6

https://doi.org/10.1038/s44386-024-00004-z

Predicting drug combination response surfaces

Check for updates

Riikka Huusari¹ 🖂, Tianduanyi Wang^{1,2}, Sandor Szedmak¹, Tero Aittokallio^{2,3,4} & Juho Rousu¹ 🖂

Prediction of drug combination responses is a research question of growing importance for cancer and other complex diseases. Current machine learning approaches generally consider predicting either drug combination synergy summaries or single combination dose-response values, which fail to appropriately model the continuous nature of the underlying dose-response combination surface and can lead to inconsistencies when a synergy score or a dose-response matrix is reconstructed from separate predictions. We propose a novel prediction method, comboKR, that directly predicts the continuous drug combination response surface for a drug combination. The method is based on a powerful input-output kernel regression technique and functional modelling of the response surface. ComboKR belongs to the family of functional output regression methods, where the prediction target is a function, in our case, a non-linear parametric surface. Our method thus avoids predicting discretized forms of the target as scalars, vectors or matrices, and therefore provides better interpolation and extrapolation along the surfaces. As an important part of our approach, we develop a novel normalisation between response surfaces that standardises the heterogeneous experimental designs used to measure the dose-responses, and thus allows training the method with data measured in different laboratories. Our experiments on two predictive scenarios and using two combination datasets highlight the suitability of the proposed approach especially in the traditionally challenging setting of predicting combination responses for new drugs not available in the training data.

Drug combinations are increasingly used for treatment of various diseases, especially blood cancers and solid tumours^{1–3}. In contrast to monotherapies, combination therapies offer the advantages in overcoming intrinsic and acquired resistance in cancer treatment, enhancing drug responses via synthetic lethality, and reducing unwanted side-effects by lowering the dose of individual drugs in the combination^{4,5}.

In pre-clinical stages, drug combinations are typically measured in cell lines using dose-response assays. High-throughput screening enables one to measure the responses of pairwise drug combinations at a few selected concentrations of the two drugs (e.g., 5×5 or 8×8 dose-response matrices)⁶. There are several efforts to conduct large-scale drug combination screens in various cancer types⁷⁻⁹, which have resulted either in fully or partially measured dose-response matrices.

The synergistic effects of drug combinations are often evaluated by summary synergy scores, calculated by the divergence between the measured drug combination responses and the expected non-interaction responses of the single drugs over the full matrix⁶. Multiple synergy models have been proposed to score such divergence based on different assumptions of the expected non-interaction response, for example, the highest single agent (HSA) model¹⁰, Bliss independence model¹¹ and Loewe additivity model¹². However, there are disagreements in terms of synergy when using different synergy models, due to large differences in drug concentrations and maximum response values across studies¹³.

Since drug combination synergy is evaluated based on multi-dose combination responses, often tested in multiple cancer cell lines with distinct oncogene addictions, large-scale screening of combination effects is required for the systematic discovery of new effective and selective combinations. However, to screen pairwise combinations among 100 drugs at 5 different concentrations in 10 cell lines would already require more than one million experimental tests. To speed up such a resource and time-

¹Department of Computer Science, Aalto University, P.O. Box 11000 (Otakaari 1B), FI-00076 Espoo, Finland. ²Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, FI-00290, Finland. ³Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, N-0310 Oslo, Norway. ⁴Oslo Centre for Biostatistics and Epidemiology (OCBE), Faculty of Medicine, University of Oslo, N-0317 Oslo, Norway. ©e-mail: riikka.huusari@aalto.fi; juho.rousu@aalto.fi

consuming combination discovery process, machine learning models are needed to narrow down the massive combinatorial search space^{14,15}.

A large proportion of current research focuses on the prediction of drug combination synergy rather than dose-response values^{15–19}. Some research works focus on the prediction of drug combination responses at selected concentrations²⁰⁻²³. A major advantage of predicting directly the dosecombination responses is that different synergy models can be applied to the predicted dose-response matrices in post-analysis, making the prediction task independent of a specific synergy metric. One such prediction method is the comboLTR²³, which predicts directly the scalar-valued dose-combination responses by applying latent tensor-based polynomial regression (LTR)²⁴. The drug combination dose-response values can be seen as discrete measurements sampled at different concentration values from a continuous dose-response surface where the response value is a function of drugs' concentrations. Both synergy score prediction and dose-response prediction can be seen as predictions based on the underlying surfaces: if the full continuous dose-response surface is known, the dose-response matrices can be sampled from the known surface at given concentrations, and the synergy scores can then be derived based on the sampled dose-response matrices. Thus, the direct prediction of a dose-response surface is a more general task than either predicting the single dose-response values or the synergy scores.

Recently, Rønneberg et al.²⁵ proposed an approach called PIICM, that considers predicting full response surfaces instead of individual dosecombination responses. They proposed a probabilistic prediction model based on Gaussian process regression where the covariance matrices for pairwise drug interactions are parameterised and learned. Notably, their approach can be interpreted as a matrix completion task on the collected response matrix, as the learning system is based on response data only, without using any additional features (cell line or drug features). This means, however, that the approach can not be expected to adapt to settings where a response surface in a test set would contain a drug not seen in training data. This more challenging new drug scenario is important in practical applications since one cannot assume that responses of all the drugs of interest would been already tested before either individually or in combination.

In this work, we propose a novel approach for predicting directly the full continuous drug combination dose-response surfaces with a kernelbased functional output prediction model, called comboKR. In contrast to the PIICM model²⁵, comboKR is based on an inductive learning approach, which predicts the drug combination response surface from input drug features that are easily available from drug databases. Such an inductive machine learning approach is based on the assumption that similar drugs have similar combination surfaces (i.e. if drugs d_1 and d_1^* are similar, so are the combination surfaces of the pairs (d_1, d_2) and (d_1^*, d_2)). Moreover, to overcome the practical issues arising from the heterogeneous experimental design often used in drug combination response measurements, we propose a novel normalisation scheme for comparing drug interaction surfaces. The main goal of the normalisation scheme is to align the dose-response surfaces to be centred around the area where the response changes rapidly as concentrations change. We demonstrate that with comboKR the massive chemical space can be exploited efficiently toward finding novel effective drug combinations beyond the given drug set with known measured responses. To summarise, our contributions are as follows:

- We propose an accurate approach to drug combination response prediction that predicts the full continuous drug combination response surfaces instead of individual dose-response or synergy score values.
- Our surface-valued regression approach takes advantage of a novel normalisation scheme between drug response surfaces that solves issues arising from the heterogeneous experimental designs between and within combination studies.
- Important for novel drug combination discovery, our proposed method can be applied to new drug settings without the need to re-train the model or experimentally measure each drug response beforehand: only functions of the monotherapy responses are required.
- In comparison to the baseline LTR method²⁶ and another surfacevalued prediction approach, PIICM²⁵ (applicable only within simple

npj Drug Discovery | (2025)2:2

predictive scenarios), we show that comboKR achieves superior results, especially in the more challenging predictive scenario, where testing is performed on drugs not available in the training stage.

Overview of ComboKR

We propose a new model, comboKR, to predict the drug combination response surface for a given drug pair in a cell line. In this inductive learning approach, the predictions are made based on the drug features, that can be collected from some database. Our method also uses a paremetrised function to model the drug combination surfaces—in our experiments, we chose to use the BRAID model²⁷. With this model, we additionally take advantage of the more abundant monotherapy response data and assume that for the drugs for which a combination surface is being predicted, we know the monotherapy response model.

In this section, we first briefly introduce our predictive framework, after which we discuss the BRAID model, and how to compute similarities between two surface functions.

Surface-valued regression

Our focus is on learning to predict the full, continuous drug interaction response surfaces $y \in \mathcal{Y}$ for the drug pairs $(d_1, d_2) \in \mathcal{X}$ in a given cell line a challenging structured output prediction problem. This is an especially difficult prediction task, since practically each output—a surface—in any such data set is sampled in part or fully from different sets of concentrations than the others, and therefore the dose-combination response matrices are not directly comparable. To solve the problem, we consider adapting an approach that has sometimes been referred to as generalised kernel dependency estimation (KDE)²⁸ or input–output kernel regression (IOKR)^{29,30}.

