OXFORD

Modeling drug combination effects via latent tensor reconstruction

Tianduanyi Wang[1,2,*], Sandor Szedmak[1], Haishan Wang[1], Tero Aittokallio[1,2,3,4,5], Tapio Pahikkala[6], Anna Cichonska[1,2] and Juho Rousu[1,*]

[1]Department of Computer Science, Helsinki Institute for Information Technology HIIT, Aalto University, Espoo, Finland, [2]Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Helsinki, Finland, [3]Department of Mathematics and Statistics, University of Turku, Turku, Finland, [4]Institute for Cancer Research, Oslo University Hospital, Oslo, Norway, [5]Oslo Centre for Biostatistics and Epidemiology (OCBE), University of Oslo, Oslo, Norway and [6]Department of Computing, University of Turku, Turku, Finland

*To whom correspondence should be addressed.

Abstract

Motivation: Combination therapies have emerged as a powerful treatment modality to overcome drug resistance and improve treatment efficacy. However, the number of possible drug combinations increases very rapidly with the number of individual drugs in consideration, which makes the comprehensive experimental screening infeasible in practice. Machine-learning models offer time- and cost-efficient means to aid this process by prioritizing the most effective drug combinations for further pre-clinical and clinical validation. However, the complexity of the underlying interaction patterns across multiple drug doses and in different cellular contexts poses challenges to the predictive modeling of drug combination effects.

Results: We introduce *comboLTR*, highly time-efficient method for learning complex, non-linear target functions for describing the responses of therapeutic agent combinations in various doses and cancer cell-contexts. The method is based on a polynomial regression via powerful latent tensor reconstruction. It uses a combination of recommender system-style features indexing the data tensor of response values in different contexts, and chemical and multi-omics features as inputs. We demonstrate that *comboLTR* outperforms state-of-the-art methods in terms of predictive performance and running time, and produces highly accurate results even in the challenging and practical inference scenario where full dose–response matrices are predicted for completely new drug combinations with no available combination and monotherapy response measurements in any training cell line.

Availability and implementation: *comboLTR* code is available at https://github.com/aalto-ics-kepaco/ComboLTR.

Contact: tianduanyi.wang@aalto.fi or juho.rousu@aalto.fi

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Combination therapies, involving two or more drugs, offer several advantages over standard monotherapies, including higher treatment efficacies and overcoming resistance mechanisms by modulating multiple targets and signaling pathways. This is especially important in combating complex multi-factorial diseases, such as cancer, and cardiovascular, neurological and autoimmune disorders. Moreover, drugs in combination can often be administered in lower individual doses which, in turn, results in reduced risk of adverse reactions (Al-Lazikani *et al.*, 2012; Pemovska *et al.*, 2018). The number of US Food and Drug Administration-approved drug combinations has been continuously growing since the first approvals for co-administration of drugs to treat nervous and respiratory system disorders in 1940s (Das *et al.*, 2018). Currently, most of the ongoing research and development is focused on combinatorial therapies for different cancer types (Pemovska *et al.*, 2018). The development is, however, very challenging as the number of possible pairwise combinations increases very rapidly with the number of individual drugs,

not even mentioning the enormous size of the chemical universe that could be explored (Reymond and Awale, 2012).

Computational approaches offer cost-effective means for large-scale, fast and systematic pre-screening and prioritization of potential drug combinations for further experimental validation. Most of the machine-learning models introduced to date and benchmarked in crowd-sourced DREAM Challenge competitions (Bansal *et al.*, 2014; Menden *et al.*, 2019) focus directly on the prediction of synergistic drug combinations (Li *et al.*, 2019; Preuer *et al.*, 2018; Sidorov *et al.*, 2019; Tonekaboni *et al.*, 2018; Yang *et al.*, 2020). Nonetheless, modeling the full dose–response matrices of drug pairs offers more in-depth view of their complex response landscapes, and allows to calculate different synergy metrics as a follow-up step. This is important especially for translational applications, where knowledge of optimal dose combination regions is often critical (Ianevski *et al.*, 2020; Tang *et al.*, 2015).

Here, we introduce *comboLTR*, a new polynomial regression-based framework for modeling anti-cancer effects of drug combinations in various doses. We compare the performance of *comboLTR*

Table 1. The general scheme of LTR-based regression

Prediction	Learning
Given : $\mathbf{T}$ tensor, $\mathbf{x}$	Given : $\{(y_i, \mathbf{x}_i) \mid i \in [m]\}$
Output : y	Output : $\mathbf{T}$ tensor
$\pi(\mathbf{x}) = \langle \mathbf{T}, \otimes^{n_d} \mathbf{x} \rangle \Rightarrow y$	$\min_{\lambda, \mathbf{P}} \sum_i \|y_i - \langle \mathbf{T}, \otimes^{n_d} \mathbf{x}_i \rangle\|^2$
	$s.t. \, \mathbf{T} = \sum_{t=1}^{n_t} \lambda_t \otimes_{d=1}^{n_d} \mathbf{p}_t^{(d)}$

Note: Given an n_d-order parameter tensor $\mathbf{T}$ and data point $\mathbf{x}$, the prediction entails computing an inner product between $\mathbf{T}$ and an n_d-order tensor product of the data point with itself. Learning entails finding a factorization of $\mathbf{T} = \sum_{t=1}^{n_t} \lambda_t \otimes_{d=1}^{n_d} \mathbf{p}_t^{(d)}$ with the lowest regression error.

to random forest (RF) and recently introduced method by our groups, the *comboFM* method (Julkunen *et al.*, 2020), using the National Cancer Institute (NCI)-ALMANAC dataset (Holbeck *et al.*, 2017) generated by the US NCI. Both *comboLTR* and *comboFM* exploit the fact that dose–response matrices of drug combinations can be represented as a higher-order tensor indexed by drugs, drug concentrations and cell lines. To predict the response values within the data tensor, a highly non-linear polynomial model is needed in order to capture the multi-way interactions. To learn the parameters of multivariate high-order polynomials, tensor factorization approaches are effective. *comboFM* models the drug combination effects by learning latent factors of the tensor using factorization machines that estimate non-linear target functions using symmetric polynomials and factorized parametrization (Blondel *et al.*, 2016). On the other hand, *comboLTR* is based on latent tensor reconstruction (LTR) method (Szedmak *et al.*, 2020), which can be also considered as an alternative of factorization machines, which extends the range of functions that can be learned by removal of the assumption of the symmetry imposed on the polynomials. Moreover, due to only linear dependence on the design parameters separately, a straight, gradient-based algorithm can be applied, which can also exploit advanced update rules, e.g. ADAM (Kingma and Ba, 2015). As a consequence, *comboLTR* can process much larger datasets than *comboFM*, in the number of both examples and features, with significantly reduced running time.

In summary, this article makes the following contributions.

- We introduce *comboLTR*, a new time-efficient framework for modeling drug combination responses in cancer cell lines based on a polynomial regression model where the function learning problem is transformed into a tensor reconstruction problem, with the tensor indexed by drugs, drug concentrations and cell lines. The algorithm implements mini-batch data processing and allows learning complex, highly non-linear target functions from large-scale datasets with a constant memory complexity and linear running time in all important parameters (degree, rank, sample size and number of variables).
- We demonstrate that *comboLTR* provides highly accurate cell line context-specific results under various prediction scenarios, including more challenging and practical settings where dose–response matrix predictions are made for (i) new drug combinations with no available combination response measurements in any cell line and (ii) when response measurements of individual drugs are also lacking from the training data. Moreover, we show that drug combination synergy scores can be recovered with high accuracy based on the predicted dose–response matrices.

