DTF: Deep Tensor Factorization for Predicting Anticancer Drug Synergy

Zexuan Sun, Shujun Huang, Peiran Jiang and Pingzhao Hu*

1Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, R3E 0J9, Canada; 2School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China; 3College of Pharmacy, University of Manitoba, Winnipeg, Manitoba, R3E 0T5, Canada; 4Department of Bioinformatics & Systems Biology, Huazhong University of Science and Technology, Wuhan, 430074, China; 5Department of Computer Science, University of Manitoba, Winnipeg, Manitoba, R3E 0J9, Canada; 6Research Institute in Oncology and Hematology, CancerCare Manitoba, Winnipeg, R3E 0V9, Canada

*To whom correspondence should be addressed.

Abstract

**Motivation:** Combination therapies have been widely used to treat cancers. However, it is cost- and time-consuming to experimentally screen synergistic drug pairs due to the enormous number of possible drug combinations. Thus, computational methods have become an important way to predict and prioritize synergistic drug pairs.

**Results:** We proposed a Deep Tensor Factorization (DTF) model, which integrated a tensor factorization method and a deep neural network (DNN), to predict drug synergy. The former extracts latent features from drug synergy information while the latter constructs a binary classifier to predict the drug synergy status. Compared to the tensor-based method, the DTF model performed better in predicting drug synergy. The area under the curve (AUC) of the receiver operating characteristic was 0.92 for DTF and 0.88 for the tensor method. We also compared the DTF model with random forest and logistic regression models, and found that the DTF outperformed the two methods. A further look at the predictive performance of the DTF on the basis of individual cell lines found that the DTF showed an AUC greater than 0.90 for majority of the cell lines. Applying the DTF model to predict missing entries in our drug-cell line tensor, we identified novel synergistic drug combinations for 27 cell lines from the six cancer types. A literature survey showed that some of these predicted drug synergies have been identified in vivo or in vitro. Thus, the DTF model could be a valuable in silico tool for prioritizing novel synergistic drug combinations.

**Availability:** Source code and data is available at https://github.com/ZexuanSun/DTF-Drug-Synergy

**Contact:** pingzhao.hu@umanitoba.ca

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

Though monotherapy has contributed a lot to helping cure many human diseases, it has several evident drawbacks, such as acquired resistance or low efficiency [1, 2]. The complexity of human diseases is usually resulting from the complex interactions of different phenomic and genomic factors. Thus, single drug, which typically targets on a single protein or pathway, is usually hard to treat the complex diseases well. To solve this dilemma, there comes combinatorial drug therapy, which uses a pair of or more drugs simultaneously to treat a specific disease. The synergistic effect of certain drug pairs can potentially improve the curative effect significantly. For instance, pentamidine and chlorpromazine do not exhibit any traces of inhibiting tumor activities while being used individually, however, the combination of these two drugs is able to inhibit the growth of tumor efficiently. What’s more, the drugs used for evaluating drug synergetic effect usually employ existing drugs, which have been studied thoroughly and approved by Food and Drug Administration (FDA) for treating specific diseases. This will save lots of time for clinical trials of the safety of these drug combinations. As a result, drug combination therapies have become a more and more popular treatment option for complex diseases. However, how to identify the drug pairs with drug synergistic effect is still challenging since the search space of the drug pairs from the drugs approved by FDA is huge. It is too time-consuming and unrealistic to implement clinical assays on all drug pairs. Therefore, computational methods for predicting drug pairs with strong synergistic effect are in great demand.

Currently, there are many computational methods for predicting relevant drug pairs. These include both traditional machine learning methods and deep learning methods. For example, Sidorov et al. proposed models for drug synergy prediction based on random forest (RF) and extreme gradient boosting (XGBoost) [3]. The physicochemical properties of drugs were used as the input of the models. Zhang et al. developed a model, AuDNNsynergy, based on deep learning method, which took advantage of the gene expression, copy number and genetic mutation data coming from cancer cell lines to predict drug pairs with high synergistic effect [4].