In a nutshell, our proposed model for surface-valued prediction relies on a vector- or function-valued kernel ridge regression (KRR) problem, obtained by mapping the surfaces to the reproducing kernel Hilbert space (RKHS) $\mathcal{H}_{\mathcal{Y}}$ associated with kernel $k_{\mathcal{Y}}: \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ (see Fig. 1c). Now, instead of solving directly the structured prediction problem $f: \mathcal{X} \to \mathcal{Y}$, the learning problem has been cast as vector-valued one to learn $g: \mathcal{X} \to \mathcal{H}_{\mathcal{Y}}$, after which the prediction to $\mathcal{H}_{\mathcal{Y}}$ is mapped back to \mathcal{Y} (preimage problem). Using the closed-form solution for the KRR and assuming a normalised output kernel $k_{\mathcal{Y}}$, the final optimisation problem to obtain predictions can be written as

$$f(x) = \operatorname*{arg\,max}_{y \in \mathcal{Y}} k_y(y, Y) (\mathbf{K}_x + \lambda \mathbf{I}_n)^{-1} k_x(X, x). \tag{1}$$

Here \mathbf{I}_n is $n \times n$ identity matrix, and λ is the regularisation parameter of the kernel ridge regression model. \mathbf{K}_x stands for the $n \times n$ kernel matrix collecting all values of kernel evaluations between pairs of drugs $k_x(x_i, x_j)$, i, j = 1, ..., n and each x_i stands for a (ordered) pair of drugs, $x_i = (d_1, d_2)_i$. In this work, we have chosen to use as k_x the Tanimoto similarity computed from the MACCS fingerprints; however our framework is general and adaptable to other input data representations with suitable kernel. The shorthand $k_y(y, Y)$ refers to the vector $[k_y(y, y_1), ..., k_y(y, y_n)]$ with $\{y_i\}_{i=1}^n = Y$ the outputs of the training set; $k_x(X, x)$ is defined analogously.

Problem (1) is ill-posed in general. When this is the case, it is most often solved by restricting the search for maximum value over a candidate set C_y instead of the full output space \mathcal{Y} . Every element in the set is tried out, and the one giving the maximum value for (1) is chosen as the final prediction.

BRAID surfaces

Our learning approach considers full dose-response surfaces as the outputs of the learning problem. Yet, the dose-response data is collected via factorial experimental design by measuring response values from varying dose concentrations, resulting in data sets where these surfaces are represented with matrices collecting the measurements. Starting from the famous Loewe additivity¹² and Bliss independence¹¹, various approaches to modelling drug interactions have been proposed^{27,31-35}. In this work, in order to obtain the

(a) Drug combination data and BRAID surfaces



(b) Surface normalisation scheme



(c) Surface-valued learning problem



Fig. 1 | **Overview to the proposed approach. a** The drug combination data and associated monotherapy responses are used to fit a parametric surface on each dose-response matrix. The BRAID surface model uses the Hill equations of the two drugs, as well as two interaction parameters to model different types of response surfaces. b Illustration of surface normalisation resulting in different similarities. The surfaces S1 and S2 are

computed with $\kappa = 0$ (neutral), and S2 and S4 with $\kappa = -1.5$ (extreme antagonism) and $\kappa = 25$ (extreme synergism), respectively. **c** Finally, a surface-valued prediction problem is formulated and solved with the output kernel learning-style approach, where the output data is mapped with the help of a suitable kernel $k_y : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ to RKHS $\mathcal{H}_{\mathcal{Y}}(\phi_y \text{ is the associated feature map: } k_y(y_1, y_2) = \langle \phi_y(y_1), \phi_y(y_2) \rangle_{\mathcal{H}_{\mathcal{Y}}}$).

continuous form of the surface, we use the bivariate response to additive interacting doses (BRAID) drug interaction model²⁷ that builds on the Hill equation^{36,37} and is motivated by the Loewe additivity principle. We fit this function to each of the drug pair combinations. The model is intuitive, as it uses the Hill equation parameters of the two drugs, in addition to two interaction parameters. More concretely, the function depends on the following parameters:

• Four response parameters: the baseline response in the absence of drugs, *R*₀; the maximal responses of drugs 1 and 2 as *R*₁ and *R*₂; and

optionally also the maximal combination response R_3 . The R_3 parameter is not present in the original BRAID model introduced in ref. 27, but it is present in the implementation³⁸.

- Hill equation slope parameters for both drugs: τ_1 and τ_2 .
- Half maximal effective concentration (EC₅₀) for the two drugs: EC₁ and EC₂.
- The interaction parameter $\kappa \in]-2, \infty]$: $\kappa < 0$, $\kappa = 0$, $\kappa > 0$ for antagonism, additivity or synergy, respectively (as illustrated in Fig. 1a).

Now, the BRAID function for drugs 1 and 2 applied in dose concentrations c_1 and c_2 is written as

$$BRAID(c_1, c_2) = R_0 + \frac{R_1 - R_0}{1 + \tilde{D}_{12}^{-\sqrt{\tau_1 \tau_2}}}$$

where

$$\begin{split} \tilde{D}_{12} &= \tilde{D}_1^{1/\sqrt{\tau_1 \tau_2}} + \tilde{D}_2^{1/\sqrt{\tau_1 \tau_2}} + \kappa \sqrt{\tilde{D}_1^{1/\sqrt{\tau_1 \tau_2}}} \tilde{D}_2^{1/\sqrt{\tau_1 \tau_2}} \\ \tilde{D}_1 &= \left(\frac{c_1}{EC_1}\right)^{\tau_1}, \\ \tilde{D}_2 &= \frac{\left(\frac{R_2 - R_0}{R_1 - R_0}\right) \left(\frac{c_2}{EC_2}\right)^{\tau_2}}{1 + \left(1 - \frac{R_2 - R_0}{R_1 - R_0}\right) \left(\frac{c_2}{EC_2}\right)^{\tau_2}}. \end{split}$$

We note that in this formula it is assumed that one of the drugs is the weaker one, and the other stronger. In the practical implementation³⁹ either of the drugs could be the stronger one, resulting to a slightly more complicated formula. The implementation provides methods to fit the BRAID functions to data.

Due to the scarcity of the combination data (for example only 3×3 measurements for each combination in the NCI-ALMANAC dataset), to aid with the fitting process we additionally include the more abundant monotherapy data available from other sources if possible. We illustrate this in Fig. 1a.

The final optimisation problem of our model (Eq. (1)) is solved with the candidate set optimisation. Here, domain knowledge can be exploited in making the most suitable candidates: we make the natural assumption that the monotherapy equations for the two drugs are known. Compared to drug combination response data, monotherapy data is easier to collect, and many databases already exist for those measurements. With the monotherapy equations known, generating the candidate sets with the BRAID surface model boils down to generating suitable parameter combinations for the interaction parameters R_3 and κ .

Kernels between surfaces

Our surface-valued prediction approach relies on having a kernel, k_j , defined between the drug interaction surfaces. We choose to use the Gaussian (or RBF) kernel, defined as $k_{\text{RBF}}(z, z') = e^{-\gamma ||z-z'||^2}$ for some vectors z and z'. At first glance, it might seem attractive to use the (vectorised) dose-response matrices directly available in drug combination datasets in this kernel. However, even if those matrices were of the same size, most often the drug doses used to measure drug combination responses are widely different between any two surfaces. Thus, these measurements cannot be directly compared to each other. We instead assume that we have functions $S_i : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, mapping two dose concentrations to a response value, parameterising the drug interaction surfaces available—in our work we obtain these from the BRAID interaction model. We can now fix a set of concentrations for all drugs, $C = [[c_1, c'_1], [c_1, c'_2], \ldots, [c_N, c'_N]]$, and compare any two surfaces in for example RBF kernel with

$$k_{\gamma}(S_A, S_B) = \exp(-\gamma || S_A(C) - S_B(C) ||_{\rm F}^2)$$

where $S_A : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $S_B : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are response surface functions (e.g. BRAID). Both of them are defined for a pair of drugs on a cell line and are here queried with some concentration values, returning the response values at those dose combinations. Our shorthand notation ($S_1(C)$ and $S_1(C)$) evaluates these functions at all concentration pairs in the grid defined by *C*, which, after properly reshaping, gives two $N \times N$ matrices in the Frobenius norm evaluation in the kernel.

