

“학습하지 않은 약물도 반응 예측” GIST, 신약 후보물질 선별 돕는 AI 개발

- AI학과 이현주 교수팀, 약물의 다양한 화학적 특징을 분석하고 예측에 유용한 정보를 선별하는 AI 모델 'InFo-DRP' 개발해 학습하지 않은 새로운 약물의 반응도 예측
- 기존에 학습한 약물과 핵심 화학구조가 다른 약물 평가에서 예측 상관계수 35.8% 향상... 신약 후보물질 선별 및 실험 우선순위 결정에 활용 기대
- 국제학술지《IEEE Journal of Biomedical and Health Informatics》게재



▲ (왼쪽부터) GIST AI학과 이현주 교수, 이지상 석박사통합과정생, 염태림 석박사통합과정생, 원광대학교 전자공학과 김준우 교수

광주과학기술원(GIST·지스트, 총장 임기철)은 **AI학과 이현주 교수 연구팀**이 학습 과정에서 접하지 않은 **새로운 약물에 대해서도 암세포의 반응을 보다 정확하게 예측할 수 있는 인공지능(AI) 모델 '인포-디알피(InFo-DRP)'**를 개발했다고 밝혔다.

이 모델은 **약물의 화학적 특징을 여러 방식으로 분석하고, 새로운 약물의 반응을 예측하는 데 유용한 정보를 선별하도록** 설계됐다. 특히 **AI가 학습한 약물과 핵심 화학구조가 다른 약물을 대상으로 한 평가에서 기존 모델보다 높은 예측 성능을 보였다.**

같은 항암제라도 **암세포의 유전적 특성에 따라 약물에 대한 반응은 달라질 수 있다.** 이에 따라 약물의 화학구조와 암세포의 유전자 정보를 분석해 약물에 대한 세포의 반응을 미리 예측하는 AI 기술이 연구되고 있다.

그러나 기존 AI 모델은 학습에 사용한 약물의 반응은 잘 예측하더라도, **처음 접하는 약물이나 화학구조가 크게 다른 약물에서는 예측 성능이 떨어지는 한계가 있었다.** 실제 신약개발에서는 실험 데이터가 충분하지 않은 새로운 후보물질을 평가해야 하므로 이러한 한계를 해결하는 것이 중요하다.