However, these methods have not made use of structure of the drug synergy data as a multi-way data (e.g. a data set can be represented as a multidimensional array). In fact, multi-way data reflects a structure of multi-way relations, which can be best represented as a multi-way array, that is, so-called tensor [5]. Tensor decomposition methods are utilized to decompose a given tensor constructed from raw data to capture latent relations between variables, which can be used for discovering hidden patterns or performing classifications. However, classic algorithms to decompose tensors cannot handle those tensors with missing values, which is a common issue in predicting drug synergistic effect. Some studies were conducted to solve the problem based on novel tensor frameworks. For example, Chen and Li proposed DrugCom, a tensor-based model, which incorporated multiple different existing data sources related to drugs and diseases. DrugCom decomposed a tensor with missing values by integrating existing knowledge at the same time to get the latent information of drug synergy, and demonstrated a high prediction performance over other methods [6]. Abe et al. expanded the most well-known tensor factorization method CANDECOMP/PARAFAC (CP) as a weighted least squares problem that uses a first-order optimization approach to handle the missing values in a given tensor. The new approach was called as CP-WOPT (CP Weighted OPTimization) [5]. This approach can capture the latent structure of the data via a higher-order factorization.

Although tensor-based factorization approach is efficient to represent multi-way data, there is still a much need to improve its prediction performance. Recently, deep learning methods have been shown to predict drug pairs with high synergy scores. Our study tried to combine tensor-based framework and deep learning methods together to predict synergistic effect of drug pairs. We proposed a Deep Tensor Factorization (DTF) model, which is comprised mainly by a tensor factorization method and a deep neural network (DNN). We first use the algorithm, CP-WOPT, to decompose tensor with missing entries, then the results of the tensor decomposition are served as features to be used to train the DNN model, which can predict the synergistic effect of drug pairs. This strategy allows the DTF model to capture the structure of multi-way data and learn more latent information with the help of deep learning method, therefore to enhance its overall performance.

2 Materials and Methods

Our model design is shown in Fig.1. The DTF model for synergistic drug combination prediction is based on two sub-models: a special tensor decomposition model to decompose the tensor with missing values and a DNN model to predict the drug synergy status.

2.1 Data Collection and Preprocessing

The data we mainly used is the drug-drug synergy (DDS) data derived from O’Neill et al.’s study [7]. We got 23,062 drug combinations with the corresponding Loewe synergy scores measured among 38 drugs in 39 cell lines, which come from 6 human cancer types (Figure 1A). We also refer the Loewe synergy score as Loewe additivity.

We used the DDS data to construct the tensor. Since we have three variables, i.e. drug A, drug B and cell line, we built a 3-order tensor with three axes representing the three variables, respectively. The value of each entity of the tensor is the synergy score corresponding to a specific drug pair and cell line. We also recorded the positions of the missing values in the tensor for the convenient of tensor decomposition. For some specific cell lines, there were experiments carried out multiple times for the same drug pairs. In order to construct the three-dimensional (3D) drug-drug-cell-line tensor, we averaged these scores for the same drug-drug pairs. It

---

![Fig. 1](image-url)
Predicting anticancer drug synergy

should be also noted that the synergy score of drug A and drug B is the same as that of drug B and drug A. Therefore, for each cell line the matrix formed by the synergy of drug pairs is symmetrical. We set the diagonal of each matrix to zero since there is of course no synergistic effect for the pair of the same drugs.

We also used the DDS data to generate classification labels or synergy status for training the DNN. We select the threshold as 30, which was used in the study of Kristina Preuer et al. [8]. That is to say, if the synergy score of a given drug pair is greater than 30, the synergy status is 1, otherwise, it is 0. We treated the entities with synergy status 1 as positive samples, and those with synergy status 0 as negative samples. The numbers of positive and negative samples for each cell line are shown in Figure 2.

2.2 DTF: Deep Tensor Factorization

2.2.1 Notations

We define the dimensionality of a tensor and multi-way data as order, and refer each dimension as a mode. We use a lowercase letter (a) to denote a scalar, a boldface lowercase letter (A) to denote a vector, a boldface capital letter (A) to denote a matrix and a boldface Euler script letter (𝒜) to denote a tensor. Every element of a tensor is denoted by a lowercase letter with a subscript. For a three order tensor 𝑋, its entries can be represented as 𝑥𝑖𝑗𝑘.

Subarrays (or subfields) can be created by fixing some of the given tensor’s indices. If we fix all but one index, slices (or slabs) are created. If we fix all but two indices, there come fibers. It is easy to find out that for a third order tensor each slice is actually a matrix.

We use ⊗ to represent a multi-way vector outer product, which is a tensor and each entry of the tensor is the product of corresponding elements in vectors. For example, the vector outer product of 3 vectors, a, b, c is a 3 dimensional tensor 𝑋, where (𝑋)𝑖𝑗𝑘 = 𝑎𝑖𝑏𝑗𝑐𝑘.

