MARSY: A multitask deep learning framework for prediction of drug combination synergy scores

Mohamed Reda El Khili[1], Safyan Aman Memon[1], and Amin Emad[1,2,3,*]

[1] Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada

[2] Mila, Quebec AI Institute, Montréal, QC, Canada

[3] The Rosalind and Morris Goodman Cancer Institute, Montréal, QC, Canada

* Corresponding Author:

Amin Emad

755 McConnell Engineering Building

3480 University Street, Montreal, Quebec, Canada, H3A 0E9

Email: amin.emad@mcgill.ca

Abstract

Combination therapies have emerged as a treatment strategy for cancers to reduce the probability of drug resistance and to improve outcome. Large databases curating the results of many drug screening studies on preclinical cancer cell lines have been developed, capturing the synergistic and antagonistic effects of combination of drugs in different cell lines. However, due to the high cost of drug screening experiments and the sheer size of possible drug combinations, these databases are quite sparse. This necessitates the development of transductive computational models to accurately impute these missing values. Here, we developed MARSY, a deep learning multi-task model that incorporates information on gene expression profile of cancer cell lines, as well as the differential expression signature induced by each drug to predict drug-pair synergy scores. By utilizing two encoders to capture the interplay between the drug-pairs, as well as the drug-pairs and cell lines, and by adding auxiliary tasks in the predictor, MARSY learns latent embeddings that improve the prediction performance compared to state-of-the-art and traditional machine learning models. Using MARSY, we then predicted the synergy scores of 133,722 new drug-pair cell line combinations, which we have made available to the community as part of this study. Moreover, we validated various insights obtained from these novel predictions using independent studies, confirming the ability of MARSY in making accurate novel predictions.

33 **Introduction**

34 Cancers are complex diseases that involve various pathways and are regulated by a multitude
35 of different genes [1,2]. Despite the emerging understanding of cancers, the development of
36 effective treatments remains a prevailing challenge. Combination therapies, in which multiple
37 treatments are administered simultaneously, have emerged as an alternative to
38 monotherapies [2-4]. By simultaneously targeting multiple genes, proteins, and pathways, they
39 can reduce the probability of drug resistance while improving treatment's efficacy [3,5-7], and
40 allow for the reduction of necessary dosage per drug, reducing the risk of drug toxicity and
41 adverse effects [5,7,8].

42

43 Due to the distinct molecular and clinical characteristics of cancer types, it is necessary to
44 evaluate the response of cancer cells to different treatments and treatment strategies in each
45 cancer type. The curation of molecular profiles of cancer cell lines (CCLs) and their response to
46 monotherapies in large databases such as the Cancer Cell Line Encyclopedia (CCLE) [9] and
47 Genomics of Drug Sensitivity in Cancer (GDSC) [10] initiated the development of various
48 computational models for prediction of single drug response in CCLs [11,12] and patient tumours
49 [13,14]. More recently, large databases of synergy scores of drug combinations (mainly drug-pairs)
50 in CCLs such as DrugComb [15] have been curated based on results of many high-throughput drug
51 screening studies. These databases include different synergy scores to quantify the observed
52 effectiveness of a drug combination compared to its expected effectiveness if they are not
53 interacting [16]. In spite of these efforts, the sheer size of possible (drug combination, CCL) tuples
54 and the cost of drug screening experiments have resulted in sparse datasets with many missing

values. Computational models that can accurately predict the synergy scores of drug combinations and impute these datasets can have a significant impact in this domain.

Various models have been proposed to achieve this goal. In one study [17], the authors used various machine learning (ML) models and obtained their best predictive performance using Random Forests (RF) and Extreme Gradient Boosting (XGB) models. Other studies have also explored ML models such as linear regression, Support Vector Machine (SVM), LASSO, RF, XGB, Extremely Randomized Trees, and TreeCombo (which is an extreme gradient boosted tree-based approach) [18-20]. In addition to these traditional ML models, several deep learning (DL) models have been proposed. One of the earliest methods, DeepSynergy [21], used a fully connected neural network composed of two hidden layers and showed significant improvement compared to then state-of-the-art ML models such as RF and XGB. A more recent model, MatchMaker [8], showed that learning two distinct representations for each drug-CCL pair in a combination can improve the performance.

Here we present MARSY, a multitask DL model that predicts the level of synergism between drug-pairs tested on CCLs. Using gene expression to characterize CCLs and drug-induced signatures to represent each drug, MARSY learns a distinct set of embeddings to obtain multiple views of the input features. Precisely, a representation of the entire combination and a representation of the drug-pair are learned in parallel. These embeddings are then fed to a multitask network that predicts the synergy score of the drug combination alongside single drug responses. A thorough evaluation of MARSY revealed its superior performance compared

77 to various state-of-the-art and traditional computational methods. A detailed analysis of the
78 design choices of our framework demonstrated the predictive contribution of the learned
79 embeddings by this model. Using MARSY, we then predicted the synergy scores of 133,722 new
80 drug-pair CCL combinations, which can be used to guide future drug screening and
81 pharmacogenomics studies. Moreover, we validated various insights obtained from these novel
82 predictions using independent studies, confirming the ability of MARSY in making accurate
83 novel predictions.
84

85 **METHODS**

86 **Prediction of drug-pair synergy scores and single drug response using MARSY**

87 MARSY (Multitask drug pAiR Synergy) is a DL-based model that seeks to learn latent
88 representations (embeddings) capable of predicting the synergy score of two drugs in a CCL. As
89 input, it receives feature vectors of two candidate drugs and a feature vector of a CCL (Figure
90 1). The architecture includes two parallel and separately parameterized encoders with a
91 bottleneck layer, and one multi-task predictor. The first encoder (ENC_{Pair}) receives the
92 concatenation of the feature vectors of the drug-pair to learn a (drug 1, drug2)-specific
93 embedding, while the second encoder (ENC_{Triple}) receives all three feature vectors to learn a
94 (drug1, drug2, CCL)-specific embedding. These embeddings are then concatenated and
95 provided as input to the predictor, which predicts the synergy score of the two drugs, along
96 with the single drug response of each drug in that CCL (performing three tasks simultaneously).
97 The inclusion of single drug response predictors as auxiliary tasks ensures that the
98 representations learned by the two encoders are constrained to capture biologically and

99 chemically important information corresponding to each drug, improving its generalizability
100 and performance.
101
102 ENC_{Pair} accepts a drug-pair's signature (a vector of length 3912) as input and consists of two
103 fully connected hidden layers of width 1024 and 2048. ENC_{Triple} accepts a vector of length 8551
104 corresponding to the concatenation of a drug-pair's signatures and a CCL's gene expression
105 profile as input and consists of two fully connected hidden layers of width 2048 and 4096. The
106 first hidden layer of both encoders is a bottleneck layer to perform dimensionality reduction,
107 while the second hidden layer has a larger width to increase the capacity of the model. The
108 relative width of layers in these two encoders are chosen to approximately match the length of
109 inputs to each encoder. A linear activation function is used for the first layer, while a Rectified
110 Linear Unit (ReLU) function is used for the second layer of both encoders. Both encoders use
111 dropout regularization with a probability of 0.2. The embeddings obtained from these two
112 encoders are concatenated and provided as input to the predictor ($PRED_{Resp}$). $PRED_{Resp}$ consists
113 of two fully connected hidden layers of width 4096 and 1024 and an output layer of width 3 to
114 predict the synergy score of the drugs and their individual drug response. A ReLU activation
115 function is used for the first two layers while a linear activation function is selected for the
116 output layer, given the regression nature of the prediction task. Finally, a dropout
117 regularization with a probability of 0.5 is used for the predictor. Details of hyperparameter
118 tuning and training are discussed below.

