AI-Powered Prediction of Synergistic Drug Combinations for Cancer Treatment

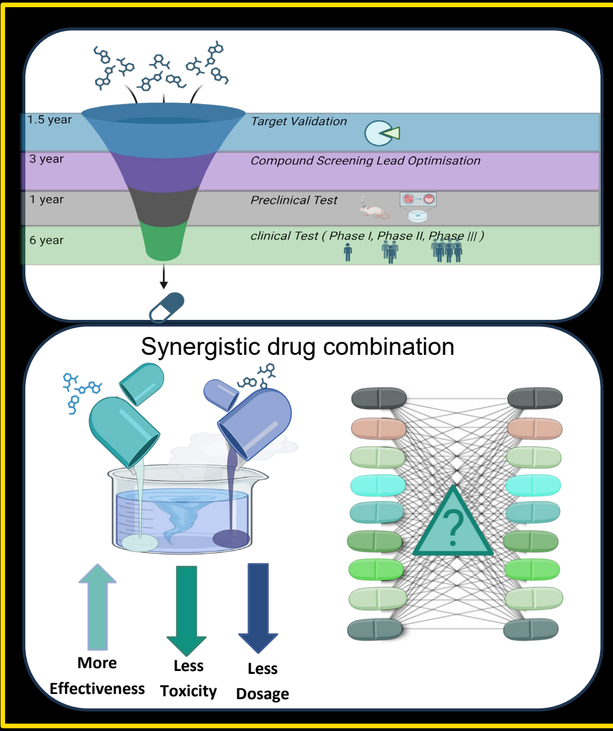
Parvin Mansouri[1], Muhammad Javad Heydari[1], Bryan Lye[1], Thomas Marsland[1], Antoine De Weck[2], Fatemeh Vafaee[1]

1. School of Biotechnology & Biomolecular Sciences, UNSW
2. School of Clinical Medicine, UNSW

UNSW Biomedical AI VAFAEE LAB

Abstract:

Combinational drug therapies hold great promise for treating complex diseases like cancer, but identifying effective drug combinations remains a significant challenge due to the vast number of possible pairings. Although in vitro testing on cancer cell lines has provided valuable insights into the effects of drugs, exploring the entire combinatorial space is impractical, limiting the discovery of potentially life-saving treatments. Artificial intelligence (AI) offers a transformative solution to this challenge. While most studies focus solely on the chemical structures of drugs, we recognize that a drug's mechanism of action is highly dependent on its targets. To address this, we developed a novel approach using graph neural networks (GNNs) to model drug-target interactions by integrating drug structural information (i.e., molecular graph representations) with protein target features extracted from a transformer model pre-trained on proteins' primary structures. Additionally, we incorporated a contrastive learning technique that leverages drug bioactivity data, derived from the descriptions of bioassays processed using large language models (LLMs).By combining these features with gene expression data from cancer cell lines, we trained advanced machine-learning models capable of predicting synergistic drug combinations with greater accuracy and biological relevance. Our framework represents a significant step toward utilizing AI to accelerate the development of combinational therapies for cancer treatment.