- *comboLTR* can work with large feature sets, including chemical descriptors and multi-omics cell line features, such as gene expression, copy number variation (CNV), CRISPR-Cas9 gene knock-outs and proteomics data.

2 Materials and methods

2.1 Notation

In the text, $\otimes$ denotes the tensor product of vectors, $\langle , \rangle$ is the inner product and $\| \, \|$ is the norm in a Hilbert space $\mathcal{H}$. The notation $\langle , \rangle$ is also applied for the Frobenius inner product of tensors, $\circ$ denotes the pointwise product of tensors with the same shape of any order. 1_m is a vector of dimension m whose all components equal to 1. The set $1, \ldots, n$ for a given n is denoted by $[n]$. The matrix $\mathbf{D}_{\mathbf{u}}$ is a diagonal matrix whose diagonal is equal to the vector $\mathbf{u}$. $\mathbf{A}_i$ denotes the row i of matrix $\mathbf{A}$.

2.2 Data representation

In the learning problem, we have a sample of examples given by input–output pairs $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i) \mid i \in [m], \mathbf{x}_i \in \mathbb{R}^n, \mathbf{y}_i \in \mathbb{R}^{n_y}\}$ taken from an unknown joint distribution of input and output sources. The rows of the matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ contain the vectors $\mathbf{x}_i$, and similarly the rows of $\mathbf{Y}$ hold the output examples, $\mathbf{y}_i$, for all i.

2.3 Background: learning polynomial regression models

In this article, we consider learning polynomial regression models

$$
\begin{aligned}
\pi(\mathbf{x}) \quad &= \sum_{j=1}^{n} w_j x_j + \sum_{j,k=1}^{n} w_{jk} x_j x_k + \cdots \\
&+ \sum_{j_1, j_2, \ldots, j_{n_d}=1}^{n} w_{j_1, \ldots, j_{n_d}} x_{j_1} \cdots x_{j_{n_d}},
\end{aligned}
\tag{1}
$$

where w's are the regression coefficients to be learned, n is the number of input variables and n_d is the degree of the polynomial.

Polynomial regression models are known to have high representation power, capable of accurately representing continuous functions with a fixed L_∞ norm-based tolerance. This fact allows us to exploit the Stone–Weierstrass theorem and its generalizations (Prenter, 1970) to approximate those functions by polynomials on a compact subset with an accuracy not worse than a given arbitrary small error.

However, estimating high-degree multivariate polynomial functions presents challenges. An arbitrary multivariate polynomial defined on the field of real numbers can be described by $\binom{n + n_d}{n}$ parameters, where n is the number of variables, and n_d is the maximum degree of the polynomial. Thus, the complexity relating to the size of the underlying parameter tensor is $O(n^{n_d})$, which grows exponentially in the number of parameters.

This exponential complexity in the polynomial degree presents both statistical and computational challenges: there is often not enough data to reliably estimate all the coefficients, and the exponential time and space complexity forbids processing sufficiently large training sets.

The key approach to tackle this exponential complexity in higher-order factorization machines (HOFM) (Blondel *et al.*, 2016; Rendle, 2010) is a special representation of the coefficients as inner products of factors: e.g. for the second order terms

$$
w_{i_1, i_2} = \langle \mathbf{p}_{i_1}, \mathbf{p}_{i_2} \rangle = \sum_{j=1}^{n_t} p_{i_1, j} p_{i_2, j},
$$

where $\mathbf{p}_i \in \mathbb{R}^{n_t}$ encodes the participation of i'th variable in n_t factors. For higher degree terms, the same is given by a generalized inner product

$$w_{i_1,\dots,i_m} = \langle \mathbf{p}_{i_1}, \dots, \mathbf{p}_{i_m} \rangle = \sum_{j=1}^{n_t} p_{i_1,j} \cdots p_{i_m,j}.$$

This factorized representation drastically reduces the number of parameters to $O(n_d \cdot n_t \cdot n)$ (Blondel *et al.*, 2016). The HOFM was recently demonstrated to be able to accurately predict drug combination responses (Julkunen *et al.*, 2020). However, HOFMs are constrained to symmetric polynomials, i.e. functions that are invariant to permutation of features, which restricts the HOFM model as a general regression model.

In this work, we follow an alternative approach for factorizing the parameter representation, called LTR (Szedmak *et al.*, 2020), that starts from the full-order tensor representation of the unknown regression coefficients and learning a factorization into rank-one tensors (of full order) that optimizes the regression error (Table 1). Importantly, the LTR model lifts the limitation of symmetricity of the learned polynomial, and can therefore tackle a wider class of learning problems.

2.4 Tensor-based representation of polynomial functions

A polynomial function over the real numbers with degree n_d and with n variables can also be written in a compact form

$$\pi(\mathbf{x}) = \langle \mathbf{T}, \mathbf{x} \otimes_{d=1}^{n_d} \mathbf{x} \rangle, \tag{2}$$

where $\mathbf{T}$ is a symmetric tensor of order n_d, and with dimension $\times_{d=1}^{n_d} n$. If the vector $\mathbf{x}$ is given in homogeneous form, extended with a constant 1, then (2) covers all possible polynomials up to degree n_d. The tensor $\mathbf{T}$ can be given in a decomposed HOSVD form (Kolda and Bader, 2009; Lathauwer *et al.*, 2000),

$$\mathbf{T} = \sum_{t=1}^{n_t} \lambda_t \otimes_{d=1}^{n_d} \mathbf{p}_t^{(d)} \tag{3}$$

$$\text{s.t.} \quad ||\mathbf{p}_t^{(d)}|| = 1, \ \mathbf{p}_t^{(d)} \in \mathbb{R}^n, \ t \in [n_t].$$

This representation is generally not unique, see Lathauwer *et al.* (2000) and de Silva and Lim (2008). By replacing $\mathbf{T}$ with its decomposed form, the polynomial function turns into the following expressions

$$\pi(\mathbf{x}) = \sum_{t=1}^{n_t} \lambda_t \langle \otimes_{d=1}^{n_d} \mathbf{p}_t^{(d)}, \otimes_{d=1}^{n_d} \mathbf{x} \rangle = \sum_{t=1}^{n_t} \lambda_t \prod_{d=1}^{n_d} \langle \mathbf{p}_t^{(d)}, \mathbf{x} \rangle,$$

where we exploit the well-known identity connecting the inner products and the tensor products (Golub and Loan, 2013). This form only consists of terms of scalar factors, where each scalar is the value of a linear functional acting on the space $\mathcal{X}$. This transformation eliminates the difficulties, which arise in working directly with full tensors. Observe that the function π is linear in each of the vector-valued parameters, $\mathbf{p}_t^{(d)}, t \in [n_t], d \in [n_d]$.

We can further transform the polynomial representation (2.4) into a form, which does not contain any reference to tensor product. We have a following simple statement, which allows us to introduce an additional factorization within the polynomial function to reduce further the number of parameters.

Proposition 1. *The polynomial function $\pi(\mathbf{x})$ can be expressed only by the help of matrix and pointwise, Hadamard, products, namely*

$$\pi(\mathbf{x}) = \sum_{t=1}^{n_t} \lambda_t \prod_{d=1}^{n_d} \langle \mathbf{p}_t^{(d)}, \mathbf{x} \rangle = \lambda^T \circ_{d=1}^{n_d} \mathbf{P}^{(d)} \mathbf{x}, \tag{4}$$

where $\mathbf{P}^{(d)}$ is a matrix of size $n_t \times n$ for any d, whose rows are given as $\mathbf{P}_t^{(d)} = \mathbf{p}_t^{(d)}$, and λ is a vector with components $\lambda_t, t \in [n_t]$.