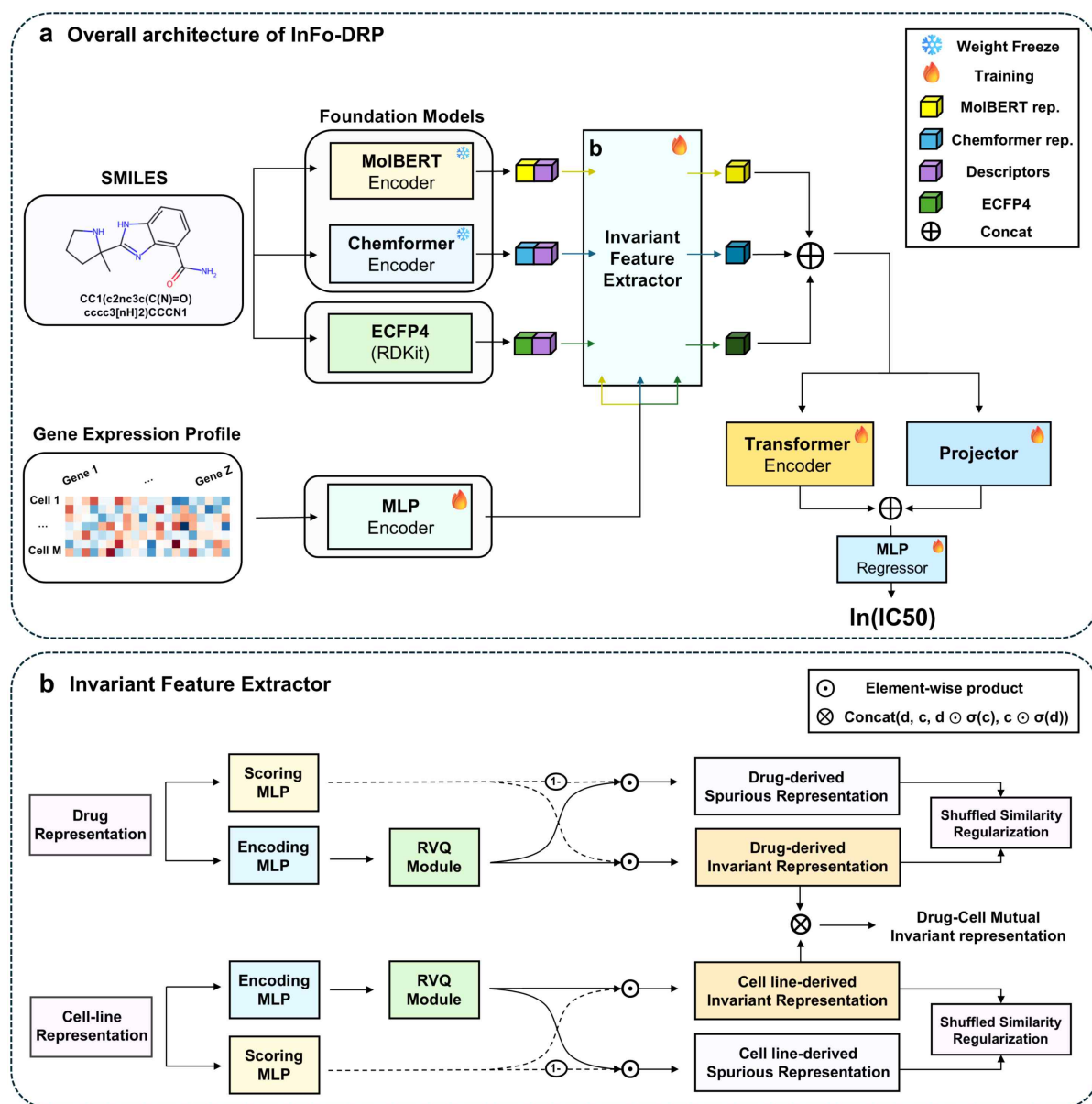
연구팀은 AI가 학습에 사용한 약물의 특징에 지나치게 의존하지 않도록, **약물의 화**

화학구조와 성질을 여러 방식으로 분석하고 예측에 필요한 정보를 선별하는 방법에 착안했다.

이를 위해 다양한 분자의 화학구조를 미리 학습한 AI 모델을 활용하고, 약물의 세부 구조와 물리화학적 특성도 함께 분석했다. 하나의 분석 방식에 의존하지 않고 서로 다른 화학 정보를 종합해 새로운 약물의 반응을 더 잘 예측하기 위해서다.

또한 '불변 학습' 기술을 적용해 학습에 사용한 약물에만 나타나는 특징의 영향을 줄이고, 새로운 약물에서도 예측에 활용할 수 있는 정보를 추출하도록 했다. 이렇게 얻은 약물 정보와 암세포의 유전자 발현 정보를 결합해 반응을 예측하도록 설계한 모델이 'InFo-DRP'다.

* 불변 학습(invariant learning): 특정 학습 데이터에만 의존하는 정보보다 데이터 환경이 달라져도 일관되게 유지되는 예측 정보를 학습하도록 하는 방법이다.



▲ 약물 반응 예측 AI 'InFo-DRP'의 작동 원리. (a) 약물의 화학구조와 세포의 유전자 발현 정보를 각각 분석한 뒤, 약물 반응을 예측하는 전체 과정을 나타낸다. 약물의 특징은 미리 학습된 AI 모델과 화학구조 분석 기술을 활용해 여러 방식으로 추

출한다. (b) '불변 특징 추출기'의 작동 원리로, 약물과 세포의 정보에서 새로운 약물의 반응을 예측하는 데 유용한 특징을 선별하는 과정을 보여준다. 이렇게 얻은 정보를 종합해 약물에 대한 세포의 반응을 예측한다.

연구진은 InFo-DRP가 처음 접하는 약물의 반응도 잘 예측하는지 확인하기 위해 **대규모 암세포 약물 반응 데이터베이스인 GDSC(Genomics of Drug Sensitivity in Cancer)**를 활용해 **기존 AI 모델 6종과 성능을 비교했다.**

먼저 **학습에 사용하지 않은 약물을 대상으로 평가했다.** 이어 학습한 약물과 **핵심 화학 골격이 겹치지 않는 약물로 더욱 엄격한 평가를** 진행했다. 이를 통해 AI가 기존에 학습한 **화학구조에 의존하지 않고** 새로운 약물의 반응을 예측할 수 있는지 확인했다.

