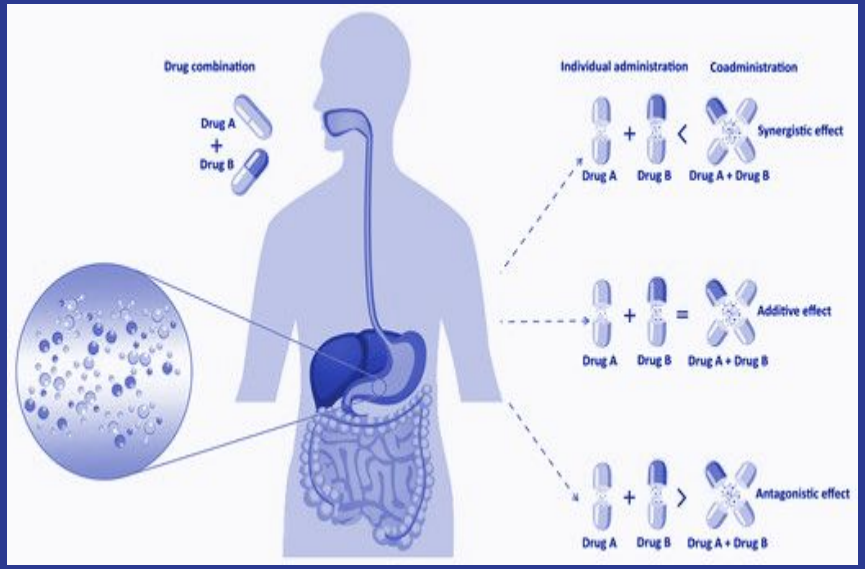
Cancer Drug Synergy Prediction

Machine Learning in Bioinformatics (INFO-I 529)
Spring 22



Group 2:
Dhanusha Duraiyan
Supriya Bidanta
Farzaneh Behzadnia
Nitanshu Joshi
Paveethran Swaminathan

Overview

Drug A → Specific response for A on cell line

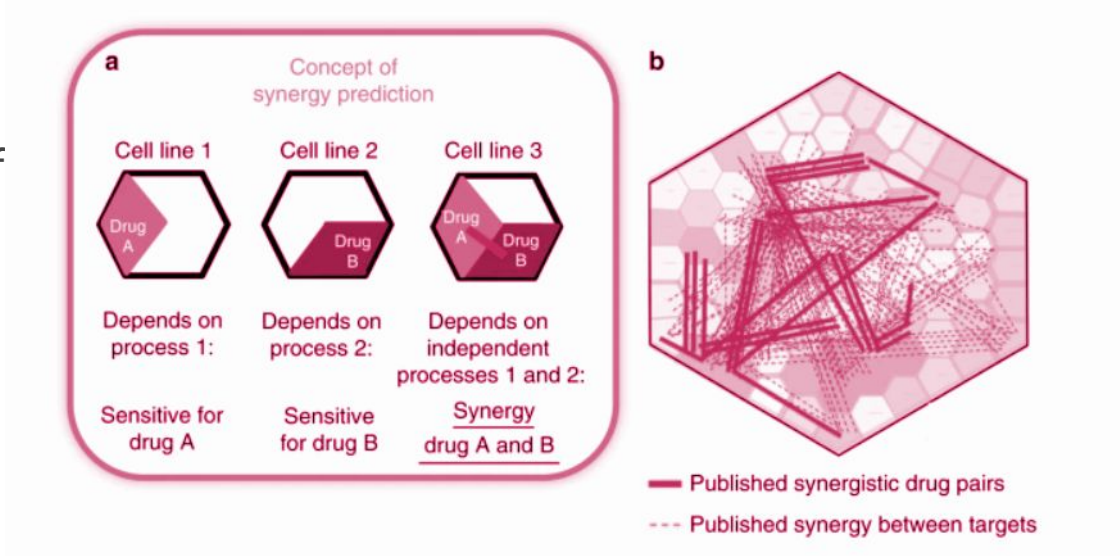
Drug B → Specific response for B on cell line

Synergy of A and B →

Combined response of

A and B on cell line.

Synergy could be additive or inhibitive.



Why synergy scores?

Synergy score - Average excess response due to drug interactions.

- Helps to predict which new (untested) drug combinations are likely to yield **synergistic behaviors** in a patient population
- Helps to **identify novel biomarkers** that may help reveal underlying mechanisms related to drug synergy
- Helps to predict whether a known **drug combination will be effective** for a specific patient.