However, this straightforward approach still has limitations. Different drugs often have different effective concentrations, meaning that two combination surfaces expressing very similar interaction profiles (e.g. similar levels of synergy) might be shifted in relation to each other so that the kernel evaluation results in non-intuitive values. To overcome this issue, we propose normalisation over the dose-concentration values to an effective concentration range [0, 1] using the Hill equation, where 0 corresponds to concentrations that have no effect on the cell growth, 0.5 indicates EC_{50} concentration, and 1 corresponds to concentrations with a maximal response. We denote kernel acting on normalised surfaces as \tilde{k}_{γ} .

With this normalisation, problems related to shifted surfaces are decreased, and the kernel \tilde{k}_y shows more realistically the differences in interaction pattern (i.e. if the combination is synergistic or antagonistic). We illustrate this difference between the two choices of kernels in Fig. 1b, where surfaces of three types of interactions (S2, S3 and S4) are compared to baseline surface S1. The three surfaces differ from S1 by one of the Hill functions (h_3 instead of h_1) and the interaction parameter κ . Intuitively, the surfaces S1 and S2 are closest together in the sense that their drug combination interaction effect is very similar ($\kappa = 0$). However, the Hill equations and especially their EC₅₀ values for the first drug are different (equation h_1 vs. h_3), which has resulted in the two surfaces being "shifted" w.r.t. each other. The Frobenius norms between the original surfaces are, in all three cases, large, and S1 is judged most similar to S3. The normalised comparison can take the shifting into account, and S2 is therefore judged to be most similar to S1.

Results

In this section, we computationally validate our proposed surface-valued kernel regression model for drug combination response prediction, comboKR. We consider two drug combination datasets: the NCI-ALMANAC⁷, and O'Neil⁴⁰. We compare our comboKR to a scalar-valued dose-response prediction baseline, LTR²⁶, and to another surface-valued prediction model, PIICM²⁵. For datasets where PIICM does not scale well, we compare to a model derived from it—we call this PIICM^{*} (see the "Methods" section for details). Due to the computational cost of comboKR arising from the use of kernel matrices, we consider all the cell lines as independent data sets, and train and test on them separately—a challenging setting that often arises in real-world personalised medicine applications. For comboKR rew^{*} comboKR raw^{*} or "comboKR r." the version with the original, simpler kernel, while "comboKR" denotes version with our normalisation scheme.

The results presented here are obtained from two representative predictive scenarios: new drug and new combo. New combo refers to the case when the test set consists of new drug combinations—however, all the drugs are available in the training set as parts of other combinations. New drug is more realistic, but also more challenging scenario, where the drug combinations in the test set always contain a drug that has not been present in any of the combinations in the training set.

Overview of the results

We present the overall Pearson correlations averaged over the cell lines between the predicted values and ground truth measurements, as well as between the Bliss and Loewe synergies calculated from them (Fig. 2). For the response values, we also display density plots in Fig. 3 where include additionally HSA baseline¹⁰. In the easier new combo scenario on the NCI-ALMANAC dataset, the scalar-valued prediction approach, LTR, slightly outperforms all the surface-valued ones. With O'Neil data the PIICM method slightly outperforms others. Yet, the differences between the methods are relatively small. However, in the more challenging new drug scenario, the situation reverses, and the surface-valued approaches mostly outperform LTR. In this more challenging setting, comboKR has a clear advantage over the other models, especially when using the concentration normalised surface kernels. PIICM inherently requires all the drugs (and cell lines) in the test set and also in training, and thus cannot be applied in this setting. With the NCI-ALMANAC dataset, all the predictive models outperform the HSA baseline in the easier scenario, while comboKR continues to outperform it also in the more difficult one. With the smaller O'Neil dataset, all the results are closer to the baseline.

It can be seen from Fig. 3 that the surface-valued models can offer improvements over the scalar-valued model in predicting the extreme New combo scenario



diction settings (NCI-ALMANAC: 60 cell lines; O'Neil: 39 cell lines). The vertical lines in the plots highlight the mean and the extrema. First row of plots shows a correlation between original and predicted combination responses. Additionally, synergy scores (Bliss and Loewe) were calculated from both ground truth

measurements as well as for the predictions. The two other rows show correlations of these synergy scores. Statistical *p*-values from two-sided Kolmogorov–Smirnov test is shown for all the competing methods compared to ComboKR (full pairwise results are available in Supplementary material).

New drug scenario

response values that are more rare in the data. The LTR model focuses on the predictions of the more abundant, higher response values, and easily overshoots the predictions of the lower response values. While this is easier to see in the new drug predictive scenario between comboKR and LTR, the same behaviour is already present in the new combo setting. With NCI-ALMANAC data, the other surface-valued method, PIICM^{*}, predicts the extreme response values better than LTR but not as well as the comboKR approaches. However, with O'Neil data—with which PIICM was originally evaluated with²⁵—PIICM outperforms the other methods.

In supplementary material, we show additional results for the correlations with the fitted BRAID surfaces, as well as full pairwise results on the statistical significance of the differences between results displayed in Fig. 2.

Performance over tissue and drug combination types

Both NCI-ALMANAC's and O'Neil dataset's cell lines originate from multiple tissue types: nine types in NCI-ALMANAC, six in O'Neil. Similarly, the tested drugs belong to three drug groups (chemotherapy, targeted and other). Details of these groupings can be found from the supplementary material.

We investigated and compared in more detail the performance of the methods on these different tissue types and different drug type combinations in the two predictive settings (Figs. 4 and 5). As before, the LTR method performs slightly better than the comboKR approaches in the easier predictive scenario with the NCI-ALMANAC dataset, while PIICM outperforms others in this setting with the O'Neil dataset. Yet, looking at drug-type combination results, it can be seen that the PIICM has less advantage with chemotherapy drugs, with which comboKR performs similarly to PIICM. In the new drug scenario, comboKR clearly outperforms comboKR r. and LTR, while with the NCI-ALMANAC dataset also comboKR variants

is larger in the challenging new drug setting, also in the easier new combo setting on the O'Neil dataset, comboKR already outperforms the simpler comboKR r. approach in some cell lines and drug type combinations.

Performance of predicting individual surfaces

The results so far have focused on the overall performance of the models. Here, we show in Table 1 the results of the pairwise comparison of individual surface predictions. The table reports the average percentages (averaged over the cell lines) of how often a method achieved a better prediction for the surfaces in the test set, w.r.t. mean squared error, compared to the other methods (the comparison is always pairwise).

Similarly to the previous results, also here it can be observed that the differences between the methods are small in the easier new combo predictive scenario (Table 1a and c). The new drug predictive scenario (Table 1b and d) highlights again the benefits of the surface-valued methods: LTR obtains most often the worst prediction for a surface. The results again highlight how using the normalised surface kernels outperforms the basic comboKR r. approach. Notably the amount of identical predictions between the two comboKR versions is almost half the amount of that in the new combo setting.

An advantage to our method is that comboKR predicts a full continuous interaction surface, from which one can sample any doseconcentration pairs to obtain dose-specific predictions or summary synergy scores. Contrary to this, the baseline LTR accurately predicts only those concentrations measured in the drug response assay separately, which can lead to predictions that do not follow a smooth interaction pattern. We illustrate this in Supplementary material.

Discussion

In this work, we have investigated the drug combination response prediction problem from the point of view of predicting entire drug combination



(d) New drug predictive scenario, O'Neil dataset.

Fig. 3 | Density plots of predictions and ground truth measurements over all cell lines for the two predictive scenarios. The titles of the plots indicate the Pearson correlation, Spearman correlation and root mean squared error (RMSE) between the predicted and the measured response values. The panels **a** and **c** show results for NCI-ALMANAC dataset, while **b** and **d** show them for O'Neil. Panels **a** and **b** show results of the new combo scenario, and **c** and **d** show results in the more challenging new drug scenario.

Article



(a) New combo experimental scenario, tissue types, NCI-ALMANAC dataset



(b) New combo experimental scenario, tissue types, O'Neil dataset



(c) New drug experimental scenario, tissue types, NCI-ALMANAC dataset



(d) New drug experimental scenario, tissue types, O'Neil dataset

Fig. 4 | Distributions of Pearson correlations of the drug-dose response prediction on the different tissues (see supplementary material for details) in the two predictive scenarios, for the comboKR variants and LTR, as well as for PIICM variants in the new combo scenario. The vertical lines in the plots highlight the mean. The violins with fewer than ten elements indicate the number of elements in parenthesis of the horizontal axis label and additionally include also the individual results, marked with crosses. The panels **a** and **c** show results for NCI-ALMANAC dataset, while **b** and **d** show them for O'Neil. Panels **a** and **b** display results of the new combo scenario, and **c** and **d** display the results in the more challenging new drug scenario.

surfaces, instead of predicting individual response values. We propose an approach based on kernel methods, which when combined with a novel surface normalisation scheme, overcomes issues arising from the heterogeneous experimental design used to measure the data. We show that casting the drug combination response prediction as a structured prediction learning problem can improve predictive performance, especially in traditionally challenging experimental settings. Namely, our method shows great promise especially in the new drug scenario, providing the opportunity to find promising drug combinations that go beyond the limited set of drugs in the training set.