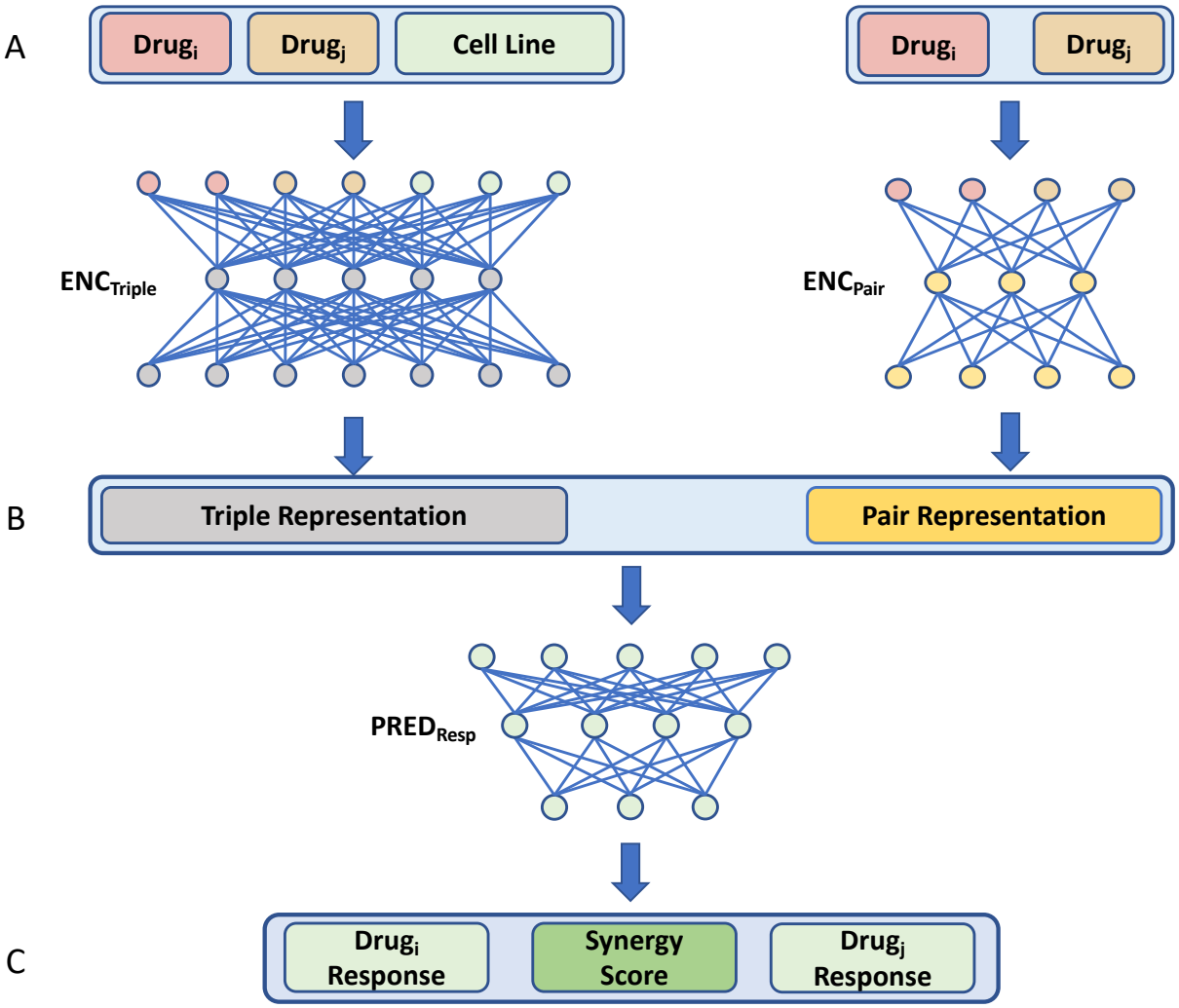

Figure 1: The overview of MARSY's architecture. A) The drug features and CCL gene expression profile are provided as input to the two encoders. Separately parameterized encoders obtain representations corresponding to the drug-pair and to the drugs and CCL triple. B) Both representations are concatenated and used as inputs to a multi-task predictor. C) The predictor predicts the synergy score of the drug combination, along with the single drug response of each drug (three outputs).

Training and hyperparameter tuning

Both encoders (ENC_{Triple} and ENC_{Pair}) and the predictor ($PRED_{Resp}$) were trained in an end-to-end fashion using a mean squared error (MSE) loss function and the Adamax optimizer [22] with the default learning rate of 0.001 and early stopping. The dropout probabilities and the activation

131 function used on the input layer of both encoders were considered hyperparameters to be
132 tuned using an independent validation set. All possible combinations of the probabilities 0.2
133 and 0.5 were assessed for both the encoders and the predictor. Furthermore, this assessment
134 was performed alternating between a ReLU and a linear activation function for the input layer
135 of the encoders. The reason we considered a linear activation function as an option for the first
136 layer is its ability to improve and simplify the optimization process. These hyperparameters
137 were evaluated on a small validation set (~1100 samples) that was excluded from the main
138 dataset used for cross-validation evaluation, in order to ensure that no data leakage occurs.
139 The rest of MARSY's architecture followed the design choices detailed above. The best
140 performing combination of hyperparameters (as evaluated on the independent validation set)
141 was used to finalize the design of the MARSY framework.
142

143 **Datasets and data cleaning**

144 In this study, the feature vector of each CCL corresponded to its baseline gene expression
145 profile (before administration of any drug), as measured in CCLE [9]. We downloaded the
146 $log_2(RPKM + 1)$ normalized RNA-seq profile of 1019 untreated CCLs (catalogued in the CCLE
147 study) from the CellMiner database [23,24]. We removed lowly expressed genes whose
148 $log_2(RPKM + 1)$ value was lower than 1 in more than 75% of all CCLs. Additionally, we
149 calculated the variance of genes across all CCLs and excluded those that had a variance smaller
150 than 0.8, which resulted in a set of 4639 features (Figure 2). Finally, a z-score normalization was
151 performed to the expression of each gene over all the CCLs, one gene at a time.
152

