“Even for the same cancer, different drugs must be used”
GIST-Seoul National University joint research team, drug responsiveness in cancer patients Development of predictive AI model... Expectation of ‘personalized treatment’

- ‘PANCDR’, developed by GIST Professor Hyunju Lee Lee’s team, overcomes the limitations of cell line data learning models and achieves high accuracy even when applied to patient data... 34% improvement in prediction performance compared to existing algorithms
- Verification of accuracy and reliability through brain tumor patient data... Published in <Briefings in Bioinformatics>, an international academic journal in the field of bioinformatics

▲ (From the left) Professor Hyunju Lee of the AI Graduate School and Juyeon Kim, a researcher in the School of Electrical Engineering and Computer Science

New drug development using artificial intelligence (AI) has been actively attempted recently as it has the advantage of shortening time and increasing the probability of clinical success by quickly discovering new drug candidates. Meanwhile, a technology to predict drug reactions in cancer patients using AI models has been developed by domestic researchers, bringing us one step closer to customized treatment that takes into account individual characteristics.

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced that Professor Hyunju Lee's research team at the AI Graduate School developed an artificial intelligence model that predicts drug response in cancer patients based on human gene expression information and drug graph information.

The artificial intelligence algorithm developed by the research team is expected to contribute to ‘patient-tailored treatment’ by recommending appropriate candidate drugs as it can predict cancer patients’ drug responsiveness with high accuracy through a model learned from cell line* data.

* cell line: A group of cells derived from a uniform tissue, meaning a lineage of cells with the same genetic characteristics. It is used to study resistance to drugs, etc.

Even if the same drug is used on a patient with the same type of cancer, the response to the drug may vary depending on the individual's genetic characteristics or mutations in the cancer cells.

Because accurate prediction of drug reactions is important in order to find the right drug for each individual, research to predict drug reactions using artificial intelligence techniques such as machine learning or deep learning has been actively conducted recently.

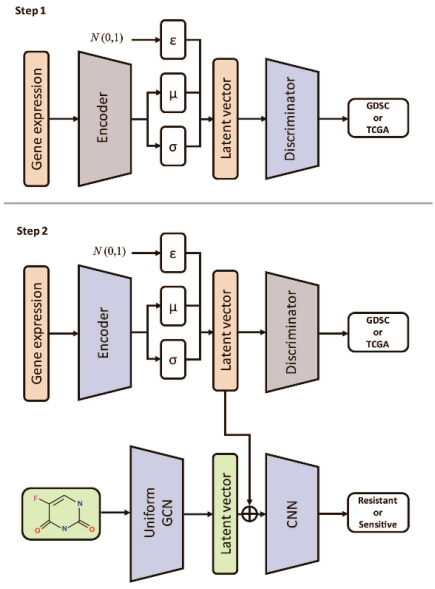
In most drug response prediction studies, the number of patient data containing drug response information is insufficient, so models are trained using cell line data with sufficiently large data. However, cell line data is significantly different from gene expression level information in patient data in that the immune system, vascular system, etc. do not exist.

Therefore, there is a limitation in that accuracy is lowered when a model learned from cell line data is applied to patient data.

The research team utilized generative adversarial neural network (GAN)* to reduce the difference in representation between cell line data and patient data in an artificial intelligence model, thereby creating a model that can accurately predict drug responses even in patient data even when learning from cell line data and developed 'PANCDR (Precision medicine prediction using an Adversarial Network for Cancer Drug Response)'.

* generative adversarial network (GAN): An algorithm that creates new data by imitating existing data, and has a structure in which two models compete against each other to achieve a goal.

The 'PANCDR' model developed by the research team alternately learns a discriminator and a drug response prediction model. In the first step, a discriminator is learned to distinguish whether a latent vector encoded by a Gaussian encoder comes from the gene expression data of a cell line or a patient.



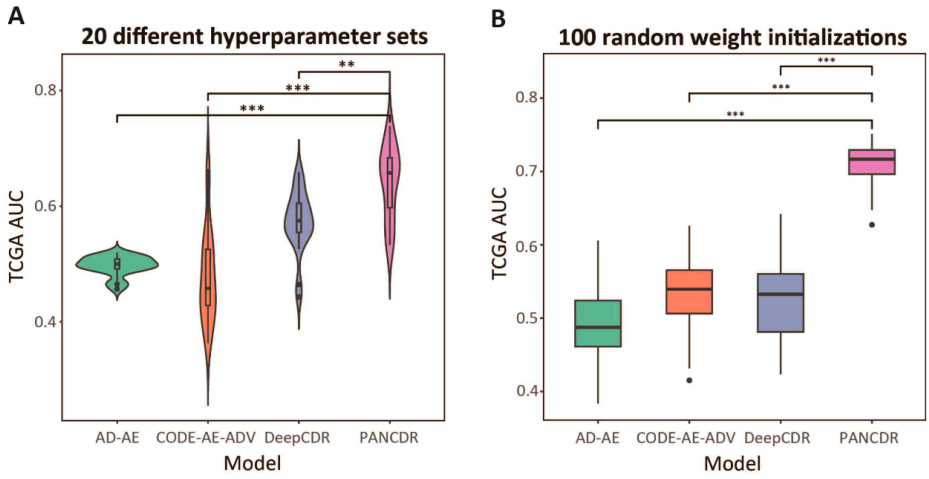
▲ PANCDR model structure. It shows learning in two stages. In the first stage, the discriminator learns to distinguish which data the latent vector comes from. In the second step, the discriminator is fooled and a drug response prediction model is trained.

In the second step, a drug response prediction model is trained so that the discriminator cannot distinguish which data it came from. At this time, large-

scale patient data with only gene expression data and no drug responsiveness was used.

The ‘PANCDR’ model (AUC* 0.7106) showed more than 34% better prediction performance than the existing drug response prediction model (AUC 0.5273) in patient data.

* AUC (Area Under the ROC Curve): This refers to the area under the ROC (Receiver Operating Characteristic) curve and indicates the performance of the classification model.



▲ Performance comparison of existing deep learning models and PANCDR. (A) When using various combinations of hyperparameters, PANCDR shows higher performance than other models. (B) When trained by randomly changing the initial weight 100 times, PANCDR shows higher and more stable performance than other models.

The research team applied the ‘PANCDR’ model to the data of pediatric brain tumor patients from the Seoul National University Hospital research team (Professor Sung-Hye Park) and selected the top five drugs predicted to have the highest reactivity. And as a result of examining existing research on this topic, it was confirmed that all five drugs were related to brain tumors, verifying the accuracy and reliability of the ‘PANCDR’ model.

Professor Hyunju Lee said, "Through this research outcome, it is possible to make predictions with high accuracy from patient data even when learning a drug response model using cell line data. This is expected to provide accurate drug response predictions for personalized treatment in the future."

This study, led by Professor Hyunju Lee of the GIST AI Graduate School and conducted by Researcher Juyeon Kim, was a joint research with Professor Sung-Hye Park of the Department of Pathology, Seoul National University College of Medicine, and was supported by the Institute for Information Technology Planning and Evaluation (IITP), and was published in the international journal 'Briefings in Bioinformatics' in the field of bioinformatics. ' was published on March 14, 2024.

GIST
Since 1993