To explore the suitability of our proposed surface-valued learning approach, we performed computational experiments on the NCI-ALMANAC and O'Neil datasets. Of the two predictive scenarios investigated in our experiments, the proposed surface-based approach achieves better predictive performance, especially in the more challenging new drug scenario. Even the more straightforward surface-valued prediction Fig. 5 | Distributions of Pearson correlations of the drug-dose response prediction on the different drug type combinations (see supplementary material for details) over the 60 (NCI-ALMA-NAC) or 39 (O'Neil) cell lines in the two predictive scenarios. The vertical lines in the plots highlight the mean. For O'Neil data, there is no "Other-Other" drug-type combination. The panels **a** and **c** show results for NCI-ALMANAC dataset, while **b** and **d** show them for O'Neil. Panels **a** and **b** display results of the new combo scenario, and **c** and **d** display the results in the more challenging new drug scenario.



(a) New combo experimental scenario, drugtype combinations, NCI-ALMANAC dataset.



(b) New combo experimental scenario, drugtype combinations, O'Neil dataset.



(c) New drug experimental scenario, drugtype combinations, NCI-ALMANAC dataset.



(d) New drug experimental scenario, drugtype combinations, O'Neil dataset.

approach outperformed the baseline LTR method on one dataset, but especially the novel concentration normalisation provides significant improvements in this challenging, yet practical setting. In personalised medicine studies, either focusing on individual cell lines or patient samples, one cannot assume that each drug has already been tested in combination with other drugs in often limited training datasets. In addition to outperforming other approaches, the suitability of our proposed comboKR in the new drug scenario is also computationally practical: unlike with the traditional methods, the proposed model does not need to be re-trained to obtain predictions when a new set of drugs is introduced to the test set; predicting only requires the monotherapy response function for the (new) drugs. Moreover, the comboKR predicts a full continuous drug interaction surface, instead of individual values that might not conform to a smooth interaction pattern. Again, this gives a practical advantage: it is easier, based on the predicted surface to experimentally validate the synergy between two drugs by using the surface to determine relevant drug concentrations.

Table 1 | Pairwise comparison of the methods

(a) NCI-ALMANAC, new combo scenario. *comboKR versions resulted in ties at rate 14.5±2.6%.

%	c.KR	c.KR r.	c.LTR	PIICM^*
c.KR		$43.1 \pm 4.0^{*}$	46.4 ± 5.1	52.9 ± 4.6
c.KR r.	$42.4 \pm 3.1^*$		44.0 ± 4.4	51.3 ± 4.2
c.LTR	53.6 ± 5.1	56.0 ± 4.4		55.4 ± 6.0
PIICM^*	47.1 ± 4.6	48.7 ± 4.2	$44.6 {\pm} 6.0$	

(c) O'Neil dataset, new combo scenario. *comboKR versions resulted in ties at rate $13.2\pm4.2\%$.

%	c.KR	c.KR r.	c.LTR	PIICM
c.KR		$50.8 {\pm} 7.3^{*}$	64.0 ± 8.2	60.2 ± 5.5
c.KR r.	$36.0 \pm 7.5^{*}$		$56.6 {\pm} 6.9$	$55.9 {\pm} 4.6$
c.LTR	36.0 ± 8.2	$43.4{\pm}6.9$		52.6 ± 5.3
PIICM	$39.8 {\pm} 5.5$	$44.1 {\pm} 4.6$	$47.4 {\pm} 5.3$	

(b) NCI-ALMANAC, new drug scenario. *comboKR versions resulted in ties at rate 7.3±3.2%.

%	c.KR	c.KR r.	c.LTR
c.KR		$56.2 \pm 5.7^{*}$	$75.8 {\pm} 6.1$
c.KR r.	$36.6 \pm 4.3^*$		67.7 ± 6.5
c.LTR	24.2 ± 6.1	32.3 ± 6.5	

(d) O'Neil dataset, new drug scenario. *comboKR versions resulted in ties at rate $7.3\pm5.6\%$.

%	c.KR	c.KR r.	c.LTR
c.KR		$78.4{\pm}12.0^{*}$	$80.3 {\pm} 9.7$
c.KR r.	$14.2 \pm 9.6^{*}$		41.0 ± 15.7
c.LTR	$19.7 {\pm} 9.7$	59.0 ± 15.7	

The table reports average percentages over all cell lines. How often did the method on row give a better prediction for a surface, w.r.t. mean squared error, than the method on column; rows with high values (darker colouring; similarly columns with low values with lighter colouring) indicate better performance for the method. The two comboKR versions with original and normalised surface kernels use the same candidate sets, and so their predictions might be identical, resulting in ties in the rankings. Tie counts are not included in the tables, so cells marked with asterisk do not add to 100%.

$1000 E_1$ The union of continues and anage in the databate contracted in the work
--

Dataset	#Cell lines	#Drugs	#Surfaces	Measurements
NCI-ALMANAC	60	104	311 527	3 × 3; one set
O'Neil et al	39	38	22 527	4 × 4; four sets

The table reports also the amount and sizes of the drug combination matrices measured in the data.

Surface sampling can also be used to suggest doses for experimental testing with highest likelihood for revealing synergy between two drugs.

Additionally, we observed that the surface-valued methods, in general, were better suited for predicting extreme response values (Fig. 3). This was most clearly observed in the new drugs setting. The extreme responses are often most informative for identifying synergistic (or antagonistic) interactions between two drugs, so their prediction is critically important for drug combination discovery. Our comparison with the HSA baseline showcases the overall difficulty of drug combination response prediction: even such a simple baseline can perform relatively well, most likely due to the presence of extreme combination responses that can be captured by most models. Indeed, HSA is equal to the concept of independent drug action (IDA) that can explain many of the clinically beneficial drug combinations, even without pre-clinical synergy³. Such a simple baseline can thus provide a useful first guess for modelling novel drug interaction patterns.

Our experiments give promising results for using a structured prediction approach in the drug combination response prediction, motivating future research. It will be important to investigate ways to make the ComboKR model more scalable on big high-throughput screening data consisting of millions of data points. A more scalable model would allow more variety in the predictive scenarios: if trained over multiple cell lines, a fully new cell line could be included in the predictive stage. The obvious bottleneck for applications to multiple cell lines is the sampling of the large kernel matrix on inputs, consisting of two Kronecker products: $K_c \otimes K_d \otimes K_d$. In addition to efficient algorithms and parallelisation strategies, for example, kernel approximations could be investigated to speed up the computations. Another avenue to pursue would be to follow⁴¹, and investigate the scalability using or generalising the proposed Kronecker product vec-trick. Finally, it would be interesting to investigate and compare different drug combination surface models. In current work, we have used the BRAID surfaces, but our method could be applied to any model.

Methods

Drug combination datasets

In this study, we consider two drug combination datasets: the NCI-ALMANAC⁷, and O'Neil⁴⁰—see Table 2 for a summary of them.

The O'Neil dataset provides a screen of 38 unique drugs in pairwise combinations on 39 diverse cancer cell lines (see Supplementary material). The combination data consists of 4×4 drug combination measurements on the cell lines; in total there are 22,527 of such drug–drug-cell combinations. (We note that the full data size is 22,727⁴⁰, but 22,527 is the amount of the surfaces with measurements on 4×4 grid. The other 200 are measured at non-conforming grids.) In this work, the median response values over the four measurement replicats are considered to be the groundtruth responses for the experiments. In addition to the combination data, the dataset contains monotherapy measurements for the drugs on the cell lines, typically on six measurement replicats.