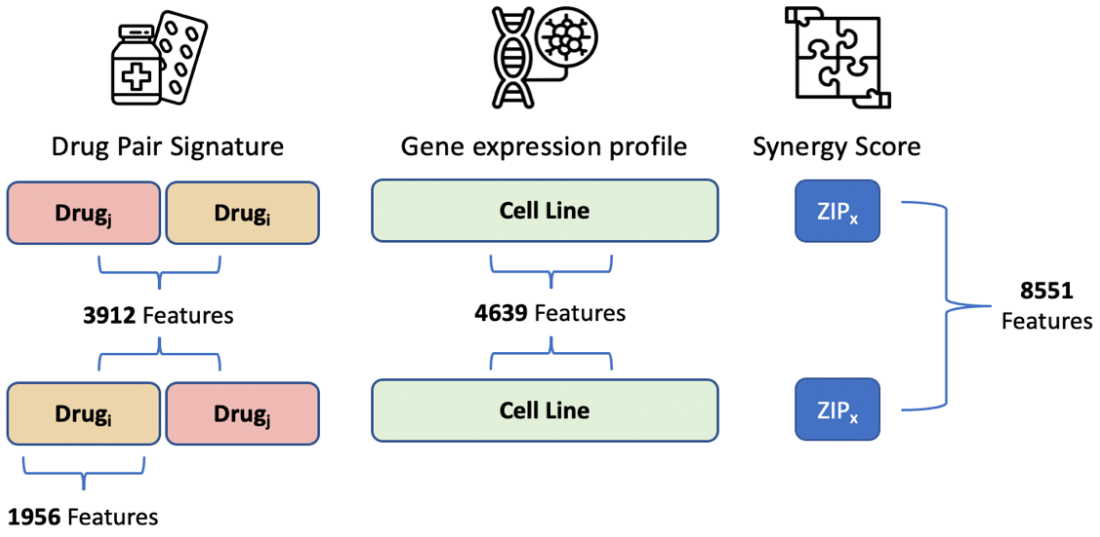

Figure 2: Overview of the number of features representing each element of the input. To ensure the embeddings learned by MARSY are invariant to order of the drug-pairs, for each such pair, an additional sample is included by inverting their order while keeping the same associated synergy score.

To represent drugs (as inputs to the model) we used their signatures corresponding to the differential gene expression profile of two CCLs (MCF7 and PC3) measured 24 hours after treatment with that drug with the dosage of $10\ \mu M$. We downloaded level 5 signatures from the Library of Integrated Network-Based Cellular Signatures (LINCS) [25], and only used experimentally measured values corresponding to 978 landmark genes. The two CCLs, the time after treatment, and the drug dosage were selected to maximize the number of samples in our study. In instances where replicate signatures existed for the same CCL, same drug, same dosage, and same time-point, we calculated their mean to obtain a consensus signature. We concatenated the MCF7 and PC3 signatures of each drug, resulting in a vector of length 3912 to represent each drug-pair (Figure 2).

170 We obtained drug synergy scores (and single drug responses) from the DrugComb database
171 (v1.5) [15]. DrugComb is one of the largest publicly available datasets for drug synergy scores with
172 a total of 739,964 drug combinations obtained from 8397 unique drugs. We downloaded data
173 corresponding to triples of (drug1, drug2, CCL) for which we had both drug signatures (from
174 LINCS) and CCL gene expression profiles (from CCLE). In this work, we focused on two measures
175 of synergism, Zero Interaction Potency ("ZIP") and "S_{mean}". ZIP calculates the synergism of a
176 drug-pair under the assumption that there is no potency between the single drug responses [26].
177 On the other hand, S_{mean} calculates synergism of a drug-pair assuming that the single drug effect
178 is equivalent to the average of the single drug responses [27]. As opposed to ZIP, S_{mean} is not
179 calculated for each combination of dosages; instead, a drug sensitivity measure is computed
180 directly from the dose response curve of each drug and used as the drug-pair's response. Thus,
181 S_{mean} serves not only as a measure of drug-pair synergy but also of its sensitivity, which aims to
182 illustrate the actual response of a CCL [27], and hence is more biologically informative.
183
184 Since DrugComb contains (sometimes inconsistent) replicates, it is important to carefully
185 process such cases. To identify samples with inconsistent replicates, we used two conditions.
186 First, we only kept samples for which at least 60% of the replicates had the same synergy score
187 sign. Second, we only kept the samples for which the standard deviation of the synergy score
188 over replicates did not exceed 0.1. Finally, to obtain a single consensus synergy score for each
189 sample, we calculated the replicates' median score. The final dataset, comprised of samples
190 that were present in all three datasets above and passed the filtering steps, contained 43,181
191 (drug1, drug2, CCL) triple samples corresponding to 670 unique drugs, 2353 unique drug-pairs

192 and 75 unique CCLs. Finally, in order to ensure that the trained model is invariant to the order
193 of the drugs in its input space, for each triple sample we constructed a new sample by changing
194 the order of the drugs (Figure 2), resulting in a total of $86{,}362$ triple samples. It is important to
195 note that the samples with known synergy scores above only correspond to 0.128% of all
196 possible combinations of 670 drugs and 75 CCLs, further emphasizing the importance of
197 transductive computational models to complete this sparse 3-dimensional tensor.
198
199 We used relative inhibition (RI), available in DrugComb, as the response of each drug in a CCL.
200 RI represents the normalized area under the dose-response curve after a log_{10} transformation
201 [28]. More precisely, RI is the proportion of this area under the dose-response curve to the
202 maximal area that one drug can achieve at the same range of dosages. This metric represents
203 the overall inhibition effect of a drug and has shown to be a robust way to represent drugs'
204 sensitivity [29]. Additionally, the RI metric allows the comparison of drug responses obtained
205 using different dose ranges [28].
206
207 **Baseline models**
208 We considered several baseline models for benchmarking the performance of MARSY. Off-the-
209 shelf models included LASSO, ElasticNet, Support Vector Machine Regression (SVM), Random
210 Forests (RF) and a multi-layer perceptron (MLP) with the same depth and width as MARSY.
211 State-of-the-art algorithms, developed by others for the same task, included TreeCombo [19],
212 DeepSynergy [21], and MatchMaker [8]. All models were trained on the same dataset (with identical
213 folds across models) and evaluated similarly to ensure a fair comparison. Hyperparameters of

these models were tuned using the same independent validation set used to tune MARSY's hyperparameters. The model that performed best on the validation set was used for cross-validation analysis on the main dataset. In addition to the baselines above, we performed an extensive ablation study to assess the effect of each component of MARSY's architecture on its performance. The details of the ablation study are provided in Results.

Cross-validation, data splitting, and evaluation metrics

The performance of MARSY (and other methods) was evaluated using a five-fold cross validation (CV) approach combined with two data splitting strategies. The main dataset described above was randomly divided into five folds (based on each data split strategy described below); four folds were used to train each model using the hyperparameters obtained from the independent validation set (not used as part of these folds) and the remaining fold was used for performance evaluation. This was repeated five times so that each fold is set as the test set once. The mean and standard deviation of performance metrics on these test sets were used for evaluation.

Two data splitting strategies were used corresponding to the transductive setup of this study: leave-triple-out and leave-pair-out. In the former strategy, random triples (drug-pair and CCL) were selected randomly to construct the five folds. In the latter strategy, the folds were randomly selected based on drug-pairs, to ensure that the pairs of drugs are not seen together in the training set. To ensure a fair comparison, we used the exact same folds for all models and evaluated them in a similar manner. We evaluated the performance of MARSY and other

methods in predicting the continuous-valued drug synergy scores using Root Mean Squared Error (RMSE), Spearman's Rank Correlation Coefficient (SCC), and Pearson Correlation Coefficient (PCC). Additionally, we converted drug synergy scores to binary classes (synergistic or antagonistic) using different thresholds (0, ±1, ±5, ±10 and ±20), and used area under the receiver operating characteristic (AUROC) to evaluate the performance. For a threshold t, synergy scores above $+t$ were considered synergistic and values below $-t$ were considered antagonistic.

