

# “같은 암이라도 다른 약 써야”

## GIST-서울대 공동연구팀, 암환자 약물 반응성 예측 AI 모델 개발... ‘개인 맞춤형 치료’ 기대

- GIST 이현주 교수팀 개발 ‘PANCDR’, 세포주 데이터 학습 모델의 한계 극복해 환자 데이터에 적용해도 정확도 높아... 기존 알고리즘 대비 예측 성능 34% 향상
- 뇌종양 환자 데이터 통해 정확도 및 신뢰도 검증... 생명정보학 분야 국제학술지 <Briefings in Bioinformatics> 게재



▲ (왼쪽부터) AI대학원 이현주 교수, 전기전자컴퓨터공학부 김주연 연구원

인공지능(AI)을 활용한 신약 개발은 신약 후보군을 신속히 발굴함으로써 시간을 단축하고 임상 성공 확률을 높일 수 있다는 장점이 있어 최근 활발히 시도되고 있다. 이러한 가운데 AI 모델을 활용해 암환자의 약물 반응을 예측하는 기술이 국내 연구진에 의해 개발돼 개개인의 특성을 고려한 맞춤형 치료에 한 발짝 다가설 수 있게 됐다.

광주과학기술원(GIST, 총장 임기철)은 AI대학원 이현주 교수 연구팀이 사람의 유전자 발현 정보와 약물 그래프 정보를 기반으로 암환자의 약물 반응을 예측하는 인공지능 모델을 개발했다고 밝혔다.

연구팀이 개발한 인공지능 알고리즘은 세포주\* 데이터로 학습한 모델을 통해 암환자의 약물 반응성을 높은 정확도로 예측할 수 있어 적합한 후보 약물 추천에 의한 ‘환자 맞춤형 치료’에 기여할 것으로 기대한다.

\* 세포주(Cell line): 균일한 조직에서 유래된 세포의 집단으로, 동일한 유전적 특징을 가지는 세포의 계통을 의미한다. 약에 대한 저항력 등을 연구하는 데 이용된다.

동일한 유형의 암 환자에 같은 약물을 사용하더라도 개인의 유전적 특성이나 암세포의 돌연변이에 따라 약물의 반응이 달라질 수 있다.

각 개인에게 맞는 약물을 찾기 위해서는 **정확한 약물 반응 예측이 중요하기 때문**에 최근에는 머신러닝이나 딥러닝 같은 **인공지능 기법을 사용해 약물의 반응을 예측하려는 연구가 활발히** 진행되고 있다.

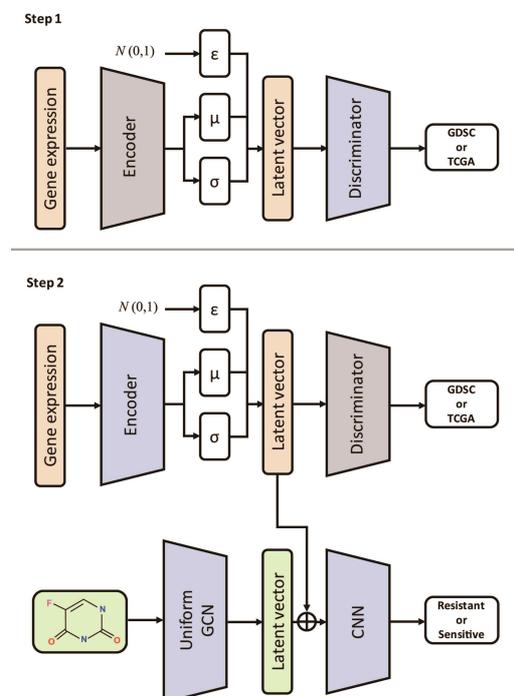
대부분의 약물 반응 예측 연구에서는 약물 반응 정보가 존재하는 환자 데이터의 수가 부족하여 데이터가 충분히 많은 세포주 데이터로 모델을 학습시킨다. 그러나 세포주 데이터는 면역계, 혈관계 등이 존재하지 않다는 점에서 환자 데이터의 유전자 발현량 정보와는 큰 차이가 있다.

따라서 **세포주 데이터로 학습시킨 모델을 환자 데이터에 적용했을 때 정확성이 낮아지는 한계**가 있다.

연구팀은 **적대적 생성 신경망(GAN)\*을 활용**하여 인공지능 모델에서 세포주 데이터와 환자 데이터 상호 간 표현(representation)의 차이를 줄임으로써 **세포주 데이터로 학습하더라도 환자 데이터에서도 정확한 약물 반응을 예측할 수 있도록 한 모델 'PANCDR(Precision medicine prediction using an Adversarial Network for Cancer Drug Response)'을 개발**했다.

\* **생성적 적대 신경망(Generative Adversarial Network, GAN):** 기존의 데이터를 모방해 새로운 데이터를 만드는 알고리즘으로 두 개의 모델이 서로 목표를 달성하기 위해 적대적으로 겨루는 구조를 지니고 있다.

연구팀이 개발한 'PANCDR' 모델은 판별자(discriminator)와 약물 반응 예측 모델을 번갈아 가며 학습시키는데, 1단계에서는 가우시안 인코더(Gaussian encoder)가 인코딩한 잠재 벡터(latent vector)가 세포주의 유전자 발현 데이터에서 온 것인지 환자의 유전자 발현 데이터에서 온 것인지 구분하는 판별자를 학습시킨다.