The NCI-ALMANAC dataset provides systematic screening of drug combinations among 104 FDA-approved anticancer drugs on the 60 NCI-60 human tumour cell lines covering 9 different tissue types (see Supplementary material). In this dataset, drugs have been screened at either 5 or 3 concentrations, resulting in 5×3 or 3×3 drug combination dose-response matrices. As the number of larger dose-response matrices is much fewer than 3 × 3 matrices, in order to have consistent dose–response matrix size in method evaluation, the large dose-response matrices were subsampled to 3×3 by keeping the entries corresponding to the largest dose concentrations. Monotherapy responses at combination dose concentrations were also included in the NCI-ALMANAC dataset for the 104 drugs, where single-drug responses were measured at various concentrations. Different concentration values were determined for different drugs⁷. Thus, each drug combination response surface is represented by the measured responses by a 4 × 4 matrix containing both combo- and mono-responses. When duplicate measurements of the same drug combination on the same cell line appeared, the median of measurements was taken.

Fig. 6 | The drug dose combination counts in NCI-ALMANAC dataset. The two axes refer to the two drugs in combination, for which the corresponding dose concentration counts are displayed in the bar plots. The counts of the drug dose combinations are displayed in the matrix, where a darker colour indicates a higher count.



NCI-ALMANAC data has been collected with a standard factorial experimental design. The 3-by-3 (or 4×4), dose–response matrices described above are independent of each other in a sense, that the dose levels between experiments (matrices) used to collect the data are not the same. Notably, even if two response matrices both consider the same drug as one factor, the concentrations used in measurements might differ between them. To illustrate this, Fig. 6 shows that here is a large amount of different concentration combinations in the NCI-ALMANAC data. On average, any given dose combination is found from only 0.44% of the drug response matrices. The most common concentration combination is present at 12.4% of the matrices (the full distribution is shown in Supplementary Material). It is very rare that for any two surfaces, all nine concentration combinations would match. Thus, directly comparing any two dose-response matrices in the dataset can very rarely be done.

NCI-ALMANAC reported two endpoints calculated differently by using time zero measurement as reference or not⁴². For the percent growth of cells with time zero as reference ("PercentGrowth" as reported in the dataset), the responses range from –100 to 100. However, for this endpoint, the calculation processes are different when the percent growth of test cells is lower or higher than the time zero measurement of cell growth. Whereas for the percent growth of cells using only control values ("Percent-GrowthNoTZ" as reported in the dataset) as a reference, the response calculation is consistent and ranges from 0 to 100. Thus, for simplicity and consistency of data, the endpoint denoted as "PercentGrowthNoTZ" was used as the response value. If there were missing values for the endpoint, it was calculated from the raw measurements reported in the dataset. For O'Neil dataset, we used directly the viability measurements as the responses that were provided in the data⁴⁰. The values mostly range in [0, 1], sometimes higher values than 1 are present.

Our proposed comboKR, as well as the compared PIICM method, rely on accurate surface modelling in the training stage. Since the NCI-ALMANAC drug combination response data only contains three monotherapy measurements for both drugs, the NCI-60 single drug monotherapy response data where typically a drug is measured at multiple concentrations was also integrated as part of the surface model fit procedure, in order to help improve the model performance with a better estimate of monotherapy dose-response functions. We remark that the O'Neil dataset directly includes additional monotherapy viability measurements in addition to the combination data. To train the machine learning models to predict the dose-response values, we consider the commonly used 166-bit 2D structure MACCS molecular fingerprints as the input data.

The ComboKR model

Our proposed ComboKR model is an adaptation of an approach that has sometimes been referred to as generalised kernel dependency estimation (KDE)²⁸ or input–output kernel regression (IOKR)^{29,30}.

In this approach, operator-valued kernels (OvKs) are used to solve a regression problem to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_{\mathcal{Y}}$ induced by a traditional scalar-valued kernel $k_{\mathcal{Y}}$ defined for the output data in \mathcal{Y} . OvKs are associated with vector-valued reproducing kernel Hilbert spaces (vv-RKHSs), containing functions that map the data to the vector-valued (or function-valued) output space. Thus, they are a natural choice to solve for the function $g : \mathcal{X} \to \mathcal{H}_{\mathcal{Y}}$.

In this context, the operator-valued kernel is a function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{L}(\mathcal{H}_{\mathcal{Y}})$ in which $\mathcal{L}(\mathcal{H}_{\mathcal{Y}})$ denotes the set of linear operators from $\mathcal{H}_{\mathcal{Y}}$ to $\mathcal{H}_{\mathcal{Y}}$ —if the output space $\mathcal{H}_{\mathcal{Y}}$ is finite-dimensional, i.e. $\mathcal{H}_{\mathcal{Y}} = \mathbb{R}^{p}$, then $\mathcal{L}(\mathcal{H}_{\mathcal{Y}}) = \mathbb{R}^{p \times p}$. In practise, the most common operator-valued kernel to use is the separable (or decomposable) kernel, which can be written as $\mathcal{K}(x, z) = k_x(x, z)\mathbf{T}$ with $k_x : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ being a traditional scalar-valued kernel on the input data, and $\mathbf{T} \in \mathcal{L}(\mathcal{H}_{\mathcal{Y}})$, which in turn is often chosen to be the identity.

Operator-valued kernels generalise the usual scalar-valued kernels, notably also the representer theorem. Thus, the solution to the regularised learning problem considered in the IOKR,

$$g(x) = \min_{f \in \mathcal{H}_{X}} \sum_{i=1}^{n} \left\| y_{i} - g(x_{i}) \right\|^{2} + \lambda \| g \|_{\mathcal{H}_{X}}^{2}$$

where \mathcal{H}_X denotes the vv-RKHS associated with \mathcal{K} , can be written as

$$g(x) = \sum_{i=1}^{n} K(x, x_i) \mathbf{c}_i$$

Here $c_i \in \mathcal{H}_{\mathcal{Y}}$ are the multipliers to be learned. Like in the usual scalar-valued case, the closed-form solution

$$\operatorname{vec}(\mathbf{C}) = (\mathbf{G} + \lambda \mathbf{I})^{-1}\operatorname{vec}(\Phi(Y))$$

can be obtained, in which $\Phi(Y) = [\phi_v(y_1), \phi_v(y_2), \dots, \phi_v(y_n)]$ and $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3]$..., \mathbf{c}_n], both of size $d \times n$ if the size of feature space associated with kernel k_v is denoted with d. G is the $nd \times nd$ operator-valued kernel matrix with block-wise structure. Now,

$$g(x) = \mathcal{K}(x, X) \operatorname{vec}(\mathbf{C}) = \mathcal{K}(x, X) (\mathbf{G} + \lambda \mathbf{I})^{-1} \operatorname{vec}(\Phi(Y)),$$

where $\mathcal{K}(x, X) = [\mathcal{K}(x, x_1), \dots, \mathcal{K}(x, x_n)]$

After solving for g, the final predictions can be obtained from the preimage problem

$$\underset{y \in \mathcal{Y}}{\operatorname{arg\,min}} \parallel \phi_{y}(y) - g(x) \parallel^{2}$$

or from

$$\underset{y \in \mathcal{Y}}{\arg \max} \left\langle \phi_{y}(y), g(x) \right\rangle$$

if one assumes that the output kernel k_{ν} is normalised, i.e. $k_{y}(y, y) = 1 \ \forall y \in \mathcal{Y}$. It is now possible to take advantage of the form of the separable kernel matrix, and the Kronecker product vec-trick, to obtain the final optimisation problem

$$f(x) = \operatorname*{arg\,max}_{y \in \mathcal{Y}} k_y(y, Y) (\mathbf{K}_x + \lambda \mathbf{I}_n)^{-1} k_x(X, x). \tag{2}$$

Here I_n is $n \times n$ identity matrix, and K_x stands for the $n \times n$ kernel matrix $[k_x(x_i, x_j)]_{i,j=1}^n$. The shorthand $k_y(y, Y)$ with $Y = \{y_i\}_{i=1}^n$ the training set outputs, refers to the vector $[k_v(y, y_1), \dots, k_v(y, y_n)]; k_x(X, x)$ is defined analogously.