RESULTS

Performance of MARSY in prediction of drug synergy scores in a leave-triple-out data split

As the first evaluation study, we focused on the five-fold CV with leave-triple-out setup, discussed in Methods (Table 1). MARSY outperformed all baseline models using all performance metrics. In particular, MARSY achieved a Pearson correlation coefficient of PCC = 0.886 on ZIP and PCC = 0.864 on S_{mean}. The substantial difference between the performance of MARSY and the MLP with a similar depth and width, reveals the importance of learning separate representations for input pairs and triples, and using a multi-task predictor to constraint these representations using single drug priors. In this evaluation, all state-of-the-art algorithms outperformed off-the-shelf methods, yet did not achieve MARSY's performance.

Table1: The performance of MARSY and baseline methods using 5-fold leave-triple-out evaluation. The folds are the same across different models for a fair comparison. Best performance values are in bold-face and underlined. The mean and standard deviations are calculated across the folds. Models are sorted based on their ZIP PCC values.

Model	ZIP			S_{mean}		
	SCC	PCC	RMSE	SCC	PCC	RMSE
MARSY	**0.780** (±0.010)	**0.886** (±0.005)	**5.36** (±0.19)	**0.836** (±0.002)	**0.864** (±0.003)	**8.42** (±0.09)
MatchMaker	0.742 (±0.004)	0.873 (±0.006)	6.11 (±0.319)	0.810 (±0.003)	0.840 (±0.005)	9.84 (±0.23)
TreeCombo (XGBoost)	0.737 (±0.004)	0.870 (±0.003)	5.73 (±0.17)	0.817 (±0.002)	0.852 (±0.002)	8.77 (±0.06)
DeepSynergy	0.701 (±0.003)	0.869 (±0.004)	5.78 (±0.18)	0.803 (±0.003)	0.843 (±0.002)	9.14 (±0.12)
MLP	0.675 (±0.002)	0.840 (±0.004)	6.30 (±0.15)	0.781 (±0.011)	0.817 (±0.009)	9.72 (±0.24)
SVM	0.689 (±0.004)	0.783 (±0.006)	7.65 (±0.19)	0.766 (±0.005)	0.779 (±0.004)	10.67 (±0.07)
Random Forests	0.419 (±0.003)	0.646 (±0.016)	9.04 (±0.21)	0.619 (±0.006)	0.640 (±0.005)	12.98 (±0.05)
LASSO	0.342 (±0.009)	0.433 (±0.013)	10.41 (±0.21)	0.561 (±0.011)	0.525 (±0.013)	14.22 (±0.14)
Elastic Net	0.342 (±0.009)	0.432 (±0.012)	10.42 (±0.21)	0.562 (±0.011)	0.525 (±0.013)	14.22 (±0.15)

Next, we sought to determine the ability of MARSY (and other models) in predicting the synergistic and antagonistic drugs in each CCL (a binary classification task). Since the ZIP scores and S_{mean} are continuous values, we used five thresholds (0, ±1, ±5, ±10 and ±20) with an increasing degree of strictness to binarize these values. For a specific threshold $\pm t$, a score (ZIP or S_{mean}) of a triple (two drugs and a CCL) larger than $+t$ was considered to be synergistic, smaller than $-t$ was considered to be antagonistic, and values between these two thresholds were considered low-confidence and were discarded. Figure 3A and Table S1 show the AUROC of MARSY and other baselines models as a function of these thresholds, revealing the superior performance of MARSY in prediction of ZIP score in a classification formulation.

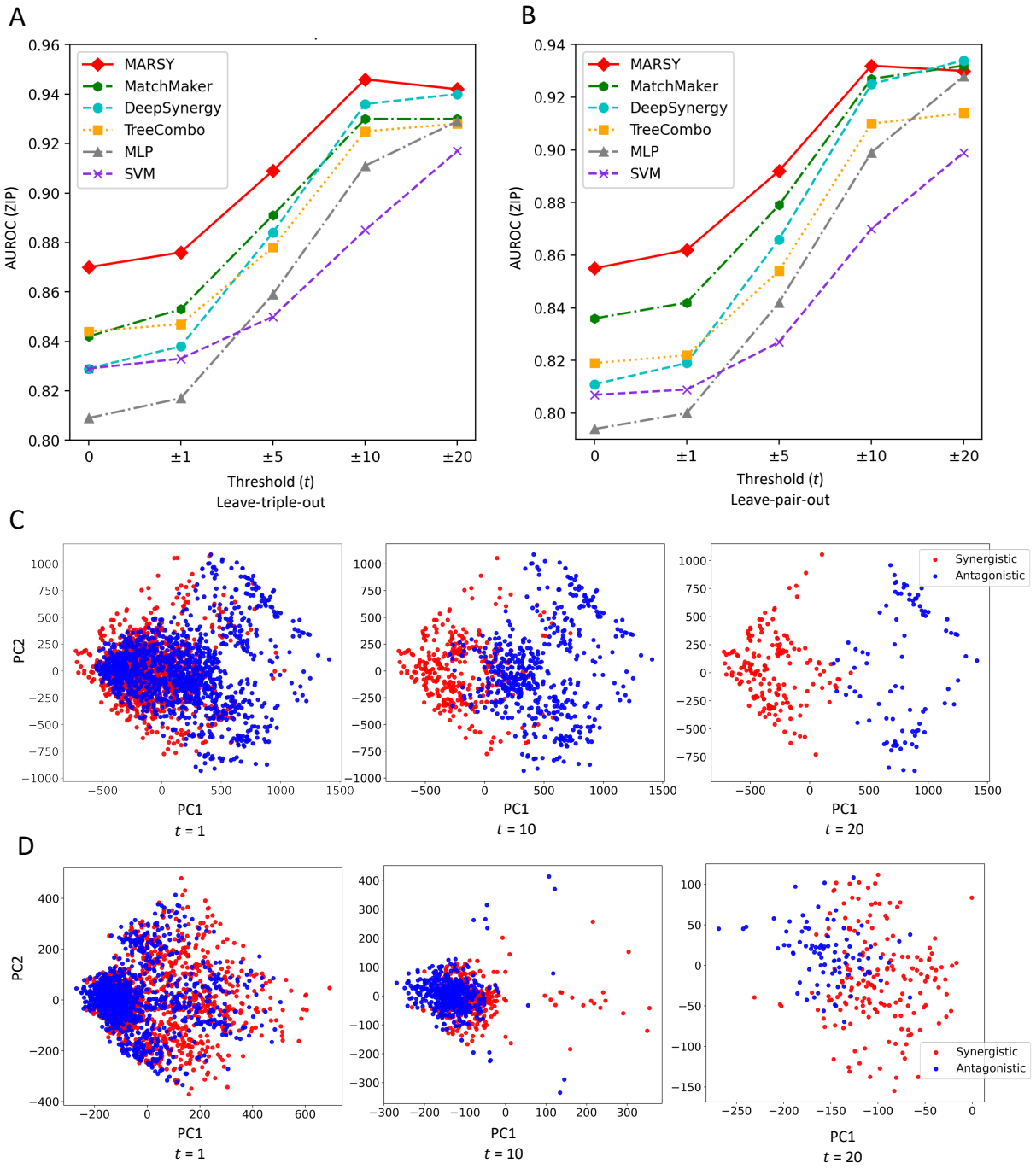

Figure 3: Performance of MARSY and its embeddings in identifying highly antagonistic and synergistic drugs. Panels A) and B) show classification performance of different methods in terms of AUROC for ZIP score for leave-triple-out and leave-pair-out five-fold cross validation, respectively. Panels C) and D) show principal component analysis (PCA) of synergistic and antagonistic drug-pairs for different thresholds (*t*) for SK-MEL-28 cell line based on their triple representations (panel C) and their pair representations (panel D) learned by MARSY.

