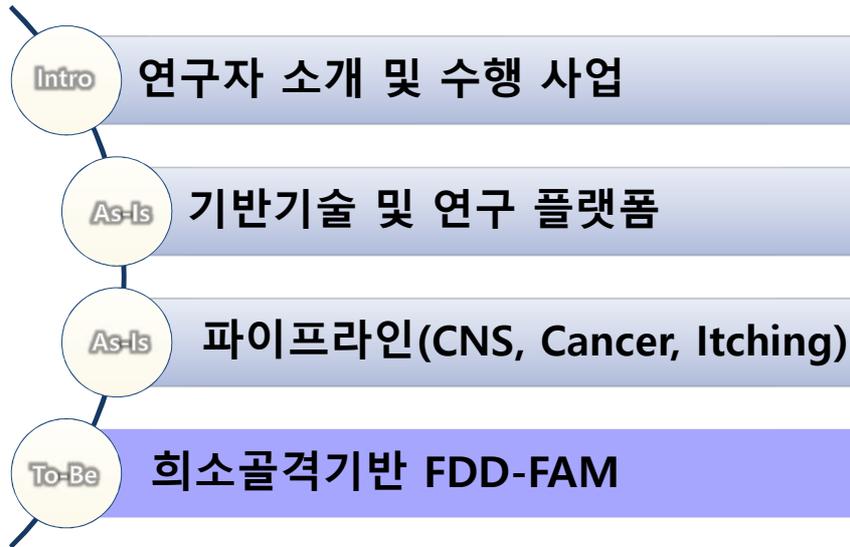
Mi-hyun Kim

희소골격 약물 파이프라인 및 공개된 보유 특허

<p>보유 파이프라인 1</p> <p>당모사 기법으로 개발된 신경독성 억제제 등록특허 (가진대 단독) 국내 등록 10-1881115 경쟁우위 뇌조직 지향성이 높은 당모핵에 기반하여 개발됨 (모핵특허 >> 물질 그 자체)</p>
<p>보유 파이프라인 2</p> <p>희소 약물 골격에 기반한 뇌염증 억제 치매치료제 등록특허 (가진대/일병원) 7개국 등록 10-1850062 경쟁우위 in vivo 인지능력개선, nanomolar 뇌염증 억제능, 희귀 모핵 물질 원천성 및 IP 확보</p>
<p>보유 파이프라인 3</p> <p>베타 차단제를 모사한 아세틸콜린에스테라제 저해제 등록특허 (가진대 단독) 국내 등록 10-2104804 경쟁우위 신경보호 효능과 아세틸콜린에스테라제 저해능 이중 기능</p>
<p>보유 파이프라인 4</p> <p>ATP 비경쟁적 GSK3β 알로스테릭 저해제 기반 다중 질환 치료제 (기분장애, 자폐증, 치매, 치매진단-Theranotics) 출원특허 (가진대/울산대/이산병원) 10-2018-0170985 경쟁우위 GSK3β 저해효능과 염증억제능 보유</p>
<p>보유 파이프라인 5</p> <p>SMYD2 저해에 기반한 삼중음성 유방암/비소세포성 폐암 타겟팅 항암제 (First-in-Class) 출원특허 (가진대 단독) 10-2022-000039 경쟁우위 아스트라제네카 AZ-505 보다 우수한 in vivo/in vitro 효능과 차별적인 약리 스펙트럼</p>

'α-to-ω of chemistry'

Mi-hyun Kim



'α-to-ω of chemical'

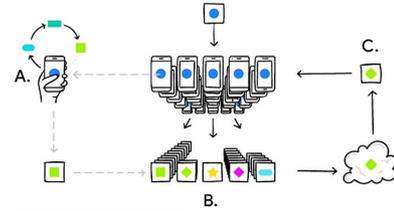
Mi-hyun Kim

01 연구의 필요성: 연합학습(FL)의 개요와 효용성

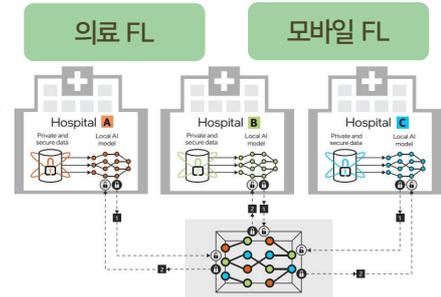
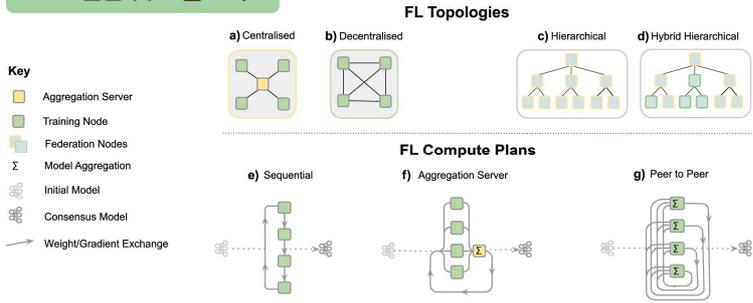
K-MELLODDY for AI/DT-DDI
Mi-hyun Kim