평가 결과, InFo-DRP는 두 조건 모두에서 비교 모델 가운데 **가장 높은 예측 상관계수(PCC)*를 기록했다.**

특히 **핵심 화학 골격이 다른 약물을 대상으로 한 평가에서 예측 상관계수는 0.592**로, 기존 비교 모델 중 최고치인 0.436보다 **35.8% 높았다.** 학습에 사용하지 않은 약물을 대상으로 한 평가에서도 **0.617**을 기록해 기존 모델의 최고치인 **0.489**보다 **26.2% 높았다.**

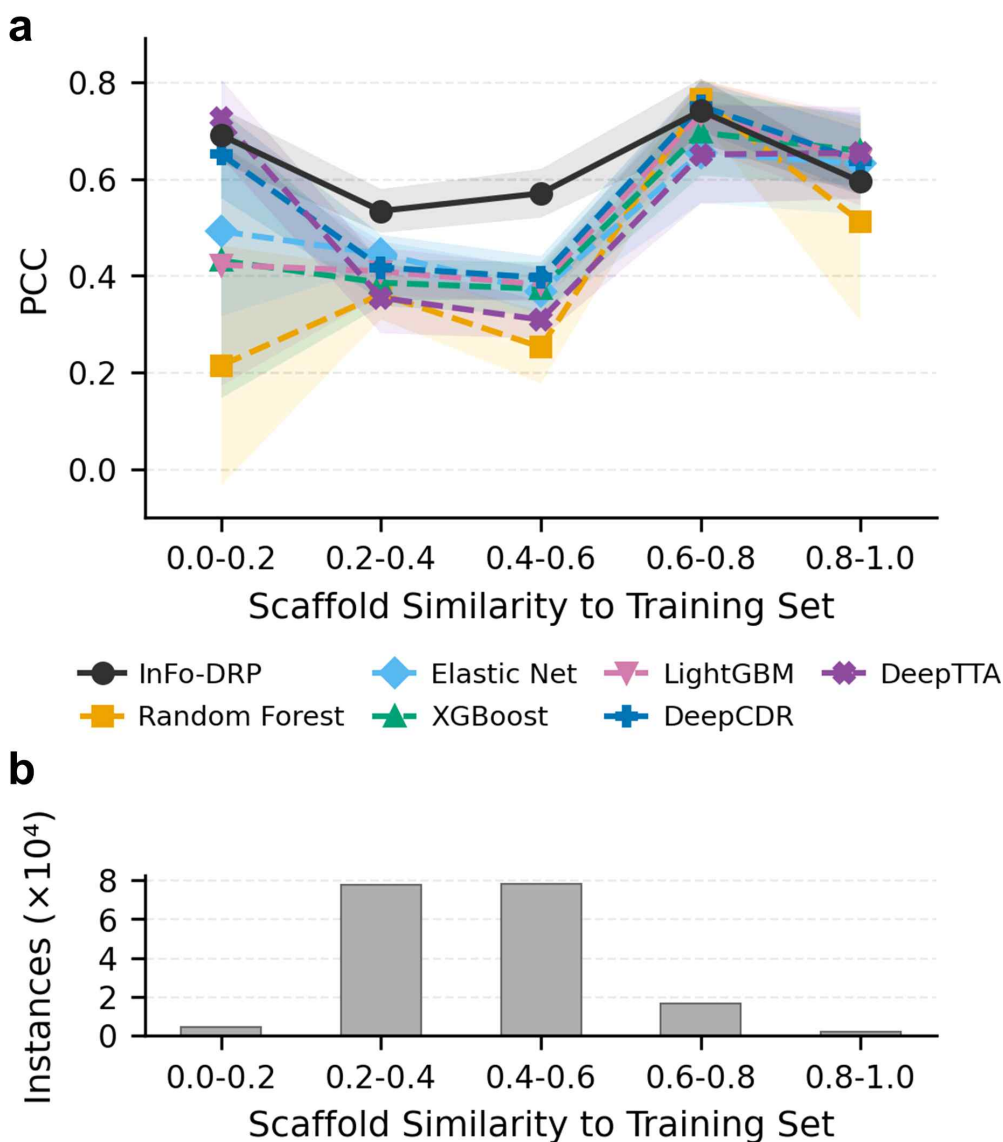
두 평가에서 기존 AI 모델 6종보다 높은 예측 성능을 보였으며, 이러한 차이는 통계적으로도 유의한 것으로 확인됐다.