Performance of MARSY in prediction of drug synergy scores in a leave-pair-out data split

Next, we sought to evaluate the performance of MARSY, when folds in the five-fold CV are determined only based on drug-pair identities (i.e., a drug-pair in the test set is never observed in the training set). Similar to the leave-triple-out framework, MARSY outperformed all other models (Table 2), reaching a Pearson correlation coefficient value of PCC = 0.875 for prediction of ZIP score and PCC = 0.841 for prediction of S_{mean}. Similar to the leave-triple-out, MatchMaker, TreeCombo and DeepSynergy were the best performing models after MARSY. Figure 3B and Table S1 show the AUROC of different models for different thresholds, further demonstrating the superior performance of MARSY for the majority of the thresholds.

Table2: The performance of MARSY and baseline methods using 5-fold leave-pair-out evaluation. The folds are the same across different models for a fair comparison. Best performance values are in bold-face and underlined. The mean and standard deviations are calculated across the folds. Models are sorted based on their ZIP PCC values.

Model	ZIP			S_{mean}		
	SCC	PCC	RMSE	SCC	PCC	RMSE
MARSY	**<u>0.749</u>** (±0.011)	**<u>0.875</u>** (±0.005)	**<u>5.62</u>** (±0.15)	**<u>0.809</u>** (±0.007)	**<u>0.841</u>** (±0.011)	**<u>9.06</u>** (±0.45)
MatchMaker	0.720 (±0.006)	0.864 (±0.007)	6.23 (±0.12)	0.788 (±0.009)	0.816 (±0.015)	10.34 (±0.46)
TreeCombo (XGBoost)	0.689 (±0.005)	0.856 (±0.006)	6.00 (±0.15)	0.775 (±0.010)	0.815 (±0.011)	9.69 (±0.37)
DeepSynergy	0.676 (±0.006)	0.860 (±0.008)	5.95 (±0.16)	0.762 (±0.004)	0.804 (±0.011)	10.17 (±0.42)
MLP	0.642 (±0.030)	0.822 (±0.023)	6.68 (±0.45)	0.744 (±0.029)	0.752 (±0.091)	11.36 (±0.24)
SVM	0.649 (±0.008)	0.773 (±0.012)	7.79 (±0.14)	0.737 (±0.011)	0.752 (±0.013)	11.15 (±0.42)
Random Forests	0.413 (±0.005)	0.650 (±0.012)	9.02 (±0.17)	0.608 (±0.016)	0.627 (±0.013)	13.14 (±0.33)
LASSO	0.333 (±0.019)	0.430 (±0.007)	10.43 (±0.17)	0.555 (±0.013)	0.519 (±0.012)	14.29 (±0.39)

Elastic Net	0.333 (±0.018)	0.429 (±0.008)	10.44 (±0.18)	0.555 (±0.013)	0.512 (±0.019)	14.29 (±0.40)

Detailed evaluation of MARSY's architecture

One aspect of MARSY's architecture is its multitask predictor. We replaced this predictor with a single-task predictor that only predicts drug synergy scores. This modification resulted in a lower performance based on all three measures and both data split strategies (Supplementary Table S2). For example, MARSY had a 6.12% higher SCC in leave-triple-out and a 5.05% higher SCC in leave-pair-out 5-fold CV for prediction of ZIP score. These results further demonstrate the importance of including a multi-task predictor in the architecture of MARSY. In fact, changing the predictor of DL baseline models also improved their performance (Supplementary Table S2).

Next, we sought to determine the role of input representations and encoders' architecture (Figure 1A) on MARSY's performance. Figure 3C and 3D (visually) show that both triple and pair representations are able to separate the synergistic and antagonistic drug-pairs. To systematically evaluate their role, we implemented six alternative architectures in which the predictor was identical to MARSY's multitask predictor, but the latent embeddings were learned using different number and types of encoders. Supplementary Figures S1-S6 show the architecture of these models and Table 3 shows their performance. These results show that overall, MARSY's choice of encoders is superior compared to these alternatives. In particular, using both Triple and Pair encoders is superior to only using the Triple encoder (Model1) or

replacing the Triple or Pair encoders with a CCL encoder (Models 2 and 3). Similarly, it performs better compared to using separate encoders for each drug (Models 4-6).

Table 3: The performance of different combination of embeddings using 5-fold leave-pair-out CV. Folds are the same across different models for a fair comparison. Best performance values are in bold-face and underlined. The mean and standard deviations are calculated across the folds. Architecture of the models are provided in Figure 1 and Supplementary Figures S1-S6.

Model	Num. Encoders	Encoder Type	ZIP			S_{mean}		
			SCC	PCC	RMSE	SCC	PCC	RMSE
MARSY	2	Pair, Triple	**0.741** (±0.018)	**0.871** (±0.009)	5.70 (±0.14)	**0.808** (±0.010)	**0.840** (±0.008)	**9.07** (±0.35)
Model1	1	Triple	0.679 (±0.024)	0.848 (±0.011)	6.22 (±0.18)	0.762 (±0.007)	0.801 (±0.012)	10.17 (±0.44)
Model2	2	CCL, Pair	0.706 (±0.018)	0.857 (±0.009)	6.52 (±0.13)	0.793 (±0.010)	0.826 (±0.007)	10.85 (±0.34)
Model3	2	CCL, Triple	0.734 (±0.017)	0.867 (±0.012)	5.77 (±0.18)	0.804 (±0.011)	0.837 (±0.007)	9.17 (±0.35)
Model4	3	Drug1, Drug2, CCL	0.714 (±0.019)	0.861 (±0.008)	6.26 (±0.15)	0.793 (±0.009)	0.828 (±0.007)	10.72 (±0.34)
Model5	3	Drug1, Drug2, Triple	0.737 (±0.013)	**0.871** (±0.008)	5.80 (±0.15)	0.798 (±0.008)	0.830 (±0.009)	9.94 (±0.44)
Model6	4	Drug1, Drug2, CCL, Pair	0.736 (±0.016)	0.870 (±0.008)	**5.69** (±0.14)	0.800 (±0.010)	0.833 (±0.005)	9.38 (±0.31)

Effect of hyperparameters and input features on MARSY's performance

As described in Methods, several aspects of MARSY's architecture were design choices selected a priori without hyperparameter tuning. To assess how these choices influence MARSY's performance, we conducted a study based on leave-pair-out 5-fold CV in which we ran MARSY with 576 different combinations of multiple hyperparameters to predict the ZIP score. These hyperparameters and their options included the learning rate (0.01, 0.001, 0.0001), the

