Data collection and preprocessing

1 Drug combination screening datasets

In this paper, we utilized the NCI ALMANAC and O'Neil drug screening datasets to train the deep learning models. The NCI ALMANAC dataset consists of combo scores for various permutations of 104 FDA-approved drugs, representing their impact on tumor growth in NCI60 human tumor cell lines. For evaluating the synergy score of two drugs on a specific tumor cell line, we utilized the average combo-score of two drugs at different doses, employing a 4-element tuple: $<D_A, D_B, C_C, S_{ABC}>$. On the other hand, the O'Neil dataset, obtained from the DrugComb platform, served as another drug combination screening dataset. We extracted the processed dataset from this platform, which provided average synergy Loewe scores. These scores assessed the synergy score of two drugs on a given tumor cell line, employing a 4-element tuple: $<D_A, D_B, C_C, S_{ABC}>$.

2 Multi-omic data of cancer cell lines

In this study, multi-omic data comprising RNA-seq, copy number variation, gene methylation, and gene mutation data for a total of 1,489 genes were incorporated into the model. These data were obtained from the Cell Model Passports and CCLE databases. Through the identification of overlapping cell lines from these databases, we identified 42 cell lines that were present in both the Cell Model Passports database and the NCI ALMANAC dataset, which are A498, A549/ATCC, ACHN, BT-549, CAKI-1, DU-145, EKVX, HCT-116, HCT-15, HOP-62, HOP-92, HS 578T, IGROV1, K-562, KM12, LOX IMVI, MCF7, MDA-MB-231/ATCC, MDA-MB-468, NCI-H23, NCI-H460, NCI-H522, OVCAR-3, OVCAR-4, OVCAR-8, PC-3, RPMI-8226, SF-268, SF-295, SF-539, SK-MEL-2, SK-MEL-28, SK-MEL-5, SK-OV-3, SNB-75, SR, SW-620, T-47D, U251, UACC-257, UACC-62, UO-31. Additionally, we found 24 cell lines that were present in both the Cell Model Passports database and the O'Neil dataset, which are A2058, A2780, A375, CAOV3, HCT116, HT144, LOVO, MDAMB436, NCI-H460, NCIH1650, NCIH2122, NCIH23, OV90, OVCAR3, RKO,

RPMI7951, SK-OV-3, SKMEL30, SKMES1, SW-620, SW837, T-47D, UACC62, VCAP.

3 Multi-omic data and clinical phenotypes datasets from ROSMAP

Following the acquisition of the ROSMAP AD datasets, they were reformatted into 2-dimensional data frames, with columns dedicated to sample identifiers, such as IDs and names, and rows corresponding to probes, gene symbols, and gene IDs. To successfully integrate the multi-omic data with clinical information, it was essential to match identical samples across the various datasets. This required standardizing the row data—including probes, gene symbols, and gene IDs—into a unified gene-level format, either by aggregating gene-specific measurements or by resolving duplicates from gene synonyms. Genes were subsequently mapped to a reference genome to ensure precise annotation within the multi-omic datasets. Gene counts were then normalized across the datasets, with missing values imputed using zeros or negative ones as needed. After aligning all columns to standard sample IDs and rows to standardized gene IDs, and ensuring a consistent number of samples and genes, the data was prepared for integration into Graph Neural Network (GNN) models, where epigenomic, genomic, transcriptomic and proteomic data were employed as features for nodes.

4 KEGG Signaling Pathways

About 59,241 gene-gene interactions for over 8,000 genes across different signaling pathways were collected from KEGG database. And 48 signaling pathways in the KEGG dataset were selected as the ground truth of subgraphs in the gene-gene interactions, which were AGE-RAGE signaling pathway in diabetic complications, AMPK signaling pathway, Adipocytokine signaling pathway, Apelin signaling pathway, B cell receptor signaling pathway, C-type lectin receptor signaling pathway, Calcium signaling pathway, Chemokine signaling pathway, ErbB signaling pathway, Estrogen signaling pathway, Fc epsilon RI signaling pathway, FoxO signaling pathway,

Glucagon signaling pathway, GnRH signaling pathway, HIF-1 signaling pathway, Hedgehog signaling pathway, Hippo signaling pathway, Hippo signaling pathway - multiple species, IL-17 signaling pathway, Insulin signaling pathway, JAK-STAT signaling pathway, MAPK signaling pathway, NF-kappa B signaling pathway, NOD-like receptor signaling pathway, Neurotrophin signaling pathway, Notch signaling pathway, Oxytocin signaling pathway, PI3K-Akt signaling pathway, PPAR signaling pathway, Phospholipase D signaling pathway, Prolactin signaling pathway, RIG-I-like receptor signaling pathway, Rap1 signaling pathway, Ras signaling pathway, Relaxin signaling pathway, Signaling pathways regulating pluripotency of stem cells, Sphingolipid signaling pathway, T cell receptor signaling pathway, TGF-beta signaling pathway, TNF signaling pathway, Thyroid hormone signaling pathway, Toll-like receptor signaling pathway, VEGF signaling pathway, Wnt signaling pathway, cAMP signaling pathway, cGMP-PKG signaling pathway, mTOR signaling pathway, p53 signaling pathway. After preprocessing the dataset, 17,259 gene-gene interactions for 1,489 genes were obtained. And 28,843 gene-gene interactions for 2,099 genes were obtained for ROSMAP AD dataset.

5 Drug-Target interactions derived from DrugBank

Drug-target information was extracted from the DrugBank database (version 5.1.5, released 2020-01-03). In total, 15,263 drug-target interactions were obtained for 5,435 drugs/investigational agents and 2,775 targets. Further, 17 drugs with known targets to the 1,489 genes previously identified were selected for use in our model for NCI ALMANAC, specifically: Celecoxib, Cladribine, Dasatinib, Docetaxel, Everolimus, Fulvestrant, Gefitinib, Lenalidomide, Megestrol acetate, Mitotane, Nilotinib, Paclitaxel, Romidepsin, Sirolimus, Thalidomide, Tretinoin, Vorinostat. And 10 drugs with known targets to the 1,489 genes previously identified were selected for use in our model, specifically: Dasatinib, Erlotinib, Lapatinib, Sorafenib, Sunitinib, Vorinostat, Geldanamycin, Metformin, Paclitaxel, Vinblastine.