Proof. The matrix-vector product $\mathbf{P}^{(d)}\mathbf{x}$ yields a vector with components $(\langle \mathbf{p}_t^{(1)}, \mathbf{x} \rangle, \dots, \langle \mathbf{p}_t^{(n_d)}, \mathbf{x} \rangle)$, and after a rearrangement, the original form can be restored. □

2.5 LTR—basic form

comboLTR is built upon the LTR-based polynomial regression method (Szedmak *et al.*, 2020). LTR exploits the representation shown in Proposition 1, which leads to the following optimization problem:

$$\min \quad \frac{1}{mn_y} \sum_{i=1}^{m} ||\mathbf{y}_i - \left(\circ_{d=1}^{n_d} \mathbf{P}^{(d)} \mathbf{x}_i \right) \mathbf{D}_\lambda \mathbf{Q}||^2$$
$$+ \frac{C_p}{n_t n_d n} ||\mathbf{P}^{(d)}||^2 + \frac{C_q}{n_t n_y} ||\mathbf{Q}||^2 \tag{5}$$
$$\text{w.r.t.} \quad \lambda, \mathbf{Q}, \mathbf{P}^{(d)}, d \in [n_d],$$

where C_p and C_q are penalty constants, and matrix $\mathbf{Q}$ projects the vector given by the polynomial function of dimension n_t into the output space.

2.6 Reparametrization of the polynomial representation

In the LTR model, the predictor is implemented via a polynomial function π acting on vectors $\mathbf{x}$ of dimension n. The parameter space corresponding to matrices $\mathbf{P}^{(d)}, d \in [n_d]$ has dimension $n_d n_t n$ which is large enough to fit the polynomial to a non-linear function with complex structure, but it requires a large sample to achieve a proper estimation of those parameters. The LTR framework can be extended to increase the flexibility, and in the same time, to reduce the dimension of the parameter space. To this end, let the polynomial function of (4) be reformulated

$$\pi(\phi(\mathbf{x})) = \sum_{t=1}^{n_t} \lambda_t \prod_{d=1}^{n_d} \langle \mathbf{v}_t^{(d)}, \mathcal{A}(\mathbf{U}^{(d)T} \mathbf{x}) \rangle$$
$$= \lambda^T \circ_{d=1}^{n_d} \mathbf{V}^{(d)} \mathcal{A}(\mathbf{U}^{(d)T} \mathbf{x}), \tag{6}$$

where $\mathcal{A}$ is a pointwise activation function, and the matrix $\mathbf{U}^{(d)T}$ is a linear transformation projecting the original input vector into a space with lower dimension, n_k, for each $d \in [n_d]$. That projection can enforce a bottleneck within the polynomial function. This modification preserves the linear dependence on the matrix valued parameters. The expression $\phi(\mathbf{x}) = \mathcal{A}(\mathbf{U}^{(d)T}\mathbf{x})$ might be viewed as a layer of a neural network. The main difference is that the layers within the LTR are joined by a polynomial function in a parallel way instead of being connected sequentially.

The following table summarizes the matrices describing the extended LTR problem.

Polynomial	parameters	$\mathbf{V}^{(d)}$ $d \in [n_d]$	$\in \mathbb{R}^{n_t \times n_k}$,
Output	projection	$\mathbf{Q}$	$\in \mathbb{R}^{n_t \times n_y}$,
Scaling		$\mathbf{D}_\lambda = \text{diag}(\boldsymbol{\lambda})$	$\in \mathbb{R}^{n_t \times n_t}$,
In layers	projection	$\mathbf{U}^{(d)}$ $d \in [n_d]$	$\in \mathbb{R}^{n \times n_k}$,
	scaling	$\mathbf{D}_{\lambda^{(u)}} = \text{diag}(\boldsymbol{\lambda}^{(u)})$	$\in \mathbb{R}^{n \times n}$,
	value	$\mathbf{A}^{(d)} = \mathcal{A}\left(\mathbf{X}\mathbf{D}_{\lambda^{(u)}} \mathbf{U}^{(d)}\right)$	$\in \mathbb{R}^{m \times n_k}$,
Full	polynomial	$\mathbf{F} = \circ_{d=1}^{n_d} \mathbf{A}^{(d)} \mathbf{V}^{(d)T}$	$\in \mathbb{R}^{m \times n_t}$,
Error		$\mathbf{E} = \mathbf{Y} - \mathbf{F}\mathbf{D}_\lambda \mathbf{Q}$	$\in \mathbb{R}^{m \times n_y}$.

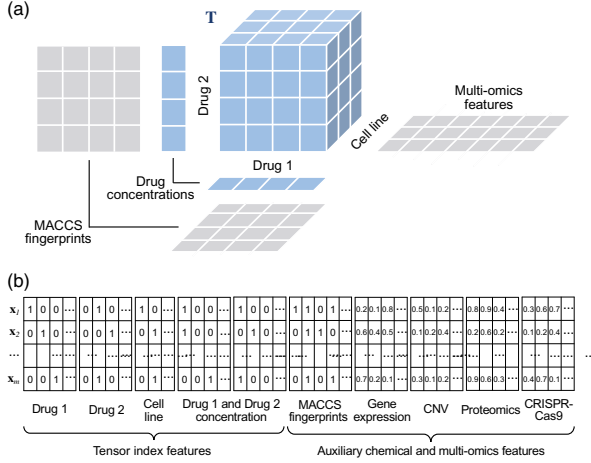The parameter λ corresponds to the singular values of the tensor decomposition, see in (3).

Fig. 1. Illustration of the drug combination response tensor and its feature representation. (**a**) Drug combination responses form a fifth-order tensor indexed by drugs, their concentrations and the cell lines. (**b**) The drug combination response tensor can be flattened into a tensor index featurematrix via one-hot encoding and accompanied by chemical and biological information

The extended LTR problem now takes the following form

$$
\begin{aligned}
\min \quad & \frac{1}{2mn_y}||\mathbf{Y} - \mathbf{FD}_\lambda\mathbf{Q}||_F^2 + \frac{C_\lambda}{2n_t}||\lambda||_2^2 \\
\text{w.r.t.} \quad & \mathbf{U}^{(d)}, \lambda^{(u)}, \mathbf{V}^{(d)}, \lambda, \mathbf{Q}, d = [n_d], \\
\text{s.t.} \quad & ||\mathbf{U}_j^{(d)}|| = 1, ||\lambda^{(u)}|| = 1, \\
& ||\mathbf{V}_t^{(d)}||_2 = 1, ||\mathbf{Q}_t|| = 1, \\
& j \in [n], t \in [n_t], d \in [n_d],
\end{aligned}
\tag{7}
$$

where C_λ is a penalty constant relating to the scale factor λ.

2.7 Projection-based algorithm
The optimization problem of (7) is solved by an iterative algorithm, which maintains the constrains imposed on the rows of parameter matrices by projecting them onto the unit sphere.