329 optimizer (Adam, Adamax), the input activation function (Linear, ReLU), the batch size (64,
330 128), the dropout (with or without) and the width and depth of the encoders and the predictor
331 (see Supplementary Table S3 in Supplementary File S1 for the specific choices).
332
333 Our analysis revealed that a large learning rate (0.01) results in poor performance with Adamax
334 (Figure 4A) and also results in non-convergence of the Adam optimizer. However, the two
335 smaller learning rates do not suffer from the same issue. When the learning rate is selected
336 appropriately (the default 0.001 or 0.0001), the effects of other hyperparameters are relatively
337 small (Figure 4B and 4C). In particular, other than the learning rate, number of encoder layers
338 and the choice of optimizer had the largest effect, where 2 encoder hidden layers resulted in
339 1.4% higher median SCC (0.3% higher median PCC) compared to 3 hidden layers and Adamax
340 resulted in 1.1% higher median SCC (0.5% higher median PCC) compared to Adam (Figure 4C
341 and Supplementary Figure S7). In conclusion, learning rate seems to have the most effect
342 (consistent with our experience in other related studies [12]), but when large learning rates are
343 excluded, the results are not too sensitive to the choice of hyperparameters. However, if
344 computational complexity is not an issue, marginal improvements can be achieved using
345 hyperparameter tuning based on an independent validation set.
346

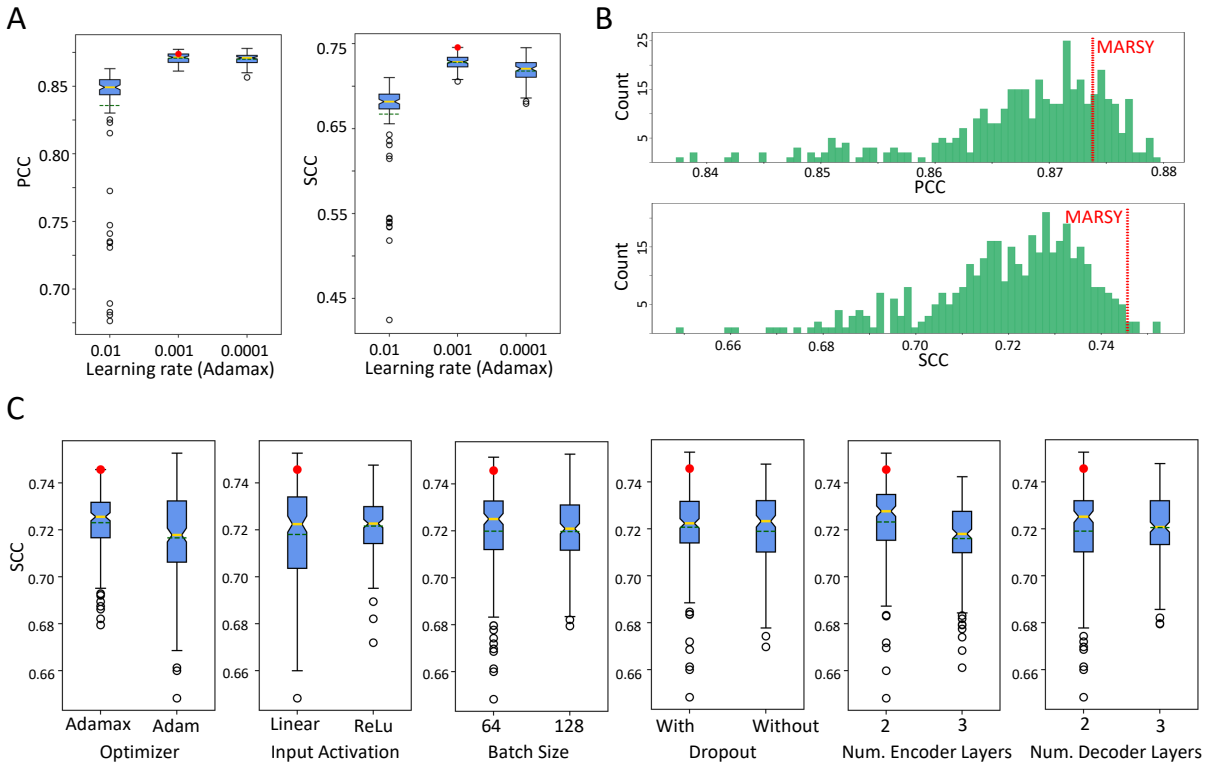

Figure 4: The effect of hyperparameters on performance of MARSY for prediction of ZIP score in leave-pair-out 5-fold CV. In boxplots, red circle represents MARSY, solid horizontal yellow line shows the median and dashed green line shows the mean. A) Effect of learning rate on the performance of MARSY with Adamax optimizer. B) The histogram of performance metrics based on different hyperparameter options, excluding a learning rate of 0.01. C) The boxplots show the distribution of SCC for different hyperparameter options. Runs with a learning rate of 0.01 are excluded due to the large number of non-converging runs.

MARSY uses LINCS drug signatures to represent each drug. We next asked how the performance of MARSY changes, if alternative drug signatures (e.g., based on their chemical structure) is used instead. We obtained the chemical representations of drugs from the DrugBank database [30] and used the RDKit python library [31] to extract two types of chemical feature representations: “Morgan fingerprints” and “molecular descriptors”. To make the evaluation consistent, we only conserved samples from DrugComb for which we could obtain both chemical feature representations and LINCS drug signatures, along with the gene

363 expression of the CCLs, resulting in 73,768 samples. Table 4 shows the performance of MARSY
364 with different drug features using this dataset. Considering all three metrics on the two
365 evaluation setups, LINCS signature performed better compared to Morgan fingerprint and
366 molecular descriptor. We also concatenated different features to provide a combination of
367 them as input to the model. Combination of Morgan fingerprints and molecular descriptors
368 performed better than each of them individually in leave-triple-out, but its performance was
369 inferior to Morgan fingerprints in leave-pair-out. Additionally, combining all three types of
370 representations performed slightly better than LINCS signature alone (and the other
371 representation choices), however one shortcoming of this combination is that it requires
372 availability of multiple type of representations for each drug, which sometimes limits the
373 number of datapoints for analysis (as was the case in our study).
374

375 **Table 4:** The performance of MARSY using distinct types of drug features. Evaluation is
376 performed using 5-fold leave-triple-out and 5-fold leave-pair-out on the prediction of the ZIP
377 synergy score. Best performance values are in bold-face and underlined. The mean and
378 standard deviations are calculated across the folds.
379

Drug Features	Leave-Triple-Out			Leave-Pair-Out		
	SCC	PCC	RMSE	SCC	PCC	RMSE
Molecular descriptors	0.702 (±0.014)	0.849 (±0.005)	5.57 (±0.12)	0.676 (±0.016)	0.841 (±0.009)	5.67 (±0.35)
Morgan fingerprints	0.739 (±0.001)	0.858 (±0.007)	5.24 (±0.13)	**0.729** (±0.011)	0.854 (±0.009)	5.36 (±0.21)
Molecular descriptors & Morgan fingerprints	0.746 (±0.003)	0.861 (±0.007)	5.16 (±0.15)	0.714 (±0.016)	0.846 (±0.012)	5.42 (±0.35)
LINCS signature	0.756 (±0.007)	0.869 (±0.005)	5.02 (±0.11)	0.725 (±0.012)	0.857 (±0.009)	**5.22** (±0.29)
All features combined	**0.757** (±0.005)	**0.870** (±0.005)	**5.01** (±0.13)	0.727 (±0.010)	**0.858** (±0.008)	**5.22** (±0.28)

380

381 **Prediction of drug synergy scores for new triples not present in DrugComb**