Say we have two same-sized tensors 𝑋 and 𝑌, which are of size 𝑙1 × 𝑙2 × … × 𝑙𝑛. We define their Hadamard (elementwise) product as 𝑋 ⊗ 𝑌 where (𝑋 ⊗ 𝑌)𝑖1⋯𝑛 = 𝑥𝑖1⋯𝑛 ⋅ 𝑦𝑖1⋯𝑛 for all 𝑖 ∈ {1, …, 𝑙1} and 𝑛 ∈ {1, …, 𝑛}.

The inner product of two same-sized tensors 𝑋 and 𝑌, is defined as

\[ (𝑋, 𝑌) = \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \cdots \sum_{k=1}^{l_n} x_{i_1\cdots i_n} y_{i_1\cdots i_n}. \]

For a 𝑙1 × 𝑙2 × … × 𝑙𝑛 sized tensor 𝑋, we define its norm as \[ \|X\| = \sqrt{(X, X)}. \] Recall that for matrices and vectors, \( \| \| \) can be referred as Frobenius-norm and two-norm, respectively. We are also able to give the definition of a weighted norm. Say 𝑋 and 𝑊 are two same-sized tensors, then we can define the 𝑊-weighted norm of 𝑋 as \[ \|X\|_W = \| WX \|. \]

The rank of a N-way tensor is 1 if an outer product of N vectors equals to this tensor.

A N-way tensor is of rank-1 if it can be strictly decomposed into the outer product of N vectors. We define the rank of a tensor 𝑅 as the minimum number of rank-one tensors which are required to get 𝑋 as their sum. For instance, a rank-R 3D tensor can, therefore, be written as \[ 𝑋 = \sum_{r=1}^{R} a_r \otimes b_r \otimes c_r = [\mathbf{A}, \mathbf{B}, \mathbf{C}] \] where the matrices A, B, C as factor matrices since they collect vectors from the rank-one components and hold them as columns. It is known that the problem of computing the rank of a tensor is NP-hard. Thus, in practice, we cannot know the exact rank of the tensor we investigate.

2.2.2 Tensor Decomposition Algorithms

Rank decomposition is one of the most popular tensor decomposition methods, which stems from the definition of tensor rank. The key idea underlying rank decomposition is to use the sum of a sequence of rank-one tensors to approximate the original tensor. CANonical DECOM-Position (CANDECOMP) and the PARAllel FACtors (PARAFAC) decompositions are the most popular rank decomposition approaches, which were proposed in different knowledge domains independently. Interestingly, both of them follow similar rules, so we usually name the methods as the CANDECOMP/PARAFAC or canonical polyadic decomposition (CPD) [9]. For a particular 3D tensor (Figure 3), the CPD algorithm is to optimize:

\[ \min_{\mathbf{X}} \|\mathbf{X} - \tilde{\mathbf{X}}\|, \text{ where } \tilde{\mathbf{X}} = \sum_{r=1}^{R} a_r \otimes b_r \otimes c_r = [\mathbf{A}, \mathbf{B}, \mathbf{C}] \] (2)

More details of the algorithm can be found in [9].

2.2.3 CP-WOPT

Equivalently, for third-order tensors, the CP decomposition can be treating as optimizing the objective error function as below:

\[ f(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \frac{1}{2} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \sum_{k=1}^{l_3} (x_{ijk} - \sum_{r=1}^{R} a_r b_r c_r)^2. \] (3)

When a tensor has missing values, similarly, we can do the same thing as equation (3). What we need to do is to define a nonnegative weighted tensor 𝑊, which is the same size as 𝑋. For each element of 𝑊, namely 𝑊𝑖𝑗𝑘 is 1, if the corresponding element of 𝑋, i.e. 𝑥𝑖𝑗𝑘 is known, otherwise, it is 0. We are able to get the weighted version of (3) as follow:

\[ f_WXYZ(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \frac{1}{2} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \sum_{k=1}^{l_3} w_{ijk} (x_{ijk} - \sum_{r=1}^{R} a_r b_r c_r)^2. \] (4)

Fig. 2. Sample size for all cancer cell lines
We can easily generalize this to $N$-way tensor, and get the $N$-way weighted objective function. Let $X$ be a tensor of size $l_1 \times l_2 \times \cdots \times l_N$ and assume the rank of $X$ is $R$. Then we can rewrite the objective function using matrices as follow:

$$f_w(A^{(1)}, A^{(2)}, \ldots, A^{(n)}) = \frac{1}{2} \|[X - [A^{(1)}, A^{(2)}, \ldots, A^{(n)}]]]_W\|^2.$$

Our goal is to find matrices $A^{(n)} \in \mathbb{R}^{m_n \times R}$, for $n=1, \ldots, N$ that minimize the equation (5). The gradient of the function can be computed using the approach developed in [10]. After having the function and gradient, we can use any gradient-based optimization method [11] to solve this optimization problem.