- **Step 1** Let $\ell = 0$, and the learning speed be $0 < \gamma < 1$.
- **Step 2** Initialize the parameters

$$
\begin{aligned}
& \lambda^{(u)}[\ell] = 1_n, \lambda[\ell] = 1_{n_t}, \\
& \mathbf{U}^{(d)}[\ell]_{jk}, \mathbf{V}^{(d)}[\ell]_{tk}, \mathbf{Q}[\ell]_{ts} \sim \mathcal{N}(0,1), \\
& d \in [n_d], j \in [n], k \in [n_k], t \in [n_t], s \in [n_y],
\end{aligned}
\tag{8}
$$

where $\mathcal{N}(0,1)$ is the standard normal distribution.
- **Step 3** Normalize the rows of the optimization parameters by L_2 norm. Only the vector $\lambda[\ell]$ will be unnormalized.
- **Step 4** Set the scale value for the unnormalized vector λ by assuming that all other parameters are fixed. Compute

$$
\mathbf{F}[\ell] = \left(\circ_{d=1}^{n_d} A\left(\mathbf{XD}_{\lambda^{(u)}[\ell]}\mathbf{U}^{(d)}[\ell]\right)\mathbf{V}^{(d)T}[\ell]\right)
\tag{9}
$$

and solve the linear least square problem for $\lambda[\ell]$

$$
\min_{\lambda[\ell]} \quad ||\mathbf{Y} - \mathbf{F}[\ell]\mathbf{D}_{\lambda[\ell]}\mathbf{Q}[\ell]||^2.
\tag{10}
$$

- **Step 5** Compute the value of the objective function of the extended LTR problem given by (7)

$$
\mathbf{h}^{(uv)}[\ell] = ||\mathbf{Y} - \mathbf{F}[\ell]\mathbf{D}_{\lambda[\ell]}\mathbf{Q}[\ell]||^2.
\tag{11}
$$

- **Step 6** Compute the partial gradients of $\mathbf{h}^{(uv)}$ by applying (15),

$$
\begin{aligned}
& \nabla_{\mathbf{U}^{(d)}}h^{(uv)}, \nabla_{\mathbf{V}^{(d)}}h^{(uv)}, d \in [n_d], \\
& \nabla_{\lambda^{(u)}}h^{(uv)}, \nabla_\lambda h^{(uv)}, \nabla_\mathbf{Q}h^{(uv)}.
\end{aligned}
\tag{12}
$$

- **Step 7** Update the parameters

$$
\begin{aligned}
\lambda^{(u)}[\ell+1] &= \lambda^{(u)}[\ell] &+ \gamma\nabla_{\lambda^{(u)}[\ell]}h^{(uv)}, \\
\lambda[\ell+1] &= \lambda[\ell] &+ \gamma\nabla_{\lambda[\ell]}h^{(uv)}, \\
\mathbf{U}^{(d)}[\ell+1] &= \mathbf{U}^{(d)}[\ell] &+ \gamma\nabla_{\mathbf{U}^{(d)}[\ell]}h^{(uv)}, d \in [n_d], \\
\mathbf{V}^{(d)}[\ell+1] &= \mathbf{V}^{(d)}[\ell] &+ \gamma\nabla_{\mathbf{V}^{(d)}[\ell]}h^{(uv)}, d \in [n_d], \\
\mathbf{Q}[\ell+1] &= \mathbf{Q}[\ell] &+ \gamma\nabla_{\mathbf{Q}[\ell]}h^{(uv)}.
\end{aligned}
\tag{13}
$$

- **Step 8** Normalize the optimization parameters in L_2 norm.
- **Step 9** $\ell = \ell + 1$.
- **Step 10 Go to** Step 5.

For large-scale applications, the above algorithm is further extended by partitioning the training examples into mini-batches, and processing them sequentially. A single run of the cycle based on mini-batches is taken as an epoch, and repeated. To reduce the variance caused by the partition, a momentum-based update can be applied, e.g. Nestorov Accelerated Gradient method, or the ADAM method frequently applied for Deep Neural Networks (Kingma and Ba, 2015; Nesterov, 2005; Polyak, 1964).

2.7.1 Gradients
Let the matrix $\mathbf{H}^{(d)} \in \mathbb{R}^{m \times n_k}$ contain the partial derivatives of the activation function with respect to the components of the matrix $\mathbf{XD}_{\lambda^{(U)}}\mathbf{U}^{(d)}$, where we exploited that the activation function is a pointwise map of the matrix in its argument. We exploit the following expressions to shorten the gradient formulas

$$
\begin{aligned}
\mathbf{F}_{\backslash d} &= \circ_{b=1, b \neq d}^{n_d}\mathbf{A}^{(b)}\mathbf{V}^{(b)T}, \\
\mathbf{B}^{(d)} &= \mathbf{EQ}^T \circ \mathbf{F}_{\backslash d}, \\
\mathbf{C}^{(d)} &= \mathbf{X}^T\left(\mathbf{B}^{(d)}\left(\mathbf{V}^{(d)} \circ \lambda 1_{n_k}^T\right) \circ \mathbf{H}^{(d)}\right).
\end{aligned}
\tag{14}
$$

Then, the gradients are expressed in a compact form

$$
\begin{aligned}
\nabla_{\lambda^{(u)}}h^{(uv)} &= \frac{1}{mn_y}\sum_{d=1}^{n_d}\left(\mathbf{C}^{(d)} \circ \mathbf{U}^{(d)}\right)1_{n_k}, \\
\nabla_{\mathbf{U}^{(d)}}h^{(uv)} &= \frac{1}{mn_y}\left(\mathbf{C}^{(d)} \circ \lambda^{(u)}1_{n_k}^T\right), \\
\nabla_{\mathbf{V}^{(d)}}h^{(uv)} &= \frac{1}{mn_y}\left(\mathbf{B}^{(d)T}\mathbf{A}^{(d)} \circ \lambda 1_{n_k}^T\right), \\
\nabla_\lambda h^{(uv)} &= \frac{1}{mn_y}(\mathbf{F} \circ \mathbf{EQ}^T)^T1_m + \frac{C_\lambda}{n_t}\lambda, \\
\nabla_\mathbf{Q}h^{(uv)} &= \frac{1}{mn_y}(\mathbf{F}^T\mathbf{E} \circ \lambda 1_{n_y}^T).
\end{aligned}
\tag{15}
$$

2.8 Dataset
In order to evaluate the performance of *comboLTR*, we used the drug combination responses in human cancer cell lines from the NCI-ALMANAC study (Holbeck *et al.*, 2017). To exploit more data sources, especially multi-omics profiles of the cancer cell lines, we filtered the data to include only the cell lines for which gene expression, CNV, CRISPR-Cas9 gene knock-outs and proteomics data were available. The resulting dataset consisted of 828 324 response measurements of 5035 drug combinations and 15 396 monotherapies in 19 cancer cell lines originating from 9 tissue types. Each drug combination has been screened using 4×4 dose–response matrix design. The response measurements are given in the form of percentage growth of the cell lines with respect to a control. The distribution of our drug combination response dataset in 19 cell lines was identical to the distribution of all drug combination responses from the NCI-ALMANAC study, as shown in Supplementary Figure S1.
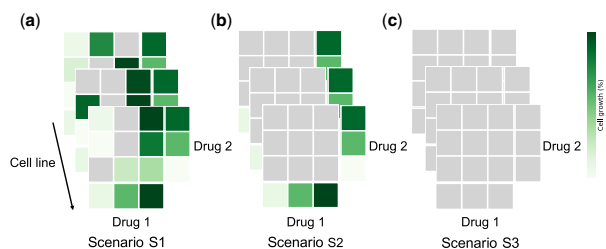
Fig. 2. Illustration of different drug combination response prediction scenarios. (**a**) Filling in the gaps in partially measured dose–response matrices (S1); predicting dose–response matrices of new drug combinations (**b**) with monotherapy responses (S2) and (**c**) without monotherapy responses (S3)

2.9 Feature representation

Each drug combination response is uniquely determined by five components, i.e. two drugs, their concentrations and a cell line. Such drug combination responses indexed by quintuplets form a fifth-order tensor (Fig. 1a). To flatten the higher-order tensor into a feature matrix, each such quintuplet is assigned a unique codeword by one-hot encoding the five components. The resulting tensor index features are similar to ones used in recommender systems (e.g. recommending movies to users). In addition to the tensor index features, the feature matrix also consists of auxiliary features, such as chemical and cell line descriptors, to include more available data sources (Fig. 1b).