382 Next, we used our trained model to predict the ZIP synergy scores of 133,722 new (drug-pair,
383 CCL) triples that were not present in DrugComb. These triples correspond to all possible
384 combinations of 69 unique drugs that appeared in at least 10% of the drug-pairs in our training
385 set (Supplementary Table S4). Given these synergy scores, we first asked which drug-pairs show
386 a synergistic effect on all 75 CCLs. We identified seven drug-pairs that had a minimum ZIP score
387 larger than 2 across all CCLs. The majority of these corresponded to the combination of
388 vincristine (a chemotherapy agent) with other drugs. In particular, two of these combinations
389 corresponded to tyrosine kinase inhibitors (TKI) lapatinib (mean ZIP = 28.6) and imatinib (mean
390 ZIP = 17.9) combined with vincristine. Several independent studies have shown the synergistic
391 effect of these drugs. For example, these two TKIs sensitized KBV20C oral cancer cells to
392 vincristine, and their combination up-regulated apoptosis and reduced cell viability [32]. In
393 addition, lapatinib significantly increased the efficacy of vincristine in epidermoid carcinoma C-
394 A120 cells that overexpress multidrug resistance–associated protein 1 (MRP1) [33]. Another
395 noteworthy combination corresponded to docetaxel (chemotherapy) and veliparib (a PARP
396 inhibitor), (mean ZIP = 16.3). The siRNA knockdown of PARP1 in PC3 cell lines has been shown
397 to enhance docetaxel activity [34], and MARSY predicted the ZIP score of docetaxel and PARP-
398 inhibitor veliparib in PC3 to be 9.8 (reflecting synergism).

399

400 Next, we sought to identify drug combinations that show synergistic effect in a tissue-specific
401 manner. For this purpose, we focused on breast tissue, since it had the highest number of CCLs

402 in our cohort (n = 26). To identify breast cancer-specific synergistic combinations, we required
403 a drug-pair to not be antagonistic in any of the CCLs of breast cancer (minimum ZIP in the tissue
404 > 0) and then ranked the remaining drug-pairs based on the difference between their mean ZIP
405 score in CCLs of breast cancer and the mean ZIP score of CCLs of other tissues (Supplementary
406 Table S5). The top-ranked drug-pair corresponded to combination of paclitaxel (a
407 chemotherapy widely used in treating breast cancer) and ruxolitinib (a selective JAK1/2 kinase
408 inhibitor). A one-sided Mann Whitney U test also showed this difference to be statistically
409 significant (Benjamini-Hochberg false discovery rate = 7.76E-3). Ruxolitinib has been shown to
410 enhance the efficacy of paclitaxel in a synergistic manner in ovarian cancer cells [35]. Recent
411 studies of triple negative breast cancer (TNBC) samples has shown that inhibition of JAK1/2 by
412 ruxolitinib sensitizes cancer cells to paclitaxel, both *in-vitro* and *in-vivo* [36,37]. A phase I clinical
413 trial of the combination of these two drugs in HER2-negative metastatic breast cancer patients
414 was recently completed and showed these two drugs to be well tolerated by the participants
415 [38]. Following these results, a phase II randomized clinical trial is currently undergoing for the
416 combination of these two drugs in TNBC patients [38]. Our results also showed that TNBC CCLs in
417 our cohort have a higher mean ZIP score for this drug combination compared to other subtypes
418 of breast cancer (difference mean ZIP = 1.4).
419
420 These examples illustrate the utility of MARSY to suggest synergistic drug combination
421 candidates for follow-up drug screening experiments.
422
423 **Discussion and Conclusion**

424 In this study, we proposed MARSY, a novel deep multi-task learning method for prediction of
425 synergy scores of drug-pairs in difference cancer cell lines. Extensive evaluations using four
426 different metrics and two methods of data splitting for cross-validation revealed the superior
427 performance of MARSY compared to various computational models designed for this task.
428 MARSY's architecture was designed to learn distinct embeddings that capture different and
429 complimentary views of the input features and are informative not just for drug synergy score
430 prediction, but also enable single drug response prediction. Given the gene expression profile
431 of CCLs as well as the LINCS signatures of each drug, the two encoders of MARSY learn different
432 views of the input features, capturing the interplay of the drugs, as well as the drugs and the
433 CCL. Our ablation study confirmed that incorporating both of these encoders is beneficial to
434 the performance of the model. Moreover, the predictor of MARSY utilizes auxiliary tasks of
435 single drug prediction to enhance the synergy score prediction performance, by enhancing the
436 embeddings learned during the end-to-end training of the model. We also observed that these
437 auxiliary tasks can be helpful to other DL-based baseline models.
438
439 Unlike many previous models that utilize drug molecular profiles, we used the signatures of
440 each drug representing the changes in the expression profiles of two cancer cell lines after
441 administration of the drug. While replacing these signatures with chemical and molecular
442 information of drugs resulted in a lower performance, integrating them together slightly
443 improved the results, showcasing the complimentary information available in each of these
444 data modalities.
445

In this study, our focus was on developing a transductive model to impute the missing values in our dataset. As such, we used MARSY to predict the synergy scores of 133,722 new (drug-pair, CCL) triples that were not present in DrugComb and validated some of these predictions and insights obtained from them using independent datasets. However, one of the remaining challenges in this domain is developing models that can predict the synergy scores for unseen CCLs. Currently, the main challenge in achieving this goal is data availability. For example, after data cleaning, pre-processing and quality control, we ended up with only 75 CCLs that were usable for method development in this study. Such small numbers do not allow deep learning model development that generalizes well to unseen CCLs. However, as new data becomes available, we expect that this shortcoming of current datasets will be resolved and accurate models that can generalize to unseen CCLs can be developed.

Funding

This work was supported by the Government of Canada's New Frontiers in Research Fund (NFRF) [NFRFE-2019-01290] (AE), by Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2019-04460 (AE), and by McGill Initiative in Computational Medicine (MiCM) (AE). This work was also funded by Génome Québec, the Ministère de l'Économie et de l'Innovation du Québec, IVADO, the Canada First Research Excellence Fund and Oncopole, which receives funding from Merck Canada Inc. and the Fonds de Recherche du Québec – Santé (AE). This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca).

Data and Code Availability

All data generated as part of this study are provided as supplementary files. An implementation of MARSY in python is freely available at: https://github.com/Emad-COMBINE-lab/MARSY.

Supplementary Files

Supplementary File S1: This file contains Supplementary Figures S1-S7, Supplementary Table S3, and the caption of Supplementary Tables S1, S2, S4, and S5, which are provided as separate xlsx files.