❖ 연합학습(FL): 데이터 보안을 유지하면서도 중앙 집중식 모델 성능에 필적하는 학습 성능을 도출

- 글로벌 파라미터(ω^G) 토폴로지 구성: Centralized, Decentralized, Hierarchical
- 연산 방법: Sequential, Aggregation, Peer-to-Peer
- 장점: 프라이버시 보호, 데이터 이동 불필요, 네트워크 효율성, 고성능
- 의료, 금융, 모바일 증강현실 서비스



FL 연산 및 토폴로지



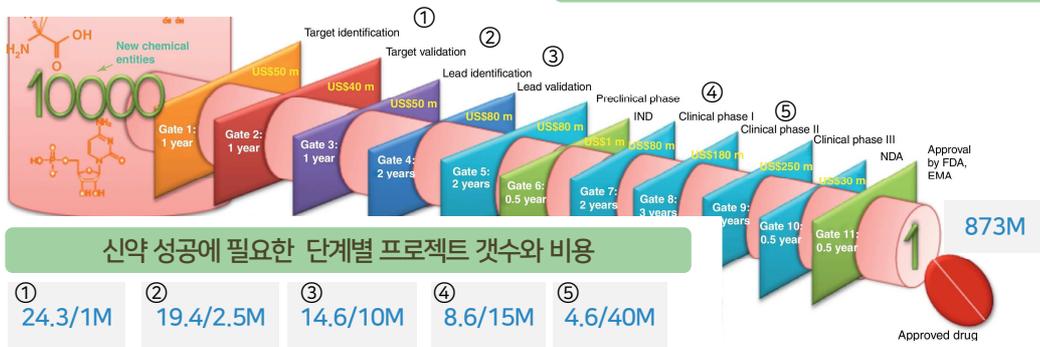
01 연구의 필요성: 제약바이오 데이터의 특성 및 FL 필요성

K-MELLODDY for AI/DT-DDI
Mi-hyun Kim

❖ 현 신약개발 패러다임과 신약 R&D 데이터 특성

- 단계 의존성(stage-dependency): 단계별 데이터간 이질성(heterogeneity), 분절성(fragmentation) 발생
- 스크리닝 및 탈락방식에 의한 최적화: 데이터 희소성(data sparsity)
- TPP(Target Product Profile) 기반 프로젝트 진행 및 잦은 중단(Go/NoGo): 데이터 무결성(data integrity) 이슈
- 바이오의료 및 신약개발 데이터는 data point가 상대적으로 작고, 관련 실험결과변수가 많은 데이터임 (EU-ChEMBL: 2.2M여개 약물, 미국-PubChem: 111M여개 화합물)
- 데이터 생산 단가가 산업분야 중 가장 높음

다출처 데이터를 활용하는 연합학습이 필요 !!!

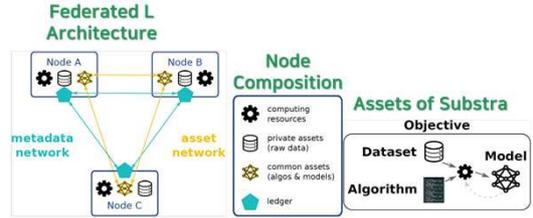


01 연구의 필요성: EU MELLODDY 프로젝트

K-MELLODDY for AI/DT-DDI
Mi-hyun Kim

연합학습 플랫폼 Substra

- 상중음성 유방암(TNBC)의 **신규 보조요법(neoadjuvant therapy)의 조직학적 반응을 예측**하는데 활용
- 구성
 - ① 자산(asset): 데이터셋, **알고리즘(제반 방법), 모델(학습 완료)**, 목적(objective)
 - ② 인터페이스: 접근 권한을 관리하는 '웹페이지', 자산을 추가 및 관리하는 CLI 및 SDK
 - ③ 이종 네트워크 구조: 자산 네트워크(asset network)와 메타데이터네트워크(metadata network)
 - ④ 렛저(ledger): 분산 원장 기술 (Distributed Ledger Technologies, DLT)이 적용된 자산관리체계