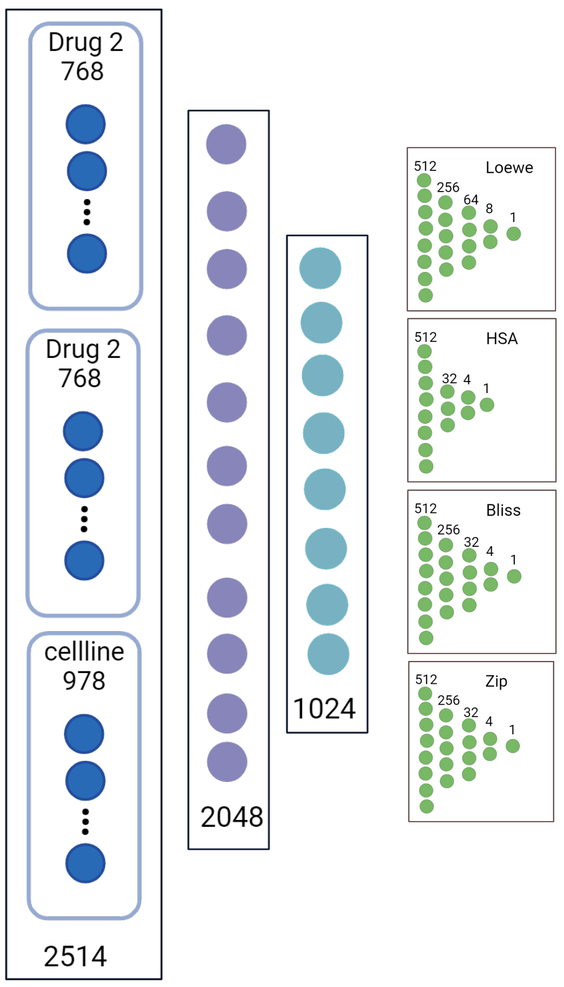
Introduction:

Traditional de novo drug discovery involves a lengthy and costly process, taking over 12 years and approximately $2 billion USD to bring a single drug to market, with a low success rate of only 5%. One promising alternative to accelerate therapeutic development is the use of drug combinations. Drug combinations, defined by the FDA as therapies involving two or more approved drugs, have shown efficacy in treating diseases such as cancer and autoimmune disorders. For instance, the combination of irinotecan and 5-fluorouracil has demonstrated a synergistic effect in colorectal cancer treatment, leveraging established pharmacological and toxicological profiles to enhance efficacy, reduce toxicity, and minimize resistance.

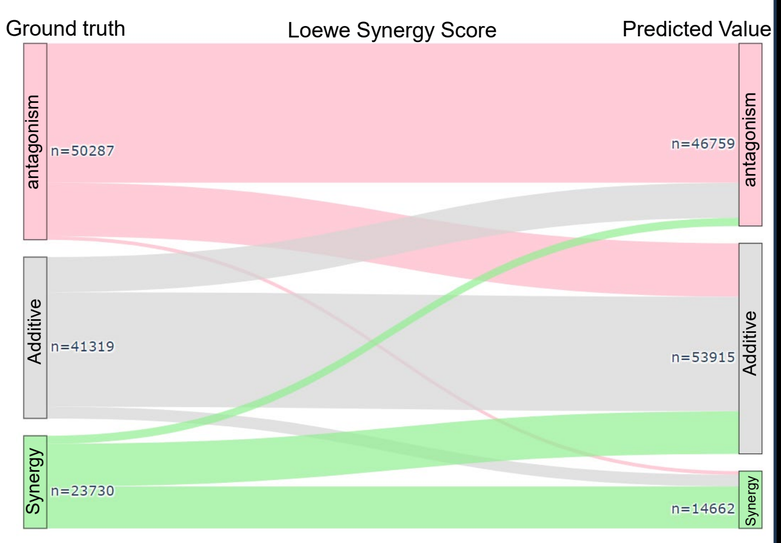
Despite these advantages, the traditional method for identifying synergistic combinations relies heavily on empirical testing, which is time-consuming, labor-intensive, and costly. The growing number of approved drugs further amplifies the combinatorial complexity, making it impractical to evaluate all possible combinations across disease contexts.

To address these challenges, this research aims to develop an AI-based model to predict and identify optimal drug combinations efficiently. By leveraging advanced computational methods, this approach seeks to streamline the discovery process, reduce resource expenditure, and maximize therapeutic benefits for patients awaiting treatment.