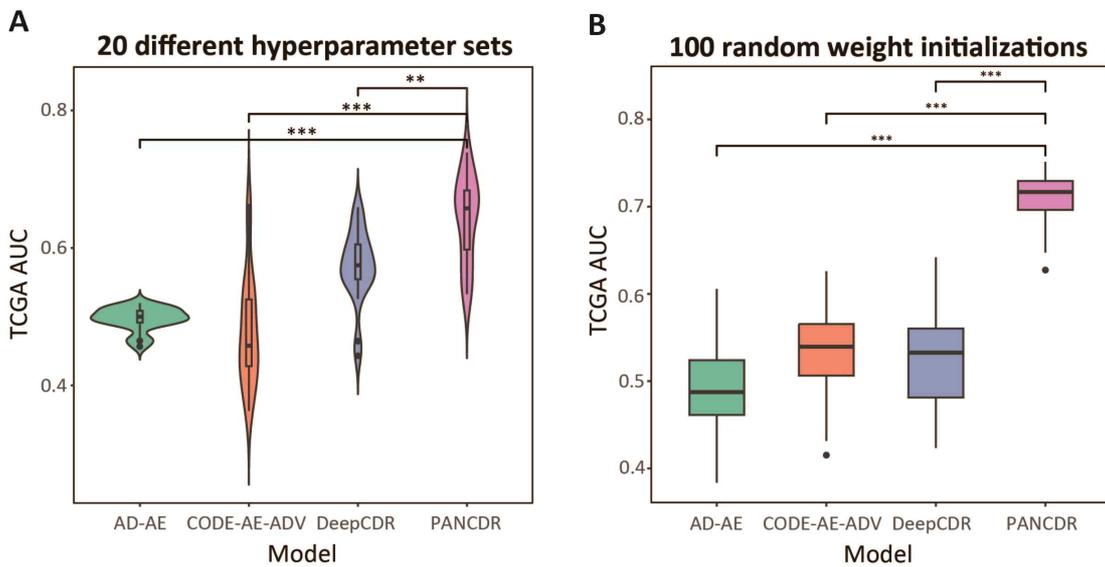


▲ **PANCDR 모델 구조** 2단계로 학습하는 모습을 보여주고 있다. 1단계에서는 판별자가 어느 데이터에서 온 잠재 벡터인지 구분하도록 학습한다. 2단계에서는 판별자를 속이는 동시에 약물 반응 예측 모델을 학습시킨다.

2단계에서는 반대로 판별자가 어느 데이터에서 온 것인지 구분하지 못하도록 약물 반응 예측 모델을 학습시킨다. 이때 환자의 데이터는 **유전자 발현 데이터만 있고 약물 반응성이 없는 대규모의 데이터를 활용**하였다.

'PANCDR' 모델(AUC\* 0.7106)은 환자 데이터에서 기존의 약물 반응 예측 모델(AUC 0.5273)보다 34% 이상 뛰어난 예측 성능을 보였다.

\* **AUC(Area Under the ROC Curve)**: ROC(Receiver Operating Characteristic) curve 아래의 면적을 의미하며, 분류 모델의 성능을 나타낸다.



▲ **기존 딥러닝 모델과 PANCDR의 성능 비교.** (A) 여러 조합의 하이퍼 파라미터를 사용했을 때, PANCDR이 다른 모델들에 비해 더 높은 성능을 보인다. (B) 초기 가중치(weight)를 랜덤하게 100번 바꿔 학습시켰을 때 PANCDR이 다른 모델들에 비해 더 높고 안정적인 성능을 보인다.

연구팀은 'PANCDR' 모델을 **서울대병원 연구팀(박성혜 교수)의 소아 뇌종양 환자 데이터에 적용하여 반응성이 가장 높게 예측된 상위 5개의 약물을 선정**했다. 그리고 이에 관한 기존 연구를 조사한 결과, **5개 약물 모두 뇌종양과 관련되어 있음을 확인하여 'PANCDR' 모델의 정확도와 신뢰도를 검증**하였다.

이현주 교수는 "이번 연구 성과를 통해 **세포주 데이터로 약물 반응 모델을 학습하더라도 환자 데이터에서 높은 정확도로 예측하는 것이 가능하다**"며 "향후 개인 맞춤 치료를 위한 **정확한 약물 반응 예측을 제공할 것으로 기대된다**"고 말했다.

GIST AI대학원 이현주 교수가 지도하고 김주연 연구원이 수행한 이번 연구는 서울대학교 의과대학 병리학교실 박성혜 교수와의 공동연구로, 정보통신기획평가원(IITP)의 지원을 받았으며 생명정보학 분야 국제학술지 'Briefings in Bioinformatics'에 2024년 3월 14일 게재됐다.

# 논문의 주요 정보

## 1. 논문명, 저자정보

- 저널명 : Briefings in Bioinformatics (IF: 9.5, 2022년 기준)
- 논문명 : PANCDR: Precise Medicine Prediction using an Adversarial Network for Cancer Drug Response
- 저자 정보 : 김주연 (제1저자, 전기전자컴퓨터공학부 석사), 박성혜 교수 (공저자, 서울대학교 의과대학), 이현주 교수 (교신저자, AI대학원, 전기전자컴퓨터공학부)