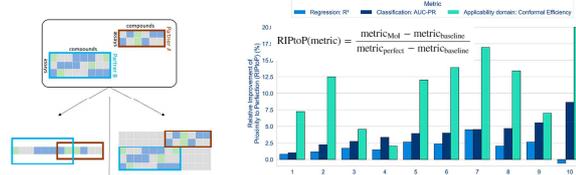


연합학습을 위한 데이터 가공 및 전처리 도구

- [MELLODDY-TUNER]
- 개인정보 보호 머신러닝을 위해 개발된 도구(전처리, 가공, 데이터 분할)
 - TFF(TensorFlow), PySyft, Flower, FedML 등 FL에 적합한 도구
- [SparseChem Library]
- 고차원 데이터에서 데이터 희소성(sparsity)을 처리하기 위해 설계된 기계 학습 도구
 - 희소 행렬 분해 및 모델 교정
- $$ECE = \sum_{m=1}^M \frac{|B_m|}{n} |Accuracy(B_m) - Confidence(B_m)|$$
- [Chemfold]
- 학습 성능 개선에 기여하기 위한 **데이터분할 도구**
 - 스피어 배제 클러스터링(sphere exclusion clustering)
 - LSH기반 분할 및 스캐폴드 트리(scaffold tree)기반 데이터 분할

연합 학습된 QSAR(구조-활성 관계) 멀티 태스킹 모델링

- >26억 기밀 R&D 데이터로부터 cross-compound federation하여 100개의 endpoint 사용
- >40,000 cross-endpoint federation (on-target 활성, 2차 약리학, 약동학) 획득
- 성능 평가 지표: FL 전후 RlPtP(Relative Improvement of Proximity to Perfection) 측정
- **5%내외 예측력(predictive power) 개선, 예측 범위(applicability domain)의 현격한 개선**



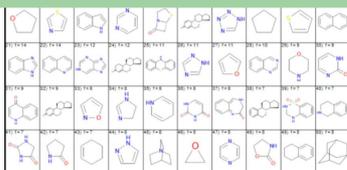
01 연구의 필요성: K-MELLODDY 를 위한 핵심 제언

K-MELLODDY for AI/DT-DDI
Mi-hyun Kim

골격 다양성을 갖는 차별적 데이터 확보

AS-IS:
다빈도 모핵(privileged scaffold),
Congeneric Data

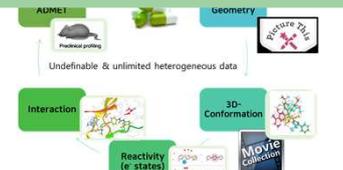
다빈도 vs 희소골격



2D/3D-분자표상과 피쳐 임베딩

AS-IS:
2D-Molecular Graph
(SMILES, Fingerprint)

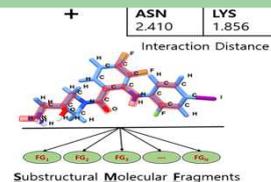
3D-Geometry/Electronic 피쳐 추가



FAM 모델의 핵심 태스크: Off-Target Toxicity

AS-IS:
Ligand 기반 예측에 충실한
QSAR/QSPR 모델

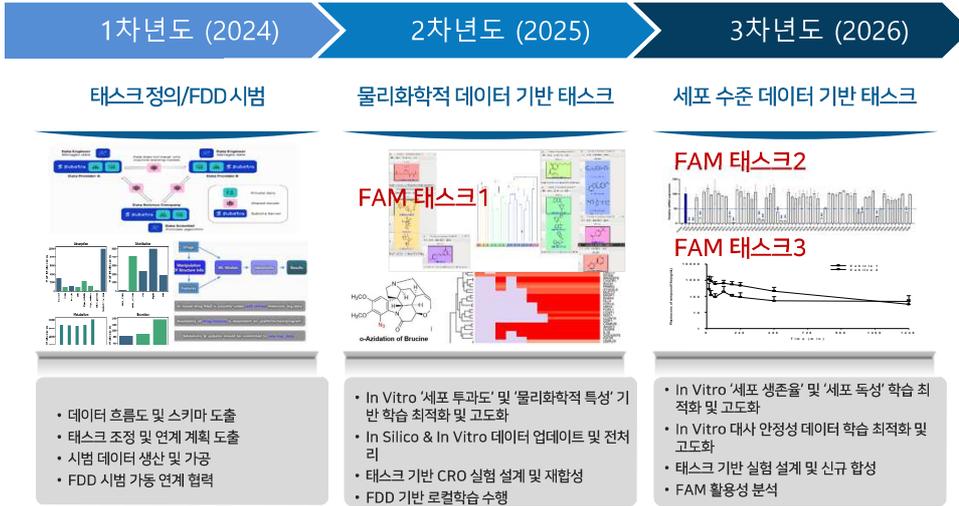
생성형AI/PLIP 패턴 기반 예측



02 연구 계획: 1단계 연차별 목표

K-MELLODDY for AI/DT-DDI
Mi-hyun Kim

기확보 희소골격 데이터 활용 FDD 구축 및 데이터 생산 확대



02 연구 계획: 2단계 연차별 목표

K-MELLODDY for AI/DT-DDI
Mi-hyun Kim

확보 데이터 기반 FAM 태스크 고도화