Review Of Literature:

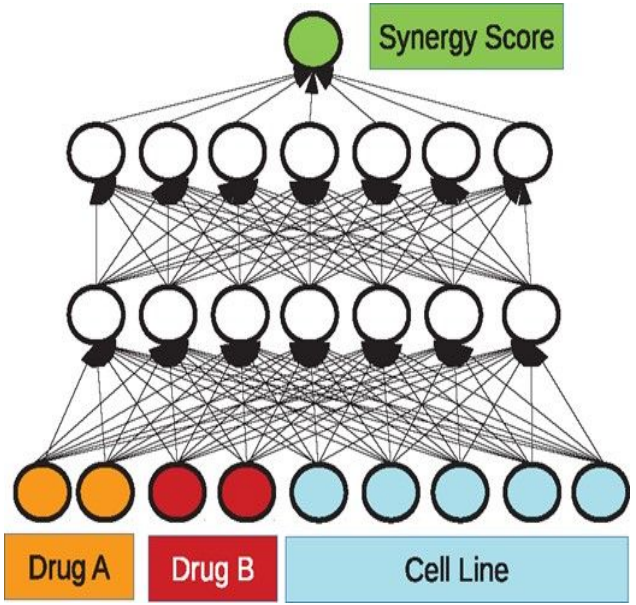
Finding synergistic drug pairs for a particular cancer type is important for improving efficacy of anticancer treatment.

Each sample:

two compounds + a cell line

General approach:

- Quantification of drug response
- Molecular feature selection or dimensionality reduction of the cellular measurements
- Machine learning model approach to predict drug synergy (ComboScores)
- Model evaluation



Datasets used in current methods:

Dataset Name	Type	# Combinations	# Drugs	# Patients/cell Lines	URL	Ref
Drug Combination Database	Clinical	1363	904	~140,000	http://www.cls.zju.edu.cn/dcdb/	125
Merck	In vitro	583	38	39	http://mct.aacrjournals.org/highwire/filestream/53222/field_highwire_adjunct_files/3/156849_1_supp_1_w2lrww.xls	126
AstraZeneca-Sanger Drug Combination Dataset	In vitro	910	118	85	https://www.synapse.org/#!Synapse:syn4231880/wiki/235645	30
NCI ALMANAC	In vitro	5,000+	105	60	https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC	78

Common features used:

1. Chemical descriptors of Drug A, Drug B
2. Drug concentration
3. Cancer cell lines
4. Morgan FingerPrint Counts (MFPC)
5. MACCS keys
6. ISIDA fragments
7. SIRMS fragments

Methods to predict synergy-scores

- Naive method: based on clinical experiences - trial and error method

- High-throughput Screening (HTS)

- Machine learning approach

Collective Machine Learning models used:

Method comparison based on their ability to predict synergy values of novel drug combination (with LODO cross-validation):

- Median Polish
- Elastic nets
- Support Vector Machines (SVMs)
- Random Forests
- Gradient Boosting Machines
- Ensemble method

Method	MSE
Deep Neural Networks	255.49
Gradient Boosting Machines	275.39
Random Forests	307.56
Support Vector Machines	398.39
Elastic Nets	420.24
Baseline (Median Polish)	477.77

Deep learning approaches:

- DeepSynergy
- AuDNNsynergy (Auto-encoder + Deep Neural Network)

Metrics used to evaluate all models:

- RMSE, MSE
- Pearson’s Correlation Coefficient, Spearman’s correlation coeff.
- Accuracy, AUC-ROC, F1-Score, Kappa score (Binary values)

Limitations

- Since the dataset has only a limited number of different drugs and cell lines all methods show difficulties to generalize well across novel drugs and cell lines.
- Even if training set size, features and classifier are high, the modeled relationship between drug synergy and features depend on training set composition and cell line properties (implicitly).
- Evaluating performance of ml models on individual cancer-cell types.
- Interpretability of models

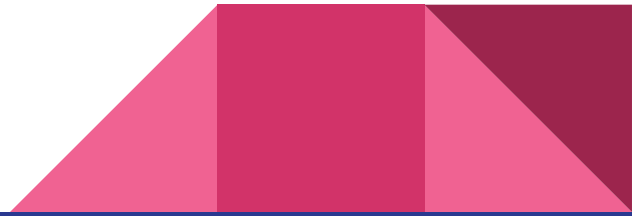
Proposal:

1. Optimal feature-model combination for every cancer-type for more reliable synergy prediction
2. Convolution Neural Network based approach to predict drug synergy scores (either numerical matrix input or converting matrix values to image representation to assign spatial relationship between features)

   Evaluate with LODO CV, RMSE + Pearson's corr score

In addition to the commonly used features, we plan to use these physico-chemical features (RDK Kit):

1. Total polar surface area (TPSA)
2. Molecular weight
3. Number of aliphatic and aromatic rings



References:

- https://doi.org/10.1093/bib/bbac075
- https://doi.org/10.1158/1535-7163.MCT-15-0843
- https://doi.org/10.3389/fchem.2019.00509
- https://doi.org/10.1093/bioinformatics/btx806
- https://www.nature.com/articles/s41598-021-90923-y

Q&A