As for chemical features, we used standard MACCS fingerprint, which consists of 166 chemical substructures. Each drug was matched with the substructures and represented as a binary feature vector describing whether a substructure was contained in the drug. Substructures present in all or none of the drugs were removed from the feature set, leaving 148 substructures in the end. For cell line features, multi-omics data including gene expression, CNV, CRISPR-Cas9 gene knock-outs and proteomics data were incorporated from DepMap data portal (Ghandi *et al.*, 2019; Meyers *et al.*, 2017; Nusinow *et al.*, 2020) to represent cell lines. Due to the large size of multi-omics data, in this case, more than 70 000 features, only 1% of the omics features with the highest variance across the 19 cell lines were selected, which resulted in 191 gene expression features, 276 CNV features, 174 CRISPR-Cas9 gene knock-out features and 69 proteomics features.

3 Results

We evaluated the performance of *comboLTR* in four practical prediction scenarios (Fig. 2):

- Filling in the gaps in partially measured dose–response matrices (S1).
- Prediction of complete dose–response matrices of new drug combinations with no available combination response measurements in any cell line, monotherapy response values present for both drugs (S2).
- Prediction of complete dose–response matrices of new drug combinations with no available combination and monotherapy response measurements in any cell line (S3).

Based on the results by Julkunen *et al.* (2020), predicting dose–response matrices of completely new drug combinations is the most challenging task from the above. Thus, we aimed to test our *comboLTR* framework under the difficult prediction scenario S2, and furthermore, under even more challenging prediction scenario S3. As shown in Figure 2, S3 has the least information available for a drug combination, since even monotherapy responses of single drugs are not present. Scenario S1 forms a relatively easy task, and thus it was considered as the reference prediction scenario. For

completeness, in addition to these three scenarios, following Julkunen *et al.* (2020), we also considered the scenario of predicting dose–response matrices of previously untested drug–drug–cell line triplets in the case where the response matrices of the same drug combination in other cell lines are known. However, as it was not our main focus in this study, the results for this prediction scenario are included in the Supplementary material only (Supplementary Tables S1 and Fig. S2). We also benchmarked the performance of *comboLTR* against RF and *comboFM*.

3.1 Model optimization via 5-fold cross validation

We applied 5-fold cross validation (CV) in all prediction scenarios in order to tune the model parameters and evaluate the predictive performance. In scenario S1, for each dose–response matrix, combination responses were randomly selected into test sets. In scenario S2, dose–response matrices of certain drug combinations in all cell lines were randomly selected into test sets. All monotherapy responses were kept in the training sets for S1 and S2 prediction scenarios. S3 is similar to S2 but with all monotherapy responses excluded from training set.

Based on our previous research, the degree of the polynomial function in *comboLTR* was set to five to model the interactions between the five components, i.e. two drugs, their concentrations and the cell line, which uniquely determine each drug combination response in the drug combination response tensor. A total of 20 drugs were randomly selected to subsample the full dataset for *comboLTR* parameter tuning. The subsampled dataset contained 31 095 response measurements of 208 drug combinations and 1901 monotherapies in 19 cell lines. The subsampled drug combination responses had almost identical distribution as our full dataset and also the whole dataset from NCI-ALMANAC study (Supplementary Fig. S1). Once parameters were determined, the performance of the model was evaluated using 5-fold CV on the full dataset, with the exception that the subsampled dataset was present in the training set only. *comboLTR* was evaluated using 5-fold CV for up to the 9th order with rank 200. Only very slight overfitting was observed in the highest order models.

We used a python implementation of RF from scikit-learn (Pedregosa *et al.*, 2011) and its default parameters for training and prediction. Parameters for *comboFM* were taken from the original publication (Julkunen *et al.*, 2020). In addition to fully evaluate the predictive performance of *comboFM* and RF, in the most challenging prediction scenario S3, their parameters were also optimized using the subsampled drug combination responses. The models with optimal parameters were then applied on the full dataset in the prediction scenario S3.

3.2 Prediction of anti-cancer effects of drug combinations

We used different feature combinations to train *comboLTR* model, as shown in Table 2. Since the one-hot encoded tensor indices are only positional features of the quintuplets, MACCS fingerprint and multi-omics data were used as auxiliary features to provide additional information on drugs and cell lines. In scenario S1, when filling in the gaps in partially measured dose–response matrices, the performance difference between feature combinations was negligible. Using only tensor index features resulted in the Pearson correlation between predicted and measured responses of 0.915, whereas adding auxiliary chemical and biological features led to the Pearson correlation of 0.922. However, in scenarios S2 and S3, when predicting the responses of completely new drug combinations, even without monotherapy responses present, adding auxiliary features clearly increased the prediction performance. In particular, in the most challenging scenario S3, using tensor indices only and additionally including auxiliary features, resulted in the Pearson correlations of 0.893 and 0.915, respectively. With the advantage of handling large feature vectors, *comboLTR* can harness data from different sources for the improved prediction performance. Thus, tensor index features, chemical and multi-omics auxiliary features were used in all further experiments.

Table 2. Performance of *comboLTR*, *comboFM* and RF under different prediction scenarios and using different features

Features	Method	S1	S2	S3
Tensor indices	*comboLTR*	0.915 ± 0.009	0.894 ± 0.002	0.893 ± 0.003
	comboFM	0.920 ± 0.010	0.914 ± 0.003	0.907 ± 0.004
	RF	0.886 ± 0.019	0.853 ± 0.010	0.858 ± 0.010
Tensor indices + MACCS	*comboLTR*	0.921 ± 0.010	0.908 ± 0.003	0.910 ± 0.003
	comboFM	0.923 ± 0.012	0.923 ± 0.005	0.913 ± 0.005
	RF	0.921 ± 0.016	0.872 ± 0.009	0.894 ± 0.005
Tensor indices + multi-omics	*comboLTR*	0.908 ± 0.014	0.909 ± 0.007	0.911 ± 0.005
	comboFM	0.910 ± 0.027	0.904 ± 0.014	0.870 ± 0.064
	RF	0.895 ± 0.019	0.859 ± 0.010	0.865 ± 0.010
Tensor indices + MACCS + multi-omics	*comboLTR*	0.922 ± 0.011	0.914 ± 0.006	0.915 ± 0.005
	comboFM	0.915 ± 0.012	0.889 ± 0.024	0.878 ± 0.064
	RF	0.923 ± 0.015	0.873 ± 0.009	0.896 ± 0.005

Note: Pearson correlations between predicted and measured drug combination responses, reported as averages across five CV folds ± SDs.

We used RF and *comboFM* as comparison methods in all prediction scenarios. Scatter plots of the predicted and measured drug combination responses are shown in Figure 3. As expected, in scenario S1, all three methods achieved comparable prediction performance. Pearson correlations for *comboLTR*, *comboFM* and RF were 0.922, 0.915 and 0.923, respectively. However, the difference in the performance of the methods became clearly visible in the most challenging and practical scenarios of predicting dose–response matrices of completely new drug combinations with and without monotherapy responses available (S2 and S3). Notably, under scenario S3, *comboLTR*, with a Pearson correlation of 0.915, clearly outperformed *comboFM* and RF (Pearson correlations of 0.878 and 0.896, respectively). It demonstrates that monotherapies play an important role in predicting higher-order interactions by *comboFM*. On the other hand, *comboLTR* produced more accurate predictions with fewer experimental measurements, which makes *comboLTR* more practical and applicable in recommending combination therapies.

To further study the prediction performance of those three methods, we investigated Pearson correlations for drug pairs in different drug classes and cell lines from different tissue types (Fig. 4). In general, *comboLTR* showed higher average Pearson correlation in most tissue types and drug classes. It was also corroborated by the violin plots that *comboLTR* shows better and more stable prediction performance across different drug classes and tissue types, particularly in the more challenging scenarios S2, and even S3 where less information was available.