Algorithm 1. The comboKR approach

Require $\mathbf{K}_{x}^{\text{train}} \in \mathbb{R}^{n_{\text{train}} \times n_{\text{train}}}$ the kernel matrix on drug pairs in training set

- 1: $\mathbf{K}_{x}^{\text{test}} \in \mathbb{R}^{n_{\text{train}} \times n_{\text{test}}}$ the kernel matrix on drug pairs between training and test sets
- 2: λ regularisation parameter for KRR
- 3: $[S_i]_{i=1}^{n_{\text{train}}}$ list of fitted (BRAID) surface functions for training data
- 4: $[h_1]_{i=1}^{n_{\text{test}}}$ and $[h_2]_{i=1}^{n_{\text{test}}}$, lists of monotherapy (Hill) functions for the drugs in test data
- 5: $[C_1]_{i=1}^{n_{\text{test}}}$, $[C_2]_{i=1}^{n_{\text{test}}}$ lists of concentration arrays on which predictions should be made

6: $\mathbf{M} \leftarrow (\mathbf{K}_{x}^{\text{train}} + \lambda \mathbf{I})^{-1} \triangleright \text{Training}$

7: $\mathbf{Z} \leftarrow \mathbf{MK}_{r}^{\text{test}}$

- 8: if use normalised output kernel then
- 9: $S \leftarrow$ sample training surfaces on normalised grid

10: end if

11: $T \in []$ \triangleright T collects predictions at queried concentrations 12: for $t \in [1, ..., n_{\text{test}}]$ do \triangleright Loop for predicting

- $Q \leftarrow$ candidate surfaces with $h_1[t]$ and $h_2[t]$ as monotherapies 13:
- 14: if use normalised output kernel then
- 15: $\mathbf{Q} \leftarrow$ sample candidate surfaces Q on normalised grid defined on $[0, 1] \times [0, 1]$
- else if use the original output kernel then 16:
- $S \leftarrow \text{sample training set surfaces on test concentration grid}$ 17: based on $C_1[t]$ and $C_2[t]$
- $\mathbf{Q} \leftarrow$ sample candidate surfaces Q on test concentration grid 18: based on $C_1[t]$ and $C_2[t]$
- 19: end if
- $\mathbf{K}_{y}^{\text{test}} = [\exp(-\gamma \parallel \mathbf{S}_{i} \mathbf{Q}_{j} \parallel_{\mathbf{F}}^{2})]_{i,j}; i \in [1, n_{\text{train}}], j \in [1, n_{\text{candidates}}]$ $i \leftarrow \arg \max [\mathbf{K}_{y}^{\text{test}}]^{\top} \mathbf{Z}[:, t] \triangleright \mathbf{Z}[:, t] \text{ denotes th column of } \mathbf{Z}$ 20:
- 21:
- [1,n_{candidates}] if use normalised output kernel then 22:

- 23: add Q_i sampled at concentration grid based on $C_1[t]$ and $C_2[t]$ to T
- 24: else if use the original output kernel then
- 25: add \mathbf{Q}_i to T

end if 26.

- 27: end for
- 28: return T

The pseudocode of the approach can be found in Algorithm 1. Our approach assumes fitted monotherapy functions (Hill equations in our experiments) for the drugs in the test set, and also fitted surface functions for the training data (BRAID surfaces in our experiments); computing these can be seen as a preprocessing step. We note that for the new combo scenario, all these Hill equations are equations of the drugs in the training set, but in the new drug scenario, the Hill equations also contain equations for previously unseen drugs.

Training of our model is done based on the closed-form solution and consists of computing $(\mathbf{K}_{x} + \lambda \mathbf{I}_{y})^{-1} k_{x}(X, x)$ from Eq. (2) (lines 6 and 7). For predicting, we consider candidate set optimisation. Here, we take advantage of the domain knowledge and the assumption that the monotherapy functions for the drugs in the test set are known. Now, the candidates can be created by directly using the parameters available in the Hill equations and by generating various combinations of the other required parameters; for the BRAID function the parameters R_3 and κ . Finally, one of the candidate surfaces is chosen as the final prediction following Eq. (2) (lines 20 and 21). It is good to note that in order to obtain predictions for a drug outside of the training set, the model requires no re-training, but only the Hill equation of the drug in question.

In this work, for our experiments, we choose k_x to be defined as $k_x((d_1, d_2), (d'_1, d'_2)) = k_d(d_1, d'_1)k_d(d_2, d'_2)$ with k_d acting on the MACCS fingerprints to be Tanimoto kernel⁴³. As discussed, for the output surfaces, we consider the RBF kernel, which conforms to the assumption of having the output kernel being normalised. We use the parameter $\gamma = \frac{1}{2\sigma^2}$ with σ equal to the mean of distances between the surfaces in the training set. Results comparing different output kernels can be found from the supplementary material.

Remark 1. For simplicity, we have chosen to use the pairwise kernel defined as $k_x((d_1, d_2), (d'_1, d'_2)) = k_d(d_1, d'_1)k_d(d_2, d'_2)$. It is not invariant to the order of the drugs, and thus, in practice results in the requirement of "doubling" the data in training by including both orders.

Formulations for invariant kernels exist (see e.g. ref. 25), but we observed negligible differences in observed performance between the kernels (Table 3).

Concentration normalisation to obtain a normalised surface kernel. The main idea of our proposed surface normalisation scheme is to standardise the concentration measurements across the different drugs so that all the concentrations are in the same range, and the surface comparisons can be made more easily. To this end, we map the dose concentrations with the help of Hill equations, individually for each drug in each cell line, to the range [0, 1], where 0 intuitively stands for "no effect", and 1 stands for "maximal effect" (see Fig. 7). More concretely, the concentration normalisation transformation CN can be written as

$$CN(c|R_0, R_{\max}, \tau, EC_{50}) = \frac{h(c|R_0, R_{\max}, \tau, EC_{50}) - R_0}{R_{\max} - R_0}$$

Table 3 | Minimum, average and maximum differences of Pearson correlations over all the cell lines on the O'Neil data results, when a non-invariant input kernel was used along with the data doubling, or an invariant kernel was considered

	New combo	New drug
Minimum difference	0.0	4.2e-05
Average difference	0.000668	0.001973
Maximum difference	0.004697	0.008049



Fig. 7 | Illustration of the concentration normalisation procedure to [0, 1] range. The concentrations too low to elicit a response are mapped to zero, EC_{50} concentration is mapped to 0.5, and concentrations close to maximal effect are mapped close to 1.

in which *c* is the original concentration, and *h* is the Hill equation with parameters R_0 , R_{max} , τ and EC₅₀ (i.e. baseline response, maximal response, slope parameter and half maximal effective concentration). The normalisation formula uses the same R_0 and R_{max} as the Hill equation.

After the concentration normalisation, the values [0, 1] can be seen to correspond to each other over all the different drugs—unlike in nonnormalised cases where a concentration might yield a very different response from two different drugs. With all drug concentrations normalised, comparing different surfaces is easy in a common grid of values from $[0, 1] \times [0, 1]$ (even if for different drugs and cell lines they map to different original concentrations). Due to the nature of the normalisation scheme, it can be expected that this common drug grid in all surfaces will focus especially on areas where the potential synergy or antagonism can best be captured (as illustrated in the second part of Fig. 1).

Experimental setup

We compare our comboKR (available at https://github.com/aalto-icskepaco/comboKR) to two baselines: LTR^{26,14} for predicting individual response values, and PIICM^{25,45} as another surface-valued prediction method based on Gaussian processes. With our comboKR, we consider both the original and normalised surfaces in output kernel computations: the former case is denoted by comboKR r.

LTR is a polynomial regression model for scalar-valued prediction that exploits higher order interactions between the views in predictions. As input for a prediction, it takes the two drugs and their concentrations. As comboKR considers kernels on input data, for LTR we use the empirical features⁴⁶ from Tanimoto kernel evaluated on the MACCS fingerprints. As for the dose-concentrations, the LTR method represents them as one-hot encoded vectors.

PIICM is a surface-valued approach based on Gaussian process regression, that uses latent GP model from bayesynergy R package⁴⁷ to model the drug response surfaces.²⁵ conducted the experiments with the O'Neil dataset also considered in our work. Thus, with the O'Neil dataset we use their training-test split in the new combo scenario, with their original surface modelling and data normalisation scheme, both of which differ from ours. Contrary to²⁵, to adhere within our setting, we train separate models for each cell line. With NCI-ALMANAC dataset we instead use the same BRAID surfaces as the baseline models as in our comboKR. Like in ref. 25, we consider a subset of concentrations to represent the surfaces since the method is computationally too heavy to run with a full set of concentrations in the data (over 60 unique dose values), and the BRAID surfaces are sampled at these concentration normalisation also resulted in a [0, 1] interval, when using the BRAID surface model, we also consider our novel normalisation scheme.