We employed a first-order optimization approach, to be specific, the L-BFGS-B algorithm proposed by Richard H et al. [12], to solve the weighted least squares problem. This algorithm functions as a gold standard tool to solve large nonlinear optimization problems with simple bounds described. It develops a limited memory BFGS matrix to approach the Hessian of the objective function. The algorithm was devised to make good use of the form of the limited memory approximation to carry out the algorithm efficiently [12].

As mentioned previously, the rank of a tensor is often unknown, but results in [13] showed that direct optimization methods have a better performance than alternating least square approaches when the rank is over-estimated. So bearing this fact in mind, we can set the number of components $R$ relatively large in the beginning and check the results of this decomposition. If the results are within our tolerance, we decrease $R$ and check the results again. If they are still good enough, we tend to choose a relatively larger $R$.

Similar to CP decomposition, the results of CP-WOPT are a sequence of rank-one tensors. Actually we can collect the vectors for each dimension and write the results in the form of factor matrices. If the tensor $X$ to be decomposed is of order 3, the results of CP-WOPT can be represented as $\{A, B, C\}$. If we pick up the $r$-th column vectors of these three matrices, the vector outer product of these three vectors is the $r$-th rank-one tensor of the results of the CP-WOPT decomposition. With these rank-one tensors, we are capable of reconstructing the original tensor. Let $X'$ be the sum of these rank-one tensors be $X'$. For each entry of $X'$, $x'_{ijk}$ can be written as the sum below:

$$\sum_{r=1}^R a_{ir} b_{jr} c_{kr}.$$

where $a_{ir}$ is the $i$-th element of the $r$-th column vector of $A$, $b_{jr}$ is the $k$-th element of the $r$-th column vector of $B$ and $c_{jr}$ is the $j$-th element of the $r$-th column vector of $C$. For each element of $X$, $x_{ijk}$, no matter it is known or unknown, the $x'_{ijk}$ of $X'$ is corresponding to it. For the known entries, $x'_{ijk}$ and $x'_{ijk}$ should be pretty close to each other, since it is the goal of our optimization. Since $x'_{ijk}$ can be represented in the form of a sum, each $x'_{ijk}$ corresponds to a sum of elements coming from factor matrices, $A$, $B$ and $C$.

### 2.2.4 Deep Neural Network

The structure of the neural network we employed in our DTF model is a fully connected neural network with $D$ layers, where the $d$-th layer contains $U_d$ neurons. We can regard this as a $D$-layer perceptron. We use $r^{(n)}_d$ to denote the input of the $n$-th sample into the $d$-th layer. Let $h(\cdot)$ be an activation function, then the result of activation is $a^{(n)}_d = h(r^{(n)}_d)$. Particularly, for the input layer, $a^{(n)}_0 = r^{(n)}_d$.

Typically, we used forward propagation to calculate the input of next layer. For instance, for the input of $(d+1)$-th layer ($0 \leq d < D$), $r^{(n)}_{d+1}$ can be given by the equation $r^{(n)}_d = XW_d a^{(n)}_d + b_d$, where $W_d$ is a $U_{d+1} \times U_d$ matrix and $b_d$ is a bias. We optimized the parameters of the DNN’s $W_d$ and $b_d$ in order to minimize the loss function $F = \sum_{n} \log(p(y^{(n)}, y^{(n)}))$, where $y^{(n)}$ is the predicted probability of the synergistic drug pairs and $y^{(n)}$ is the true synergy status. In our model, since we defined the prediction problem as a binary classification problem, we used binary cross entropy $\sum_{n} -y^{(n)} \log(y^{(n)}) + (1 - y^{(n)}) \log(1 - y^{(n)})$ as the loss function.

We employed RMSProp (Root Mean Square Propagation) to optimize the parameters of the DNN. The key of RMSProp is to first maintain the running average of the squared gradients for each weight, and then divide the gradient by square root the mean square. RMSProp has exhibited excellent adaptation of learning rate in different applications. We can regard RMSProp as an extension of Rprop, which is good at working with not only full-batch, but also mini-batches [14].