Next, we evaluated the performance of the methods in quantifying the level of synergy and identifying highly synergistic drug combinations based on the predicted dose–response matrices. To calculate the synergy scores, we applied the NCI ComboScore introduced along with the NCI-ALMANAC dataset (Holbeck *et al.*, 2017). Scatter plots and Pearson correlations between the NCI ComboScores calculated based on the complete measured and predicted dose–response matrices of the three methods are shown in Supplementary Figure S3. RF performed particularly well in the simplest scenario S1, but *comboLTR* clearly outperformed the other two methods in the more challenging scenarios, e.g. with a Pearson correlation of 0.67 compared to 0.57 (*comboFM*) and 0.46 (RF), in the S3 scenario. We also conducted discrimination analyses using the precision-recall (PR) curves (Supplementary Fig. S4) and receiver operating characteristic curves (Supplementary Fig. S5) to further evaluate the model performance in classifying drug combinations as synergistic versus non-synergistic with varying thresholds for synergy, in the three prediction scenarios. *comboLTR* showed very competitive performance in discriminating highly synergistic drug combinations across several top-% synergy thresholds in the most challenging prediction scenarios. For example, in scenario S3, the areas under the PR curve at a synergy threshold of 5% for *comboLTR*, *comboFM* and RF were 0.25, 0.16 and 0.12, respectively (Supplementary Fig. S4).

To investigate the importance of multi-omics features, contribution of each type of omics data to the model performance was evaluated by 'leave one type of omics data out' and 'adding only one type of omics data' 5-fold CVs (Table 3). First, from the feature set comprising tensor indices, MACCS fingerprints and multi-omics data, each type of omics data were excluded to test their contribution to the predictive performance. Then, the predictive performance was also evaluated by including each type of omics data into the feature set, on top of tensor indices and MACCS fingerprints. The prediction performance was relatively stable when including or excluding certain types of omics data.

To investigate the importance of individual features, after the model was trained, each feature column was randomly permuted 20 times, and the average Pearson correlation difference between the models with original and permuted feature matrices was calculated as a measure to evaluate individual feature contribution to the predictive performance (Supplementary Fig. S6.). Weights of features from trained *comboLTR* model were also extracted and plotted in the form of a heatmap of L2-norm of each feature set, including MACCS fingerprints, gene expression, CNV, CRISPR-Cas9 gene knock-outs and proteomics data (Supplementary Fig. S6). In general, the most weight has been placed on tensor indices, while MACCS fingerprints and multi-omics data had only a relatively minor contribution to the model accuracy.

4 Discussion

Drug combinations are emerging as a powerful treatment modality to combat complex multi-factorial disorders, including cancer. Machine-learning models can significantly speed-up the search for effective drug combination therapies by systematically prioritizing the most promising combinations for further experimental validation. Using existing drug combination response data, we developed *comboLTR* to efficiently recommend combination therapies for cancer. Table 2 shows that *comboLTR* produces accurate and stable predictions even without monotherapy responses present in the training data. Monotherapy responses, which contain individual drugs tested in different cancer cell lines in various concentrations, are also costly and time-consuming to obtain in the lab. Without such limitation, *comboLTR* is more applicable and practical, especially in clinical research.

We compared the performance of *comboLTR* to RF and recently introduced *comboFM* method. Both *comboLTR* and *comboFM* are based on polynomial regression. A performance difference between the two methods in prediction scenario S3, where dose–response matrices of completely new drug combinations without monotherapy responses are inferred, could be due to different forms of polynomial functions to be learned. In *comboLTR*, a complete polynomial of degree n_d is used, while in *comboFM*, it is restricted to the symmetric polynomials. Since the monotherapies induce asymmetry in the drug representation, a

(a) S1: Filling in the gaps in partially measured dose-response matrices
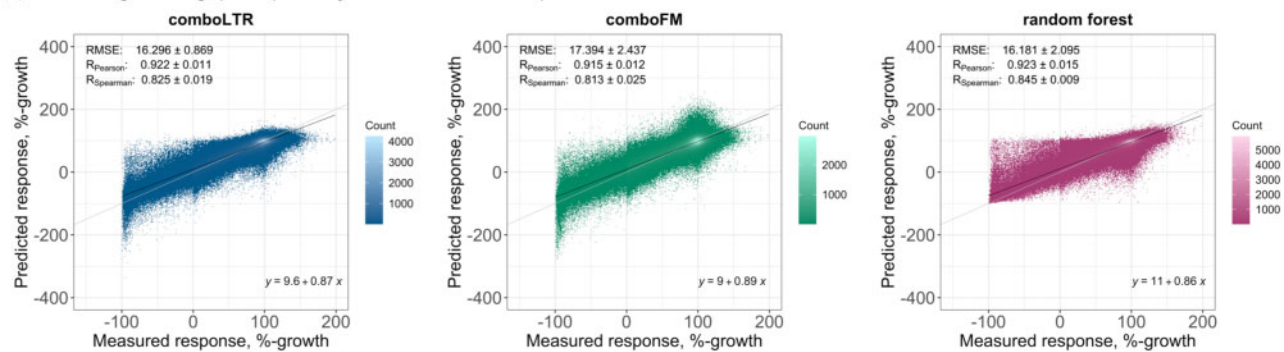
(b) S2: Predicting dose-response matrices of new drug combinations
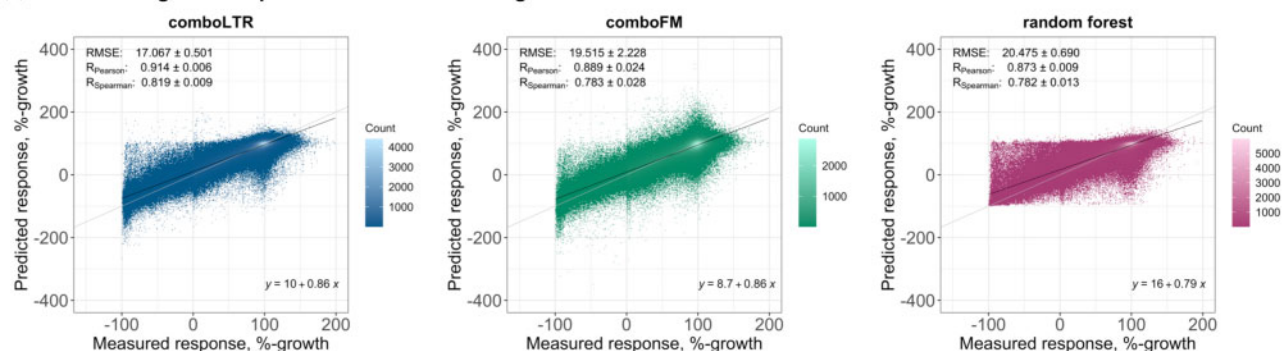
(c) S3: Predicting dose-response matrices of new drug combinations w/o monotherapy responses
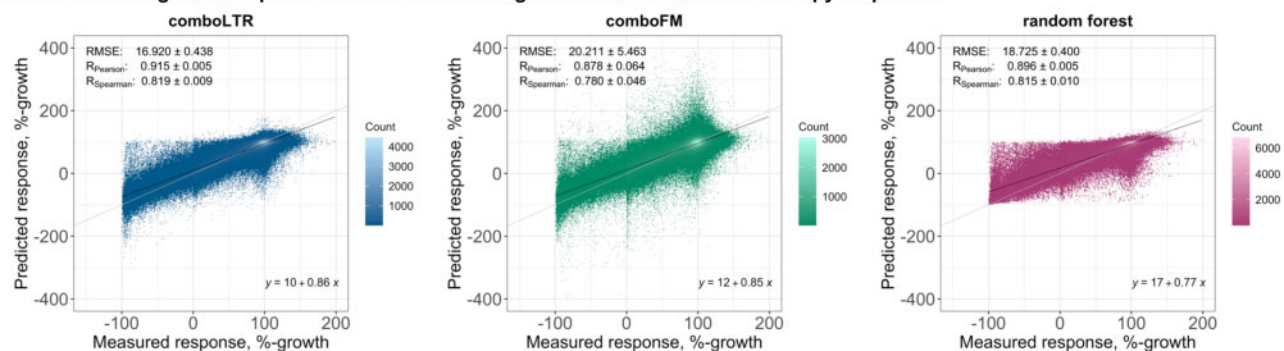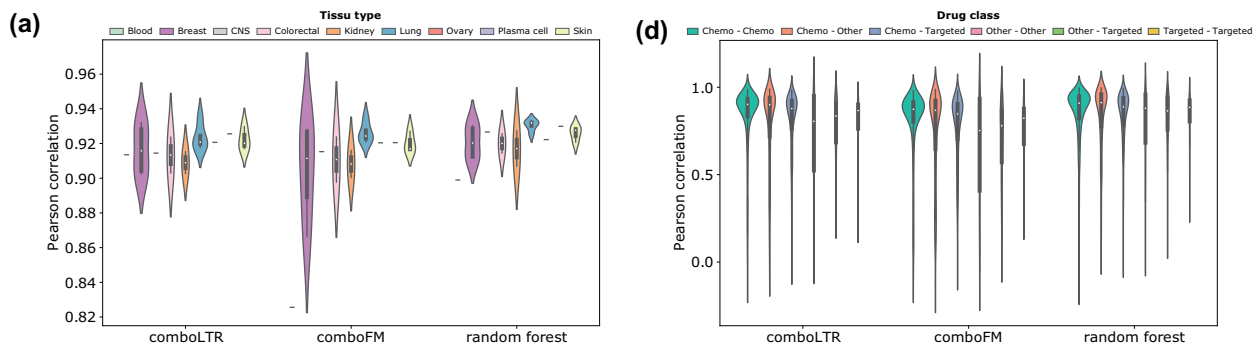