478 **REFERENCES**

479 1 Li, J., Tong, X.-Y., Zhu, L.-D. & Zhang, H.-Y. A Machine Learning Method for Drug
480 Combination Prediction. *Frontiers in Genetics* **11**, doi:10.3389/fgene.2020.01000
481 (2020).
482 2 Sun, Y. *et al.* Combining genomic and network characteristics for extended capability
483 in predicting synergistic drugs for cancer. *Nature Communications* **6**, 8481,
484 doi:10.1038/ncomms9481 (2015).
485 3 Bayat Mokhtari, R. *et al.* Combination therapy in combating cancer. *Oncotarget* **8**,
486 38022-38043, doi:10.18632/oncotarget.16723 (2017).
487 4 Madani Tonekaboni, S. A., Soltan Ghoraie, L., Manem, V. S. K. & Haibe-Kains, B.
488 Predictive approaches for drug combination discovery in cancer. *Brief Bioinform* **19**,
489 263-276, doi:10.1093/bib/bbw104 (2018).
490 5 Li, H., Li, T., Quang, D. & Guan, Y. Network Propagation Predicts Drug Synergy in
491 Cancers. *Cancer Research* **78**, 5446-5457, doi:10.1158/0008-5472.Can-18-0740 (2018).
492 6 Li, H., Hu, S., Neamati, N. & Guan, Y. TAIJI: approaching experimental replicates-level
493 accuracy for drug synergy prediction. *Bioinformatics* **35**, 2338-2339,
494 doi:10.1093/bioinformatics/bty955 (2018).
495 7 Li, S. *et al.* Prediction of Synergistic Drug Combinations for Prostate Cancer by
496 Transcriptomic and Network Characteristics. *Frontiers in Pharmacology* **12**,
497 doi:10.3389/fphar.2021.634097 (2021).
498 8 Kuru, H. I., Tastan, O. & Cicek, E. MatchMaker: A Deep Learning Framework for Drug
499 Synergy Prediction. *IEEE/ACM Trans Comput Biol Bioinform* **Pp**,
500 doi:10.1109/tcbb.2021.3086702 (2021).
501 9 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of
502 anticancer drug sensitivity. *Nature* **483**, 603-607, doi:10.1038/nature11003 (2012).
503 10 Yang, W. *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for
504 therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* **41**, D955-961,
505 doi:10.1093/nar/gks1111 (2013).
506 11 Costello, J. C. *et al.* A community effort to assess and improve drug sensitivity
507 prediction algorithms. *Nat Biotechnol* **32**, 1202-1212, doi:10.1038/nbt.2877 (2014).
508 12 Hostallero, D. E., Li, Y. & Emad, A. Looking at the BiG picture: Incorporating bipartite
509 graphs in drug response prediction. *bioRxiv*, 2021.2008.2011.455993,
510 doi:10.1101/2021.08.11.455993 (2022).
511 13 Huang, E. W., Bhope, A., Lim, J., Sinha, S. & Emad, A. Tissue-guided LASSO for
512 prediction of clinical drug response using preclinical samples. *PLoS Comput Biol* **16**,
513 e1007607, doi:10.1371/journal.pcbi.1007607 (2020).
514 14 Hostallero, D. E., Wei, L., Wang, L., Cairns, J. & Emad, A. A Deep Learning Framework
515 for Prediction of Clinical Drug Response of Cancer Patients and Identification of Drug
516 Sensitivity Biomarkers using Preclinical Samples. *bioRxiv*, 2021.2007.2006.451273,
517 doi:10.1101/2021.07.06.451273 (2021).
518 15 Zagidullin, B. *et al.* DrugComb: an integrative cancer drug combination data portal.
519 *Nucleic Acids Res* **47**, W43-w51, doi:10.1093/nar/gkz337 (2019).

16 Amzallag, A., Ramaswamy, S. & Benes, C. H. Statistical assessment and visualization of synergies for large-scale sparse drug combination datasets. *BMC Bioinformatics* **20**, 83, doi:10.1186/s12859-019-2642-7 (2019).

17 Sidorov, P., Naulaerts, S., Ariey-Bonnet, J., Pasquier, E. & Ballester, P. J. Predicting Synergism of Cancer Drug Combinations Using NCI-ALMANAC Data. *Frontiers in Chemistry* **7**, doi:10.3389/fchem.2019.00509 (2019).

18 Jeon, M., Kim, S., Park, S., Lee, H. & Kang, J. In silico drug combination discovery for personalized cancer therapy. *BMC Syst Biol* **12**, 16, doi:10.1186/s12918-018-0546-1 (2018).

19 Janizek, J. D., Celik, S. & Lee, S.-I. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. *bioRxiv*, 331769, doi:10.1101/331769 (2018).

20 Celebi, R., Bear Don't Walk, O., Movva, R., Alpsoy, S. & Dumontier, M. In-silico Prediction of Synergistic Anti-Cancer Drug Combinations Using Multi-omics Data. *Scientific Reports* **9**, 8949, doi:10.1038/s41598-019-45236-6 (2019).

21 Preuer, K. *et al.* DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **34**, 1538-1546, doi:10.1093/bioinformatics/btx806 (2017).

22 Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

23 Reinhold, W. C. *et al.* CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. *Cancer Research* **72**, 3499-3511, doi:10.1158/0008-5472.Can-12-1370 (2012).

24 Shankavaram, U. T. *et al.* CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics* **10**, 277, doi:10.1186/1471-2164-10-277 (2009).

25 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e1417, doi:10.1016/j.cell.2017.10.049 (2017).

26 Yadav, B., Wennerberg, K., Aittokallio, T. & Tang, J. Searching for Drug Synergy in Complex Dose–Response Landscapes Using an Interaction Potency Model. *Computational and Structural Biotechnology Journal* **13**, 504-513, doi:https://doi.org/10.1016/j.csbj.2015.09.001 (2015).

27 Malyutina, A. *et al.* Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLOS Computational Biology* **15**, e1006752, doi:10.1371/journal.pcbi.1006752 (2019).

28 Zheng, S. *et al.* DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Research* **49**, W174-W184, doi:10.1093/nar/gkab438 (2021).

29 Douglass, E. F. *et al.* A community challenge for a pancancer drug mechanism of action inference from perturbational profile data. *Cell Reports Medicine* **3**, 100492, doi:https://doi.org/10.1016/j.xcrm.2021.100492 (2022).

30 Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* **46**, D1074-d1082, doi:10.1093/nar/gkx1037 (2018).

31 Landrum, G. (Academic Press Cambridge, 2013).

563 32 KIM, J. Y., KIM, H. S. & YOON, S. Tyrosine Kinase Inhibitors Imatinib and Erlotinib
564 Increase Apoptosis of Antimitotic Drug-resistant KBV20C Cells Without Inhibiting P-gp.
565 *Anticancer Research* **39**, 3785-3793, doi:10.21873/anticanres.13527 (2019).
566 33 Ma, S.-l. *et al.* Lapatinib antagonizes multidrug resistance-associated protein 1-
567 mediated multidrug resistance by inhibiting its transport function. *Mol Med* **20**, 390-
568 399, doi:10.2119/molmed.2014.00059 (2014).
569 34 Wu, W. *et al.* Inhibition of PARP1 by small interfering RNA enhances docetaxel activity
570 against human prostate cancer PC3 cells. *Biochemical and Biophysical Research*
571 *Communications* **442**, 127-132, doi:https://doi.org/10.1016/j.bbrc.2013.11.027 (2013).
572 35 Han, E. S. *et al.* Ruxolitinib synergistically enhances the anti-tumor activity of paclitaxel
573 in human ovarian cancer. *Oncotarget* **9** (2018).
574 36 Lian, B. *et al.* Truncated HDAC9 identified by integrated genome-wide screen as the
575 key modulator for paclitaxel resistance in triple-negative breast cancer. *Theranostics*
576 **10**, 11092-11109, doi:10.7150/thno.44997 (2020).
577 37 Han, J. *et al.* JAK2 regulates paclitaxel resistance in triple negative breast cancers.
578 *Journal of Molecular Medicine* **99**, 1783-1795, doi:10.1007/s00109-021-02138-3
579 (2021).
580 38 Lynce, F. *et al.* Phase I study of JAK1/2 inhibitor ruxolitinib with weekly paclitaxel for
581 the treatment of HER2-negative metastatic breast cancer. *Cancer Chemotherapy and*
582 *Pharmacology* **87**, 673-679, doi:10.1007/s00280-021-04245-x (2021).
583