

## RESEARCH ARTICLE

# Combination therapy synergism prediction for virus treatment using machine learning models

Shayan Majidifar<sup>1</sup>, Arash Zabihian<sup>2</sup>, Mohsen Hooshmand<sup>1\*</sup>

**1** Department of Computer Science and Information Technology, Institute for Advanced Studies in Basic Sciences (IASBS), Zanjan, Iran, **2** Department of QA, Kimia Zist Parsian Pharmaceutical Company, Zanjan, Iran

\* [mohsen.hooshmand@iasbs.ac.ir](mailto:mohsen.hooshmand@iasbs.ac.ir)**OPEN ACCESS**

**Citation:** Majidifar S, Zabihian A, Hooshmand M (2024) Combination therapy synergism prediction for virus treatment using machine learning models. PLoS ONE 19(9): e0309733. <https://doi.org/10.1371/journal.pone.0309733>

**Editor:** Michael Nevels, University of St Andrews, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

**Received:** June 2, 2024

**Accepted:** August 16, 2024

**Published:** September 4, 2024

**Copyright:** © 2024 Majidifar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data and code of Virus Combination Therapy are freely available at <https://github.com/BioinformaticsIASBS/CombinationTherapy>.

**Funding:** This work is based upon research funded by Iran National Science Foundation (INSF) under project No.4027788. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Combining different drugs synergistically is an essential aspect of developing effective treatments. Although there is a plethora of research on computational prediction for new combination therapies, there is limited to no research on combination therapies in the treatment of viral diseases. This paper proposes AI-based models for predicting novel antiviral combinations to treat virus diseases synergistically. To do this, we assembled a comprehensive dataset comprising information on viral strains, drug compounds, and their known interactions. As far as we know, this is the first dataset and learning model on combination therapy for viruses. Our proposal includes using a random forest model, an SVM model, and a deep model to train viral combination therapy. The machine learning models showed the highest performance, and the predicted values were validated by a t-test, indicating the effectiveness of the proposed methods. One of the predicted combinations of acyclovir and ribavirin has been experimentally confirmed to have a synergistic antiviral effect against herpes simplex type-1 virus, as described in the literature.

## Introduction

Bioinformatics is an interdisciplinary domain among biology, mathematics, statistics, and computer science that tends to theoretically and practically explore the field of human health solutions [1]. In other words, it utilizes the notions and tools of computer science and engineering in the analysis or introduction of efficient solutions for working with biological, medical, and even pharmacological data and information. One of the aspects of bioinformatics is to assist the *drug discovery* industry [2]. This is because drug discovery is an expensive research area and always looks for methods that reduce the cost and time of proposing a new drug for a disease, especially in emergency situations [3]. The virus-based diseases like SARS-CoV-2 [4], Mpox [5], MERS-CoV [6, 7] confirm the necessity of introducing new treatments as fast as possible. However, the drugs need to be efficient with a low side effect [8]. To meet these goals, drug repurposing, a screening method, tries to locate new targets for the approved drugs [9]. First, it uses drugs that are approved, therefore have lower side effects, and can be trusted in

treatments. Additionally, this approach narrows down the search space and consequently the cost and time for introducing the new drug. AI approaches, especially machine learning models, are commonly used in drug repurposing. The proposed drug repurposing methods cover a wide range of approaches from machine learning, e.g., logistic regression [10], random forest, support vector machine, neural networks [11–14], a spectrum of deep learning methods [15, 16] such as DTINet [17], NeoDTI [18], HIDTI [19], MolTrans [20], TransDTI [21]. Certain drug repurposing techniques have focused on predicting new associations between viruses and antivirals [22–27]. All previous methods have only considered single-drug treatments and have not explored the synergistic effects of combining multiple drugs.

However, each drug in addition to controlling and treating properties may have side effects and therefore increasing its usage dose causes high-risk issues in the patient [28]. Moreover, using a higher dose of a drug may cause *drug resistance* and nullify the treatment's effectiveness [29]. Drug repurposing has another branch of drug-target association which uses more than one drug in the treatment of a target. It is called *combination therapy* [30] that tends to reduce the side effects of drugs, fix drug resistance, and more importantly increase the effect of the treatment, e.g., synergistic drug pairs [31]. Therefore, combination therapy aims to improve the treatment and drug efficacy [32].

The first method to check the efficacy of the combination of drug pairs is brute force search. No need to mention that this method is costly and uses a tremendous amount of time and resources. High-throughput screening is another approach to investigating combination therapy. Same as brute force it consumes time and resources tremendously. One approach to researching drugs for disease treatment is through computational methods that investigate the drug space and suggest drug pairs. Such machine learning methods have achieved significant prediction power in this research area [33].