Fig. 3. Predictive performance of *comboLTR*, *comboFM* and RF in three drug combination response prediction scenarios. Scatter plots between the predicted and measured dose–dependent drug combination effects in the form of %-growth of cancer cell lines. The predictions were made under three scenarios of (a) filling in the gaps in partially measured dose–response matrices, inferring dose–response matrices of completely new drug combinations with (b) and without (c) monotherapy responses available. Root mean squared error, Pearson correlation and Spearman correlation are reported as averages ± SDs over five CV folds. Diagonal line and linear fit are also displayed in each scatter plot. Note, different x- and y-axes ranges in the plots that are consistent across the panels

method which can cover the full range of possible polynomials has higher probability to combine both the symmetric and asymmetric relations. The symmetry restriction of *comboFM* also eliminates the monomials with higher-order occurrences of the polynomial variables, which could also reduce further the range of functions to be approximated. Thus, it could be hypothesized that *comboFM* may rely more on lower-order interactions, such as monotherapy responses. The lack of such lower-order interaction information results in a performance drop of *comboFM* in the prediction scenario S3. We note that *comboFM* is a bit more competitive without using the MACCS and multi-omics features than when using them, which may be at least in part caused by the symmetricity of the polynomials with respect to the variables. Specifically, the tensor indices and the MACCS and multi-omics features are treated alike by *comboFM*, but not *comboLTR*. Tensor indices, however, do not allow any explanation of the predictions in terms of underlying biological functions or processes, and hence models relying solely on them might not be preferable in practical use.

Furthermore, in practical clinical applications, identification of synergistic drug combinations is of high interest. Importantly, *comboLTR* outperformed *comboFM* and RF in the most challenging prediction scenarios also in the task of discriminating highly synergistic drug combinations based on the predicted complete dose–response matrices and when using a range of different top-% synergy thresholds (Supplementary Figs S4 and S5).

Since CV experiments were run on different machines, the longest running time and median memory across all prediction scenarios were recorded (Table 4). Compared to *comboFM*, which also learns from higher-order interactions, *comboLTR* is significantly more time-efficient, especially in handling large feature vectors (Table 4). When training models using the full dataset with tensor index features, MACCS fingerprints and multi-omics information, only 3.1 h were needed for a 5-fold CV using *comboLTR*, whereas *comboFM* took 39 h. Most of the memory was used for storing the data.

S1: Filling in the gaps in partially measured dose-response matrices

S2: Predicting dose-response matrices of new drug combinations
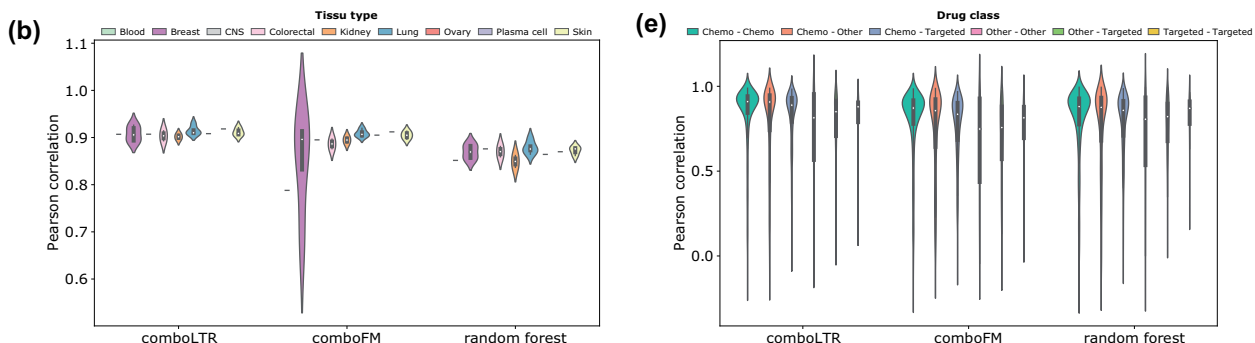
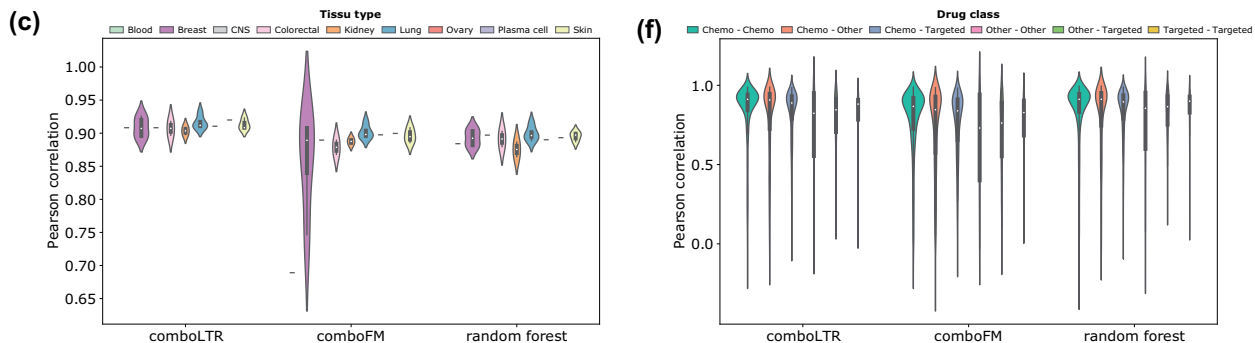## S3: Predicting dose-response matrices of new drug combinations w/o monotherapy responses