Yet, we observed that the memory requirements for the PIICM method to run with NCI-ALMANAC dataset in our setting were infeasible, as the number of drugs considered is relatively large. In order to run the method, we forced it to consider a simplified form of the drug combination covariance matrix—we call this PIICM^{*}. While this might put the method at a disadvantage, in practice, we observe that in the new combo setting its performance is close to LTR, which is close to the results obtained in ref. 25. Moreover, we observed in small-scale experiments that our modification performs comparably to, or even slightly improves, the original parametrisation (see details in Supplementary Material). The PIICM^{*} outputs predictions on the chosen concentration grid. To obtain the final PIICM^{*} predictions, we interpolate with Nadaraya–Watson kernel regression from this set of concentrations to the concentrations at which the test surfaces are measured at.

The parameters for the models are chosen with cross-validation, taking into account the specifics of the predictive scenario in each split. For our comboKR, λ is chosen from 1e-2, 5e-2, 1e-1, 5e-1, while in LTR, the order is considered 3 or 5, and rank 10 or 20. PIICM^{*} and PIICM drug rank is cross-validated over 5, 25, 50, 75 and 100. In comboKR, the normalised grid is a 11 × 11 grid, based on normalised concentrations [0, 0.1, 0.2, ..., 0.9, 0.999]. In all models, the training data is "doubled", i.e. both orders of drug pairings are included separately in the training set.

Predictive scenarios. We consider predictions in two scenarios:

- New drug: one of the drugs in the combination queried at the test stage has not been seen in any combination during the training stage. Monotherapy responses of the novel drug are assumed to be available during the training stage.
- *New combo*: the drug combination queried at the test stage has not been seen in training; however, both single drugs may have been seen in the combinations encountered during training.

In the new combo setting, the data in each cell line is divided evenly into five folds, of which one is used as test fold, and the remaining folds are used in training and validation (with O'Neil data we use the same split as in ref. 25). In the new drug setting, the test fold contains all surfaces where either drug is one of ten (NCI-ALMANAC) or four (O'Neil) randomly chosen as the new drug. The four validation folds are obtained similarly among the rest of the data used in the training, always with randomly chosen drugs. Clearly, the new combo setting is the easiet of the two prediction scenarios. We note that in both scenarios, no drug-dose combination response values of a surface queried at the test point are available in training. The task is to predict all the combination responses, instead of filling missing values inside a matrix with known entries.

In the more challenging new drugs scenario, the drug features and monotherapy responses for the novel drug are assumed to be available already in the training stage. For most of the standard models, including the LTR baseline, it means that at any time a new drug is to be included in testing, the model would need to be re-trained with the additional information. This can be very costly. An advantage of our proposed comboKR method is that this kind of re-training is not needed, and the new monotherapy response data can be introduced to the model during the test stage when the candidate surfaces are being built. In this scenario, the relevant monotherapy responses of the new drugs in the test set are included to augment the LTR training set.

We note that the PIICM method²⁵ is based on Gaussian processes, modelling the drug interaction surfaces and without considering any drug (or cell line) features when making predictions. Thus, it cannot generalise to data outside of the training set in the new drug scenario and cannot be applied there.

Obtaining the drug interaction surfaces. As a preprocessing step for surface-based approaches (PIICM^{*} and our proposed comboKR), we fit the BRAID drug interaction surfaces to the available combo data with the synergy package³⁸. Due to the scarcity of the monotherapy responses in the NCI-ALMANAC combination data, the relevant monotherapy

Count





response values in O'Neil are in range [0, 1], for better comparison to NCI-ALMANAC, they have been scaled to [0, 100] here. For RMSE in O'Neil, the first value is for scaled responses, the second value in parenthesis is for the original data.



Fig. 9 | **Examples of BRAID surfaces not fit well to the NCI-ALMANAC dataset.** On top: cell line HL-60(TB), drugs pentostatin (id 218321) and streptozotocin (id 85998). Middle: cell line NCI-H226, drugs nilotinib (id 747599) and floxuridine (id 27640). Bottom: cell line OVCAR-8, drugs capecitabine (id 712807) and bend-amustine (id 138783). The two first plots in each row show the NCI-60 monotherapy

responses in black, with blue crosses highlighting the NCI-ALMANAC monotherapy response values, while the curve displays the responses from the Hill equation from the BRAID model. Third column displays the full combo data matrix, and the final column shows the fitted BRAID surface.

responses available in NCI-60 were used to augment the combo data. Hill equations fitted to the monotherapy data are used in building the normalised output kernels. Overall, the average Pearson correlation of the fitted BRAID surfaces sampled at the measurement positions to the ground truth combination measurements was 0.9037 ± 0.0169 averaged

over the cell lines, 0.905 over all of them in NCI-ALMANAC dataset, and 0.980 ± 0.015 averaged over the cell lines, and 0.982 over all of them in O'Neil dataset (see Fig. 8 for density plots).

It can be expected that large amount of of the discrepancies are due to surface model filtering out noise from the measurements. However, some errors in fitting to the NCI-ALMANAC dataset are due to the conflicts between it and and NCI-60 dataset. Examples of this are provided in Fig. 9, where monotherapy data follows a clear trend and the monotherapy entries in combo data are outliers.

Candidate surfaces for comboKR. The simplest way to solve the preimage problem (i.e. finding the best *y* for prediction g(x) in $\mathcal{H}_{\mathcal{Y}}$) for the comboKR problem (2), is to consider a candidate set where all elements are tested out, and the one giving the highest score is selected as the prediction. In both of the predictive scenarios, we assume that the monotherapies of the drugs are available. Thus, we can easily create relevant surface candidates with the surface model by only varying the parameters κ and E_{max} that are related to the behaviour of the drug combination. In experiments, the candidates are generated by sampling κ from [-1.5, -1, -0.5, -0.1, 0.01, 0.1, 0.5, 1, 2, 10] (to capture various surface interaction profiles), and E_{max} from around the maximum values of the individual drug responses.

Performance metrics. The test sets in both predictive settings consist of drug-dose interaction surfaces that have been sampled at various differing 4×4 grids. The predictions at those concentrations were obtained as follows: in LTR they are directly predicted, in PIICM they are interpolated from the predictions in the normalised grid, and in our comboKR they are sampled from the predicted BRAID surface. Two synergy scores (Bliss and Loewe) are calculated based on these ground truth and predicted 4×4 matrices.

We report Pearson correlations of the predictions (both for raw responses and summary synergy scores) calculated for each cell line separately. Within a cell line, both the predictions and ground truth labels are vectorised, and the correlation is calculated between the two vectors. The same procedure is used when investigating different tissue types and drug combination types: all the ground truth and predicted responses within the group in a cell line are vectorised to compute the Pearson correlation.

Furthermore, we compare all the methods to each other on the level of individual surfaces, by reporting how often in the test set each of the methods obtained the best prediction for a given surface with respect to the mean squared error—Pearson correlation not being applicable if the matrix of values that is sampled from the predicted surface is constant. This happens most often in modified PIICM, but also sometimes in both comboKR versions.

Data availability

The NCI-60 and NCI-ALMANAC datasets⁷ used are publicly available at https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-60+Growth+Inhibition +Dataand https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC. The O'Neil dataset is available as supplementary material associated with ref. 40.

Code availability

The code for the comboKR method is available at https://github.com/aaltoics-kepaco/comboKR. For the competing methods, we used codes made available for LTR at https://github.com/aalto-ics-kepaco/GO_LTR and adapted PIICM from https://github.com/ltronneb/PIICM/. The software for the BRAID surfaces is available from https://github.com/djwooten/synergy/.

Received: 21 August 2024; Accepted: 16 December 2024; Published online: 03 February 2025

References

- Das, P., Delost, M. D., Qureshi, M. H., Smith, D. T. & Njardarson, J. T. A survey of the structures of us FDA approved combination drugs. *J. Med. Chem.* 62, 4265–4311 (2019).
- Pomeroy, A. E., Schmidt, E. V., Sorger, P. K. & Palmer, A. C. Drug independence and the curability of cancer by combination chemotherapy. *Trends Cancer* 8, 915–929 (2022).