### 2.2.5 Model Construction

To build the DTF model to predict synergistic drug pairs, we need to link the CP-WOPT and DNN models together. After data collection and preprocessing, we built the 3-order tensor with missing values $X$ derived from original DDS data. Note that the value for each element of $X$ is the original drug synergy score rather than the 0/1 labels which were generated by binarization. The values of the unknown entries of $X$ do not matter, since they were ignored during the computation process of CP-WOPT. In our model, we simply set them as 0, and we recorded the positions of the missing values in a position tensor $P$, which is of the same size as $X$. A particular entry of $P$ is 1, if the corresponding element of $X$ is known, otherwise is 0.

Let $R$ be the number of components, then we implemented CP-WOPT on $X$ (Figure 1B), which required three parameters: the tensor $X$ to be decomposed, the position tensor $P$ and the number of components $R$ we wish to decompose the tensor into. As aforementioned, the decomposition results of CP-WOPT can be represented using factor matrices $\{A, B, C\}$. Each matrix collects the latent information of a specific dimension. To be specific, $A$ collects all the vectors corresponding to the latent information of drug A, B corresponding to drug B, and C corresponding to cell line. For each known synergy score $x_{ijk}$, there is a sum, i.e., $\sum_{r=1}^R a_{ir} b_{jr} c_{kr}$, corresponding to it (Figure 1B). Obviously, $(a_{ir}, \ldots, a_{ir})$ is the $i$-th row vector of $A$, which can be regard as features from the latent information of drug A. We applied the same principle to drug B and cell line. Finally, we got three feature vectors $(a_{ir}, \ldots, a_{ir}), (b_{jr}, \ldots, b_{jr}), (c_{kr}, \ldots, c_{kr})$ for a particular synergy score $x_{ijk}$. We collected them all together as the features to be used for training the DNN.

If we set the number of components of CPWOPT as $R$, then for drug A, drug B and cell line, the number of dimension of each of them will be $R$, which suggests the parameters between the input layer of the DNN and the first layer is a tensor $W_0$ of size $U_1 \times 3R$. The function TRAIN as shown in Figure 4 implemented the process of training the DNN, which used the features generated by the CP-WOPT and the labels from the binarization of synergy scores, namely, the synergy status. $(W_d)$ and $(b_d)$ are the parameters of the DNN, which defined the prediction process of the whole model. During the training process, we employed the classic forward propagation algorithm to calculate the results of the model being trained and the backward propagation algorithm to derive the gradient of the parameters of the DNN.

After training the DNN model, we used the trained model to predict the synergy status of drug pairs. For each unknown entry in the original tensor $X$, it is evident that there are also three feature vectors corresponding to
Predicting anticancer drug synergy

2.4 Model and Comparison Evaluation

To evaluate whether the deep learning method, i.e., the DNN is able to do a better job at extracting the latent features generated by CP-WOPT than other traditional models, we compared it with random forest (RF) and logistic regression (LR) models for predicting synergy status of drug pairs based on the features generated by CP-WOPT. Generally speaking, CP-WOPT can be treated as a binary classifier alone. Recall that we can make use of the results to reconstruct the original tensor, and for each unknown synergy status, the sum corresponding to it can be regarded as a tensor score. Since a higher synergy score means a better synergistic effect of a given drug pair, we can treat a higher tensor score measured from the reconstructed tensor as stronger synergistic effect of a given drug pair. Hence, for a given threshold of the tensor score, the CP-WOPT can be a binary classifier. Therefore, we also added the CP-WOPT classifier as a baseline of our proposed DTF model.

It should be noted that we treated the pairs of drug A-drug B and drug B-drug A as different drug combinations in the model evaluation process, which means the sizes of the train set and test set were double. Obviously, the synergy scores of these two pairs are the same. The main reason we did this is that a classifier can be claimed to have a good performance to predict the synergy status of the original pair correctly only if it can predict the synergy status of the two pairs correctly. Distinguishing only either of the symmetrical pairs would pull down the overall performance, but distinguishing one of the symmetrical pairs correctly is at least better than classifying both of the two pairs wrongly. This can be seen as doing two independent experiments on each pair simultaneously, making our experiments more scientific.