Model



RESULT



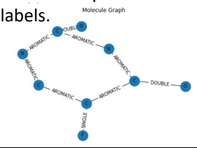
Feature representation

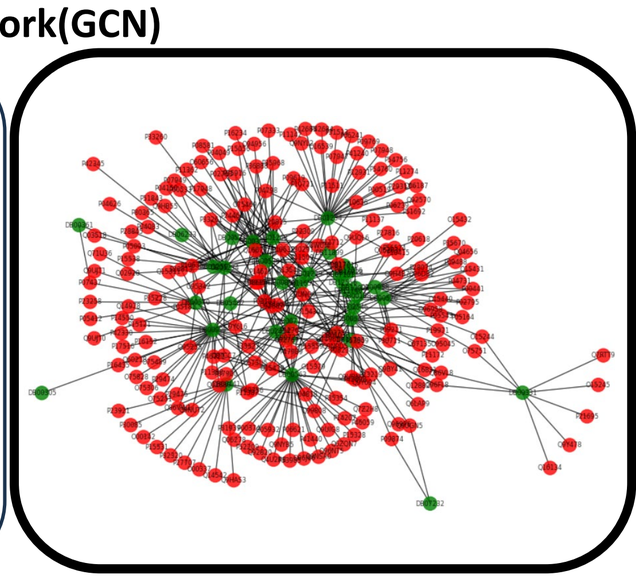
Drug Feature Representation

Protein Target Network(GCN)

Drug: Graph embedding

Molecules can be described as heterogeneous graph structures, where atoms are represented as nodes and chemical bonds as edges. Node and edge features capture the relationships between neighboring atoms in a graph, facilitating molecular representation learning. Graph neural networks (GNNs) operate on molecular graphs to iteratively up-date node representations through neighborhood aggregation, ultimately producing com-prehensive graph embeddings. These graph-level representations can then be used by a simple classifier to predict molecular properties or labels.

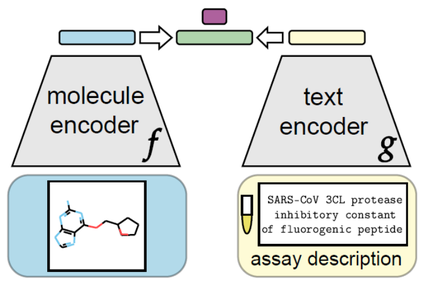




Target: ESM

ESM (Evolutionary Scale Modeling) is a cutting-edge protein language model that generates high-resolution protein embeddings directly from a protein's primary amino acid sequence. Trained using a masked language modeling approach, ESM predicts randomly masked amino acids based on their surrounding context, enabling the model to learn complex dependencies and evolutionary sequence patterns. This training enhances the model's understanding of protein structures. These embeddings are versatile and can be utilized for various downstream tasks, including accurate three-dimensional structure prediction, functional annotation, and assessing the effects of mutations, as demonstrated by models like ESMFold that leverage these embeddings to predict atomic-level protein structures.

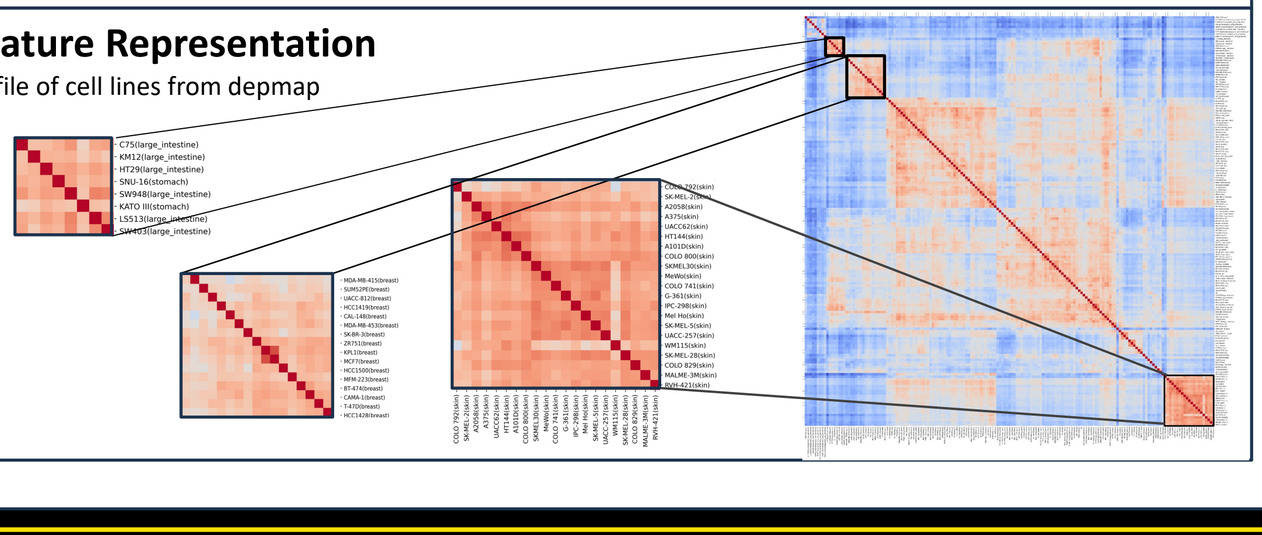
CLAMP Contrastive Language-Assay-Molecule Pre-training