Fig. 4. Predictive performance of *comboLTR*, *comboFM* and RF across tissue types and drug classes in three drug combination response prediction scenarios. Violin plots were used to characterize Pearson correlations of predicted and measured drug combination responses across tissue types (**a–c**) and drug classes (**d–f**). Note that the order of tissue types and drug classes in the legends corresponds to their order in the violin plots

Table 3. The 5-fold CV results of *comboLTR*, in the form of Pearson correlations, when leaving one type of omics data out or including only one type of omics data, compared with using the full multi-omics data, on top of tensor indices and MACCS fingerprints

Multi-omics feature combination	S1	S2	S3
Full multi-omics	0.922	0.914	0.915
Excluding gene expression	0.921	0.912	0.914
Excluding CNV	0.921	0.916	0.914
Excluding CRISPR knock-out	0.921	0.914	0.913
Excluding proteomics	0.920	0.914	0.914
Using only gene expression	0.921	0.912	0.911
Using only CNV	0.921	0.911	0.910
Using only CRISPR knock-out	0.921	0.909	0.911
Using only proteomics	0.921	0.910	0.910

Table 4. The time (h) and memory (GB) usage of *comboLTR*, *comboFM* and RF in 5-fold CV

Features	Time (h)/memory (GB)		
	comboLTR	*comboFM*	RF
Tensor indices	1.1/24	1.2/20	1.2/24
Tensor indices + MACCS	2.1/35	5.9/35	0.4/38
Tensor indices + multi-omics	3.2/44	31.5/63	5.6/54
Tensor indices + MACCS + multi-omics	3.1/74	39.0/73	2.1/74

In this work, genomic, transcriptomic and proteomic data were used to provide additional information on cancer cell lines. The complex interactions among and within multiple layers of omics

measurements form a comprehensive molecular network. For example, CNVs affect expression of genes which, together with post-translational modifications, influence the quantity of proteins. Since the full multi-omics dataset included over 70 000 features, we reduced the dimensionality of the data by selecting only 1% of each type of omics data as auxiliary descriptors characterizing cell lines (see Section 2.7).

As shown together in Tables 2 and 3 and Supplementary Figure S6, the performance increase brought by integrating multi-omics data into the model was only modest. This phenomenon could be due to that only 1% of the full multi-omics dataset was taken into account, which represented only a small part of the cell's characteristics. Besides, the complex interactions of the molecular network were not taken into consideration in this experiment. For example, different drugs at various concentrations may result in multiple perturbations to the molecular network, which will change the cellular phenotype into diverse states. The lack of integrating such complicated interactions could lead to less relevant features selected for predicting drug combination responses. One of our future aims is to select cell features based on their connections with tested drug combinations, e.g. to assign weights to features based on their interaction strength with drugs.

On the other hand, MACCS fingerprint had slightly higher associated feature weights and performance increase, when compared to the multi-omics features. MACCS fingerprint characterizes drugs by the presence or absence of specific chemical substructures. As shown in Supplementary Figure S7, different drugs in our dataset have common substructures due to the limited number of substructures defined by MACCS fingerprint. Such property is expected to be helpful especially in scenarios S2 and S3 where new drug combinations were predicted. It could be speculated that more weight would be placed on the multi-omics features in the scenario of predicting drug combination responses in the cell lines outside of the training data.

5 Conclusions

In this work, we have put forward a novel approach for predicting responses of cancer drug combinations. Our method, *comboLTR*, is based on representing high-degree polynomial regression models through learning a factorization of the parameter tensor containing the unknown regression coefficients. We demonstrated the competitive predictive performance and time efficiency of the *comboLTR* method on the large NCI-ALMANAC dataset (Holbeck *et al.*, 2017).

The results indicate that *comboLTR* is a practical tool for prediction and prioritization of new drug combinations for pre-clinical and clinical evaluation. The ability to predict full dose–response matrices enables a detailed exploration of drug response landscapes and application of different synergy models.

Acknowledgements

Funding

References

Al-Lazikani,B. *et al.* (2012) Combinatorial drug therapy for cancer in the post-genomic era. *Nat. Biotechnol.*, 30, 679–692.

Bansal,M. *et al.* (2014) A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.*, 32, 1213–1222.

Blondel,M. *et al.* (2016) Higher-order factorization machines. *Adv. Neural Inf. Process. Syst.*, 29, 3351–3359.

Das,P. *et al.* (2018) A survey of the structures of US FDA approved combination drugs. *J. Med. Chem.*, 62, 4265–4311.

de Silva,V. and Lim,L.-H. (2008) Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM J. Matrix Anal. Appl.*, 30, 1084–1127.

Ghandi,M. *et al.* (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, 569, 503–508.

Golub,G.H. and Loan,C.F.V. (2013) *Matrix Computations*. 4th edn. The Johns Hopkins University Press, Baltimore, MD.

Holbeck,S. *et al.* (2017) The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res.*, 77, 3564–3576.

Ianevski,A. *et al.* (2020) SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res.*, 48, W488–W493.

Julkunen,H. *et al.* (2020) Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nat, Commun.*, 11, 1–11.

Kingma,D. and Ba,J. (2015) Adam: a method for stochastic optimization. In: *International Conference on Learning Representations*, San Diego.

Kolda,T.G. and Bader,B.W. (2009) Tensor decompositions and applications. *SIAM Rev.*, 51, 455–500.

Lathauwer,L.D. *et al.* (2000) A multilinear singular value decomposition. *J. Matrix Anal. Appl.*, 21, 1253–1278.

Li,H. *et al.* (2019) TAIJI: approaching experimental replicates-level accuracy for drug synergy prediction. *Bioinformatics*, 35, 2338–2339.

Menden,M.P. *et al.* (2019) Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat. Commun.*, 10, 2674.

Meyers,R.M. *et al.* (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.*, 49, 1779–1784.

Nesterov,Y. (2005) Smooth minimization of non-smooth functions. *Math. Program.*, 103, 127–152.

Nusinow,D.P. *et al.* (2020) Quantitative proteomics of the cancer cell line encyclopedia. *Cell*, 180, 387–402.

Pedregosa,F. *et al.* (2011) Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, 12, 2825–2830.

Pemovska,T. *et al.* (2018) Recent advances in combinatorial drug screening and synergy scoring. *Curr. Opin. Pharmacol.*, 42, 102–110.

Polyak,B.T. (1964) Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.*, 4, 1–17.

Prenter,P.M. (1970) A Weierstrass theorem for real, separable Hilbert spaces. *J. Approx. Theory*, 3, 341–351.

Preuer,K. *et al.* (2018) DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*, 34, 1538–1546.

Rendle,S. (2010) Factorization machines. In: *ICDM '10: Proceedings of the 2010 IEEE International Conference on Data Mining*. pp. 995–1000. IEEE Computer Society.

Reymond,J.L. and Awale,M. (2012) Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.*, 3, 649–657.

Sidorov,P. *et al.* (2019) Predicting synergism of cancer drug combinations using NCI-ALMANAC data. *Front. Chem.*, 7, 509.

Szedmak,S. *et al.* (2020) A solution for large scale nonlinear regression with high rank and degree at constant memory complexity via latent tensor reconstruction. *arXiv preprint*, arXiv:2005.01538.

Tang,J. *et al.* (2015) What is synergy? The Saariselkä agreement revisited. *Front. Pharmacol.*, 6, 181.

Tonekaboni,M. *et al.* (2018) Predictive approaches for drug combination discovery in cancer. *Brief. Bioinform.*, 19, 263–276.

Yang,M. *et al.* (2020) Stratification and prediction of drug synergy based on target functional similarity. *Syst. Biol. Appl.*, 6, 1–10.