In order to evaluate and compare the models, we used stratified-sampling method to create unbiased test sets. To be specific, for each cell line, we randomly chose about 40% of the drug pairs with known synergy score as its test set and the remaining drug pairs with know synergy score as its training set. We then pulled all test sets and training sets from different cell lines as an merged test set and training set, respectively. We chose area under the curve (AUC) of the receiver operating characteristic (ROC) curve and prediction accuracy to evaluate these models. We repeated the stratified-sampling method 100 times, and the performances of the DTF model and other baseline models were evaluated 100 times, respectively. The AUC and accuracy were averaged and their standard deviations were also reported.

2.3 Software and Global Parameters

CP-WOPT was implemented in Matlab tensor_toolbox version 3.1 and the Matlab wrapper of L-BFGS-B was written by Stephen Becker [15]. We employed the Keras version 2.2.4 and scikit-learn version 0.21.3 to implement the DNN model and other machine learning models.

The tensor we constructed is of order 3, which has 3 axes, representing drug A, drug B and cell line. The dimension of drug A and drug B is 38, and that of cell line is 39. The number of components to be decomposed R was set to 1000. Hence, the dimension of the input features the DNN model is 3000. The threshold we choose to binarize drug synergy score is 30.

The DNN we constructed as a part of the DTF model has 9 hidden layers, which hold 1024, 128, 128, 128, 128, 128, 1024, 1024 hidden units, respectively. Other hyper-parameters of the DNN are shown in Table 1. We have not employed normalization, since it actually masked some information of the features, which made the training of the DNN harder. We set other parameters as default values.

| Hyperparameter | Setting |
|----------------|---------|
| Activation function | ReLU(hidden layer)  | sigmoid(output) |
| Batch size | 2048 |
| Learning rate | 0.001 |
| Optimizer | RMSprop |
| Epochs | 10 |

3 Results

3.1 Model Comparison

We used CP-WOPT algorithm to decompose the 100 tensors from the 100 training sets generated from the stratified-sampling method. After the tensor factorization, we got features of drug A, drug B and cell line and also the tensor score of each drug pair. The factorized features from the CP-WOPT were used to build the classification models of DTF, RF, LR and CP-WOPT.
Using the trained models for each of the four classification methods (DTF, RF, LR and CP-WOPT), we predicted the synergy status of the drug pairs in the test set. We collected the prediction results of the 100 numerical experiments of the four models all together to draw the summarized ROC curves of the 100 experiments for these four models, respectively. The ROCs of the models are shown in Figure 5 while the AUCs of the models, mean, standard deviation (std) and 95 percent confidence interval (95% CI) of these AUCs are shown in Table 2. As we can see, the DTF model shows a higher AUC value than other three models while its std and 95% CI are similar to other three models. For all the models, the std and 95% CI of these AUCs are very small, suggesting the proposed DTF model can predict the synergy status of the drug pairs more accurately and robustly.

### Table 2. The classification performance of the 4 methods

| Method  | AUC mean | AUC std | 95% CI      |
|---------|----------|---------|-------------|
| DTF     | 0.92     | 0.0062  | [0.90, 0.93]|
| CP-WOPT | 0.88     | 0.0062  | [0.87, 0.89]|
| LR      | 0.84     | 0.0058  | [0.83, 0.85]|
| RF      | 0.87     | 0.0061  | [0.86, 0.88]|

### 3.2 Cell Line Specific- DTF Model Performance

The model comparison shows that the DTF model has a significant improvement on the original CP-WOPT and other two traditional classification models of RF and LR. Hence, we further explored the performance of the DTF model at individual cell lines. To do this, we calculated the AUCs and accuracies of the 100 experiments for each cell line based on the DTF model. It should be noted that while calculating the accuracy, we selected 0.5 as the threshold, suggesting that a positive synergy status will be assigned to a given drug pair if its prediction probability of the synergy score is higher than 0.5. We defined accuracy as the proportion of the correctly predicted drug pairs in the test sets of each cell line.

Finally, we averaged the AUCs and accuracies of the 100 experiments for each cell line, and the results are shown in Figure 6. It can be seen that all the cell lines have the performance of AUCs and accuracies higher than 0.85, and most of them are higher than 0.90, suggesting the DTF model can predict the synergy status of drug pairs at individual cell lines well. In other words, the model is not cell line-specific.

### 3.3 Prediction of the drug pairs with unknown synergy status

The results from Section 3.2 have shown that the DTF model we built is able to learn the relationship between the constructed tensor and the known synergy status of the drug pairs. The performance of the DTF model on the test set of each cell line also shows that the model does not overfit on the train set, which means that we can reliably apply the model to predict the synergy status of the drug pairs we do not have synergy scores experimentally measured.