The CLAMP model architecture consists of a trainable molecule encoder that produces molecule embeddings and a trainable text encoder that generates bioassay embeddings. Both embeddings are assumed to be layer-normalized. CLAMP includes a scoring function, designed to return high values when a molecule m is active on a bioassay a and low values otherwise. The contrastive learning approach allows CLAMP to perform zero-shot transfer learning, enabling it to make meaningful predictions for bioassays not seen during training.

Cell line Feature Representation

Gene Expression profile of cell lines from depmap



	Epoch	LOEWE	ZIP	HSA	BLISS
MSE	1000	94.01	41.93	50.21	95.08
MAE	1000	6.83	4.35	4.91	6.03
MSE	2000	167.44	78.03	96.57	266.76
MAE	2000	9.02	5.9	6.54	10.27
MSE	3000	76.91	37.06	44.63	89.2
MAE	3000	6.14	4.09	4.65	5.7
MSE	4000	199	175	151	683.5
MAE	4000	9	8	8.04	16.49

We implemented two approaches for drug feature representation: CLAMP and GCN, leveraging message passing and aggregation within a drug-protein interaction graph. For cell line features, we utilized gene expression profiles from DepMap. By employing an early fusion technique, we designed a model architecture tailored to optimize performance across various synergy metrics.

Our methodology involved concatenating the features of both drugs and the cell line at the input layer, passing them through two shared hidden layers. The model then branched into four distinct output paths, each tailored to predict a specific synergy score through unique sequences of fully connected layers. Finally, the outputs were integrated into a single tensor for evaluation.

To ensure metric-agnostic consistency, we employed simultaneous optimization across multiple synergy metrics, achieving robust results. Notably, our approach delivered an MSE of 6.4 for the Loewe synergy score, showcasing the efficacy of our design in predicting drug synergy.

References:

- P. Seidl, A. Vall, S. Hochreiter, and G. Klambauer, "Enhancing activity predictionmodels in drug discovery with the ability to understand human language," in Inter-national Conference on Machine Learning, pp. 30458–30490, PMLR, 2023.
- L. H. Torres, B. Ribeiro, and J. P. Arrais, "Few-shot learning with transformers viagraph embeddings for molecular property prediction," Expert Systems with Applica-tions, vol. 225, p. 120005, 2023.
- K. Preuer, R. P. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, and G. Klambauer,"Deepsynergy: predicting anti-cancer drug synergy with deep learning," Bioinformat-ics, vol. 34, no. 9, pp. 1538–1546, 2018.
- V. Parvathaneni, N. S. Kulkarni, A. Muth, and V. Gupta, "Drug repurposing: apromising tool to accelerate the drug discovery process," Drug discovery today, vol. 24,no. 10, pp. 2076–2085, 2019.
- Z. Lin, H. M. Akin, R. Rao, T. Hie, J. Lu, D. Smetanin, L. Zheng, and A. Chakraborty, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, no. 6637, pp. 1123–1130, 2023.

ALGORAE PHARMACEUTICALS · CSIRO · DATA 61 · Med-Tech.AI Next Generation Graduates Program