* **예측 상관계수(PCC):** AI가 예측한 약물 반응과 실험에서 측정한 실제 반응이 얼마나 비슷한 경향을 보이는지 나타내는 지표로, 1에 가까울수록 예측값과 실제값의 상관성이 높다.

이어 **AI 모델 InFo-DRP의 예측 성능을 다른 종류의 세포와 별도의 데이터베이스에서도 검증했다.**

암세포 이외의 세포에서도 약물 반응을 예측할 수 있는지 확인하기 위해 **비(非)암성 세포주 5종을 대상으로 평가한 결과, 모두 기존 비교 모델보다 높은 예측 상관성을** 보였다.

또한 GDSC 데이터로 학습한 InFo-DRP에 **독립적으로 구축된 약물 반응 데이터베이스인 CCLE(Cancer Cell Line Encyclopedia)의 약물·세포 정보를 입력해 반응을 예측**하도록 했다. CCLE 데이터로 추가 학습하거나 보정하지 않은 상태에서도 **예측 상관계수 0.431**을 기록해 기존 비교 모델 중 최고치인 0.417보다 높았다.



▲ **화학구조 유사도에 따른 AI 모델별 약물 반응 예측 성능 비교.** (a) 평가에 사용한 약물과 학습에 사용한 약물 사이의 핵심 화학구조 유사도에 따라 AI 모델 7종의 예측 상관계수(PCC)를 비교한 그래프다. 검은색 실선이 InFo-DRP로, 평가 데이터가 많이 분포한 유사도 0.2~0.6 구간에서 기존 모델보다 높은 예측 상관성을 보였다. (b) 각 유사도 구간에 포함된 평가 데이터의 수를 나타내며, 데이터가 유사도 0.2~0.6 구간에 집중돼 있음을 보여준다.

이현주 교수는 “이번 연구는 **AI가 학습하지 않은 새로운 약물이나 기존에 학습한 약물과 화학구조가 다른 약물에서도 예측 성능을 높였다**는 점에서 의미가 있다”며 “향후 신약개발 초기 단계에서 후보물질의 반응을 미리 예측하고 실험 대상의 우선 순위를 정하는 등 **신약 후보물질 선별의 효율을 높이는 데 활용될 것으로 기대한다**”고 말했다.

GIST AI학과 이현주 교수가 지도하고 이지상 석박사통합과정생이 제1저자, 염태림 석박사통합과정생과 원광대학교 전자공학과 김준우 교수가 공저자로 참여한 이번 연구는 ▲과학기술정보통신부·한국연구재단 기초연구사업 ▲산업통상자원부·한국산업기술진흥원 산업혁신기반구축사업 ▲과학기술정보통신부·정보통신기획평가원 AI

연구 결과는 **생명의료정보학 분야 국제학술지 《IEEE Journal of Biomedical and Health Informatics》**에 2026년 9월 30일 온라인으로 게재됐다.

한편 GIST는 이번 연구 성과의 기술사업화 및 산업적 활용에 관심 있는 기업·기관의 문의를 **창업혁신진흥원 기술사업화센터(hgmoon@gist.ac.kr)**를 통해 받고 있다고 밝혔다.

논문 정보

○ 논문명, 저자 정보

- 저널명 : IEEE Journal of Biomedical and Health Informatics
(IF: 7.7, 수리·계산생물학 분야 JCR 상위 6.7%, 2025년 기준)
- 논문명 : InFo-DRP: Integrating Invariant Learning and Foundation Representations for Unseen Drug Response Prediction
- 저자 정보 : 이지상(제1저자, GIST AI학과 석박사통합과정생), 염태림(공저자, GIST AI학과 석박사통합과정생), 김준우(공저자, 원광대학교 전자공학과 교수), 이현주(교신저자, GIST AI학과 교수)