- Plana, D., Palmer, A. C. & Sorger, P. K. Independent drug action in combination therapy: implications for precision oncology. *Cancer Discov.* 12, 606–624 (2022).
- Lopez, J. S. & Banerji, U. Combine and conquer: challenges for targeted therapy combinations in early phase trials. *Nat. Rev. Clin. Oncol.* 14, 57–66 (2017).
- Jin, H., Wang, L. & Bernards, R. Rational combinations of targeted cancer therapies: background, advances and challenges. *Nat. Rev. Drug Discov.* 22, 213–234 (2023).
- Pemovska, T., Bigenzahn, J. W. & Superti-Furga, G. Recent advances in combinatorial drug screening and synergy scoring. *Curr. Opin. Pharmacol.* 42, 102–110 (2018).
- Holbeck, S. L. et al. The National Cancer Institute Almanac: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity NCI Almanac of approved cancer drug combinations. *Cancer Res.* **77**, 3564–3576 (2017).
- Jaaks, P. et al. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature* 603, 166–173 (2022).
- 9. Nair, N. U. et al. A landscape of response to drug combinations in nonsmall cell lung cancer. *Nat. Commun.* **14**, 3830 (2023).
- 10. Berenbaum, M. C. What is synergy? Pharmacol. Rev. 41, 93-141 (1989).
- 11. Bliss, C. I. The toxicity of poisons applied jointly 1. *Ann. Appl. Biol.* **26**, 585–615 (1939).
- Loewe, S. The problem of synergism and antagonism of combined drugs. Arzneim.-Forsch. 3, 285–290 (1953).
- Vlot, A. H., Aniceto, N., Menden, M. P., Ulrich-Merzenich, G. & Bender, A. Applying synergy metrics to combination screening data: agreements, disagreements and pitfalls. *Drug Discov. Today* 24, 2286–2298 (2019).
- Weinstein, Z. B., Bender, A. & Cokol, M. Prediction of synergistic drug combinations. *Curr. Opin. Syst. Biol.* 4, 24–28 (2017).
- Wu, L. et al. Machine learning methods, databases and tools for drug combination prediction. *Brief. Bioinform.* 23, bbab355 (2021).
- Abbasi, F. & Rousu, J. New methods for drug synergy prediction: a mini-review. *Curr. Opin. Struct. Biol.* 86, 102827 (2024).
- Kong, W. et al. Systematic review of computational methods for drug combination prediction. *Comput. Struct. Biotechnol. J.* 20, 2807–2814 (2022).
- Torkamannia, A., Omidi, Y. & Ferdousi, R. A review of machine learning approaches for drug synergy prediction in cancer. *Brief Bioinform.* 23, bbac075 (2022).
- Liu, H. et al. The recent progress of deep-learning-based in silico prediction of drug combination. *Drug Discov. Today* 28, 103625 (2023).
- 20. Ianevski, A. et al. Prediction of drug combination effects with a minimal set of experiments. *Nat. Mach. Intell.* **1**, 568–577 (2019).
- Julkunen, H. et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nat. Commun.* 11, 6136 (2020).
- Ling, A. & Huang, R. S. Computationally predicting clinical drug combination efficacy with cancer cell line screens and independent drug action. *Nat. Commun.* **11**, 5848 (2020).
- 23. Wang, T. et al. Modeling drug combination effects via latent tensor reconstruction. *Bioinformatics* **37**, i93–i101 (2021).
- Szedmak, S., Cichonska, A., Julkunen, H., Pahikkala, T. & Rousu, J. 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 (2020).
- Rønneberg, L., Kirk, P. D. & Zucknick, M. Dose-response prediction for in-vitro drug combination datasets: a probabilistic approach. *BMC Bioinform.* 24, 161 (2023).
- Armah-Sekum, R. E., Szedmak, S. & Rousu, J. Protein function prediction through multi-view multi-label latent tensor reconstruction. *BMC Bioinform.* 25, 174 (2024).

- Twarog, N. R., Stewart, E., Hammill, C. V. & Shelat, A. A. Braid: a unifying paradigm for the analysis of combined drug action. *Sci. Rep.* 6, 25523 (2016).
- Kadri, H., Ghavamzadeh, M. & Preux, P. A generalized kernel approach to structured output learning. In *Proceedings of the 30th International Conference on Machine Learning, ICML,* Vol. 28 (eds. Dasgupta, S. & McAllester, D.) 471–479 (PMLR, 2013).
- Brouard, C., Szafranski, M. & d'Alché Buc, F. Input output kernel regression: supervised and semi-supervised structured output prediction with operator-valued kernels. *JMLR* **17**, np (2016).
- Brouard, C. et al. Fast metabolite identification with input output kernel regression. *Bioinformatics* 32, i28–i36 (2016).
- Kong, M. & Lee, J. J. A generalized response surface model with varying relative potency for assessing drug interaction. *Biometrics* 62, 986–995 (2006).
- Lee, J. J., Kong, M., Ayers, G. D. & Lotan, R. Interaction index and different methods for determining drug interaction in combination therapy. *J. Biopharm. Stat.* **17**, 461–480 (2007).
- Whitehead, A., Whitehead, J., Todd, S., Zhou, Y. & Smith, M. K. Fitting models for the joint action of two drugs using sas (r). *Pharm. Stat.: J. Appl. Stat. Pharm. Ind.* 7, 272–284 (2008).
- 34. Zhao, W. et al. A new bliss independence model to analyze drug combination data. *J. Biomol. Screen.* **19**, 817–821 (2014).
- Wooten, D. J., Meyer, C. T., Lubbock, A. L., Quaranta, V. & Lopez, C. F. Musyc is a consensus framework that unifies multi-drug synergy metrics for combinatorial drug discovery. *Nat. Commun.* 12, 4607 (2021).
- Hill, A. V. The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves. *J. Physiol.* 40, iv–vii (1910).
- Goutelle, S. et al. The hill equation: a review of its capabilities in pharmacological modelling. *Fundam. Clin. Pharmacol.* 22, 633–648 (2008).
- Wooten, D. J. & Albert, R. Synergy: a Python library for calculating, analyzing and visualizing drug combination synergy. *Bioinformatics* 37, 1473–1474 (2021).
- 39. Synergy. Version 0.5.1 https://github.com/djwooten/synergy (2024).
- O'Neil, J. et al. An unbiased oncology compound screen to identify novel combination strategies. *Mol. Cancer Ther.* 15, 1155–1162 (2016).
- Airola, A. & Pahikkala, T. Fast Kronecker product kernel methods via generalized vec trick. *IEEE Trans. Neural Netw. Learn. Syst.* 29, 3374–3387 (2017).
- Nci-60 Methodology (accessed 1 Mar 2022); https://dtp.cancer.gov/ discovery_development/nci-60/methodology.htm.
- Ralaivola, L., Swamidass, S. J., Saigo, H. & Baldi, P. Graph kernels for chemical informatics. *Neural Netw.* 18, 1093–1110 (2005).
- Go-ltr. https://github.com/aalto-ics-kepaco/GO_LTR/tree/main/src/ go_ltr. commit d097a9e674401aa2bf3a70bdbf523717069c0036 (2024).
- Piicm. https://github.com/ltronneb/PIICM. commit 4f560595b95ef0f29ad9da53697e9e9e301dba81 (2022).
- Scholkopf, B. et al. Input space versus feature space in kernel-based methods. *IEEE Trans. Neural Netw.* **10**, 1000–1017 (1999).

 Rønneberg, L., Cremaschi, A., Hanes, R., Enserink, J. M. & Zucknick, M. Bayesynergy: flexible Bayesian modelling of synergistic interaction effects in in vitro drug combination experiments. *Brief. Bioinform.* 22, bbab251 (2021).

Acknowledgements

This work was supported by the Academy of Finland through grants 334790, 339421, 345802 (MAGITICS, MASF and AIB, respectively), as well as the Global Programme by the Finnish Ministry of Education and Culture. T.A. was supported by grants from the European Union's Horizon 2020 Research and Innovation Programme (ERA PerMed CLL-CLUE project), the Academy of Finland (grants 340141, 344698, and 345803), Norwegian Health Authority South-East (grant 2020026), the Cancer Society of Finland, and the Sigrid Jusélius Foundation. The authors acknowledge the computational resources provided by the Aalto Science-IT project.

Author contributions

R.H. and J.R. ideated the approach. T.W. handled data preprocessing, R.H. implemented the approach and performed the experiments and analysis with the help of S.S. and T.W. All authors wrote and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44386-024-00004-z.

Correspondence and requests for materials should be addressed to Riikka Huusari or Juho Rousu.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025