To do this, we first constructed the tenor using all drug pairs with known synergy scores and then built the DTF-based prediction model based on the factorized features from the tensor. We used the model to predict synergy status of the drug pairs with unknown synergy scores. Using the same cutoff of 0.5 as we used for model evaluation, we obtained a total of 196 drug pairs with the probability of synergistic effect larger than 0.5 for the 27 cell lines as shown in Supplemental Table 1. We listed the drug pairs with the highest predicted probability for each of the 27 cell lines in Table 3. Approximately 74% of these drug pairs have a predicted synergistic probability score 0.8 or higher.

### 4 Discussion

We deeply examined the 196 drug pairs with a predicted probability of synergistic effect greater than 0.5 (Supplemental Table 1), and found that 17, 20, 30, 32, and 67 of these drug pairs show synergistic effect in cell lines from melanoma, colon cancer, breast cancer, prostate cancer, lung cancer, and ovarian cancer, respectively. A further look at these predicted drug pairs at individual cell line level found that the drug synergies were predicted for only 27 out of the 39 cell lines. The top three cell lines with the largest number of synergistic drug combinations identified were VCAP, CAOV3 and T47D. Specifically, 29, 23, and 15 drug pairs were predicted to be synergistic in VCAP (prostate cancer), CAOV3 (ovarian cancer), and T47D (breast cancer), respectively. Therefore, this study has...
Predicting anticancer drug synergy

Table 3. Top predicted synergistic drug pairs for 27 cell lines

| Cell line | Cancer       | Drug A       | Drug B       | Probability |
|-----------|--------------|--------------|--------------|-------------|
| A258      | Melanoma     | Vinblastine  | Carboplatin  | 0.5025      |
| A375      | Melanoma     | Paclitaxel   | Dexamethasone| 0.9106      |
| CAOV3     | Ovarian      | Etoposide    | Dexamethasone| 0.9998      |
| COLOR120DM| Colon        | Paclitaxel   | Dexamethasone| 0.6696      |
| EFM192B   | Breast       | Vinblastine  | SN-38       | 0.5758      |
| ES2       | Ovarian      | Paclitaxel   | SN-38       | 0.8806      |
| HT144     | Melanoma     | Paclitaxel   | SN-38       | 0.8067      |
| KPL1      | Breast       | Etoposide    | Dexamethasone| 0.9481      |
| LNCAP     | Prostate     | Etoposide    | Dexamethasone| 0.6444      |
| LOVO      | Colon        | Etoposide    | Dexamethasone| 0.8763      |
| MSTD      | Lung         | Etoposide    | Paclitaxel  | 0.9989      |
| NCIH1650  | Lung         | Etoposide    | Dexamethasone| 0.8423      |
| NCH2122   | Lung         | Etoposide    | Dexamethasone| 0.9990      |
| NCIH460   | Lung         | Paclitaxel   | SN-38       | 0.9036      |
| OCUBM     | Breast       | Carboplatin  | Dexamethasone| 0.7299      |
| OV90      | Ovarian      | Vinorelbine  | SN-38       | 0.8762      |
| OVCAR3    | Ovarian      | Etoposide    | Dexamethasone| 0.9753      |
| RKO       | Colon        | Vinblastine  | Carboplatin  | 0.7283      |
| SKMEL30   | Melanoma     | Etoposide    | Dexamethasone| 0.8412      |
| SKMES1    | Lung         | Paclitaxel   | Dexamethasone| 0.8891      |
| SKOV3     | Ovarian      | Etoposide    | Dexamethasone| 0.9283      |
| SW620     | Colon        | Vinblastine  | SN-38       | 0.9157      |
| SW837     | Colon        | Vinblastine  | Mitomycin   | 0.8606      |
| T47D      | Breast       | Metformin    | SN-38       | 0.8651      |
| UWB1298BRCA1| Ovarian   | SN-38        | Etoposide   | 0.7144      |
| VCAP      | Prostate     | Vinblastine  | Topotecan   | 0.9907      |
| ZR751     | Breast       | Paclitaxel   | Dexamethasone| 0.7949      |

Predicted drug synergies for majority of the 39 cell lines from each of the six cancer types with the largest number of synergistic drug combinations identified in ovarian cancer.

Interestingly, we found many of these predicted drug pairs have been reported to be synergistic in the literature. For example, the vinblastine/carboplatin combination shows a probability of 0.98 for synergistic effect in the prostate cell line VCAP. Vinblastine, as a type microtubule inhibitor, is typically used in combination with other agents to treat cancers, such as non-small cell lung cancer, bladder cancer, brain cancer, melanoma, and testicular cancer [16]. Carboplatin, as an analog of cisplatin, is an alkylating agent and has been used to treat a number of cancer types, including ovarian, lung, head and neck, endometrial, esophageal, bladder, breast, and cervical cancers [17]. The combination of carboplatin and vinblastine together with methotrexate has demonstrated the efficacy in the treatment of patients with advanced uterine/lung cancer “unfit” for cisplatin-based chemotherapy [18]. Thus, the carboplatin and vinblastine combination may show synergistic efficacy in treating prostate cancer. Another combination, etoposide and dexamethasone, shows the highest predicted probability (0.99) of synergy among the 23 pairs in the ovarian cancer cell line CAOV3. Etoposide is a DNA topoisomerase II inhibitor and has been approved by FDA for treating testicular and lung cancers [19]. Oral etoposide has demonstrated efficacy as an advanced treatment option for platinum-resistant ovarian cancer patients [20] and also as a maintenance chemotherapy for advanced ovarian cancer patients to improve the survival outcomes [21]. Dexamethasone is a steroid used to reduce inflammation. Dexamethasone is a type of steroid medication that has powerful anti-inflammatory and immunosuppressant and has been used in cancer treatment [22]. A recent study has confirmed the efficacy of the etoposide and dexamethasone combination therapy in hemophagocytic lymphohistiocytosis treatment [23]. Therefore, combining dexamethasone and etoposide may improve the efficacy of etoposide as a single agent for the ovarian cancer treatment.

In the future, we are interested in expanding the model to incorporate more information resources into the DTF model. More effort will be put into investigating the best structure of the DNN for the sake of improving the performance of the entire model.

5 Conclusion

There are two key steps for the proposed DTF model, 1) decomposing the tensor with missing entries constructed from the original drug synergy data to generate features of drugs and cell lines using the CP-WOPT algorithm; and 2) training the DNN model using the factorized features together with the observed labels (synergistic status of the drug pairs) to predict the synergistic effect of the drug pairs with unknown synergistic scores.

The DTF method, by linking the CP-WOPT and the DNN, used only a single data source but significantly improved the performance of CP-WOPT. In addition, the DTF model showed the best prediction performance among the four algorithms used in this study, suggesting its potential as a valuable tool for predicting and optimizing synergistic drug pairs in silico and thus guiding in vitro and in vivo discovery of rational combination therapies.

Funding

This research was supported in part by Canadian Breast Cancer Foundation, Natural Sciences and Engineering Research Council of Canada, Mitacs and University of Manitoba.

References

[1] J. Lehár, A. S. Knueger, W. Avery, A. M. Heilbut, L. M. Johansen, E. R. Price, R. J. Rickles, G. F. Short Iii, J. E. Stauntion, X. Jin et al., “Synergistic drug combinations tend to improve therapeutically relevant selectivity,” Nature biotechnology, 2009.
[2] A. A. Borisy, P. J. Elliott, N. W. Hurst, M. S. Lee, J. Lehár, E. R. Price, G. Serbedzija, G. R. Zimmermann, M. A. Foley, B. R. Stockwell et al., “Systematic discovery of multicomponent therapeutics,” Proceedings of the Natl. Acad. Sci., vol. 100, no. 13, 2003.
[3] Sidorov P, Naulaerts S, Ariey F, Serbedzija, G. R. Zimmermann, M. A. Foley, B. R. Stockwell et al., “Synergistic drug combinations using NCI-ALMANAC data. Front Chem 2019;7:504076. doi:10.3389/fchem.2019.00509.
[4] Zhang T, Zhang L, Payne PRO, Li F. Synergistic Drug Combination Prediction by Integrating Multi-omics Data in Deep Learning Models. ArXiv Prepr ArXiv181107054 2018.
[5] Acar, E., Kolda, T.G., Dunlavy, D.M., & Morup, M. Scalable tensor factorizations for incomplete data. Chemometr. Intell. Lab. Syst. 106, 41–56 (2011).
[6] Huiyu Chen and Jing Li, DrugCom: Synergistic Discovery of Drug Combinations using Tensor Decomposition, IEEE International Conference on Data Mining, 2018.
[7] O’Neil J, Benita Y, Feldman L, Chenard M, Roberts B, Liu Y, et al. An unbiased oncology compound screen to identify novel combination strategies. Mol Cancer Ther 2016;15:1155–62.