Computational combination therapy in oncology is an enriched and hot topic nowadays [9, 34–38]. Preuer et al. used cancer cell line properties, i.e., gene expression, copy number, and gene mutation and drug information including, structural and molecular similarities and drug toxicity from Merck [34] and proposed a deep network to compute the synergistic score of combined drugs [36]. Zhang et al. used those entities from the NCI-ALMANAC [39] that have signaling pathways [37]. Zhang et al. [38] and Wang et al. [9] applied other deep models on new embeddings of cancer cell line properties. The former used autoencoder to drive new embeddings and the latter used kernel-based methods to extract meaningful features. Kuru et al. used two deep networks for embedding the generation of drugs and viruses from Drug-Combo [40] and the new representations were fed to a third deep network for synergistic prediction of drugs for cancer treatment [41]. Julkunen et al. utilized the NCI-ALMANAC [39] dataset and mentioned that the previous works on drug combination in oncology had not considered protein properties and biological information of drugs. Then, they used factorization machines to decompose the information into latent spaces [42]. Meng et al. used a graph learning method to estimate the synergistic effect of combination therapy [43].

As mentioned earlier while combination therapy for oncology is a hot field, there are no general studies for virus treatments using synergistic therapy. Tan et al. proposed a multiplex screening method for HIV treatments [44]. This work does not use predictive learning models and targeted treatment of a single virus. Few studies proposed combination therapy solutions for SARS-CoV-2 [33, 45]. Although both works proposed combination therapy using deep models, they are limited to SARS-CoV-2 and have no general dataset for combination therapy.

This work proposes several machine learning methods for analyzing and evaluating virus-antivirals combination therapy. To accomplish this, we create a dataset containing the characteristics of both viruses and antivirals. Then, we devise and apply several machine learning methods to evaluate the effect of AI-based methods on the subject. The results are promising,

and several new combined drugs for virus treatments are proposed. Based on our knowledge and the literature review, all research studies on virus treatment using combination therapy have been limited to experimental or single-virus treatment. Therefore, this is the first study on general virus combination therapy. The contribution of the paper is four-fold:

- First work on virus combination therapy.
- First complete dataset on virus combination therapy (CombTVir).
- Applying machine learning methods and evaluating the results.
- Applying t-test analysis for statistical analysis and prediction validation.
- Proposing new combined drugs for virus treatment. Some of these predictions have been confirmed in the literature.

The structure of the paper is as follows. Section *Dataset generation* describes the properties and aspects of the generated dataset. Section *Methods* introduces the proposed methods for combination therapy prediction. The results are reported in section *Results*. Section *Conclusion* concludes the paper.

## Dataset generation

This paper proposes a method for predicting effective antiviral combinations for treating viral diseases. The first step of this proposal is to find a suitable dataset that contains information on antivirals used for combination therapy. Unfortunately, there is currently no available dataset for viruses. Therefore, our paper's first contribution is the creation of a virus combination therapy dataset, which we call it "CombTVir" dataset.

Myhre et al. gathered and reported a list of 541 drug combinations [46–48], of which 372 combinations belong to small molecule-small molecule (SM-SM) synergism, 103 combinations belong to biotech-biotech synergism, and the remaining 66 combinations belong to other types of combinations, e.g., SM-biotech. Notably, the combination list was sourced from PubMed or clinical trials. The selected combinations are derived from experiments *in vitro*, *in vivo*, or clinical trial phases. We chose those 372 SM-SM combined drugs for the dataset. Before describing the generation of the dataset, it is necessary to clarify the modifications made to the combination therapy list. The list contains HIV and HIV-1 (there were no reported HIV-2 in the list). After analyzing the main references of HIV and HIV-1, we treated HIV-1 as equivalent to HIV. Herpes simplex virus (HSV) has two subtypes—HSV-1 and HSV-2. These subtypes are highly similar genetically [49]. Since the dataset did not indicate the HSV subtype, we assumed HSV-1 and HSV are similar in this work. Some rows in the dataset are identical, such as the combination of acyclovir with foscarnet on HSV-1, which is repeated twice. The difference between the two rows is whether they were experimented on *in vitro* [50] or not reported [51].

We selected 372 SM-SM combinations from the dataset and removed all biotech-biotech and biotech-SM combinations, resulting in 44 viruses and 211 drugs being included in the chosen combinations. Table 1 briefly reports the statistics of the dataset. With these 372

**Table 1. Combination therapy (CombTVir) dataset.**

Entity	Statistics
No of Viruses (V)	44
No. of antivirals (A)	211
No of approved interactions (AAV)	372
Sparsity	0.019%

<https://doi.org/10.1371/journal.pone.0309733.t001>

SM-SM combinations, we gathered information about them from NCBI [52] and DrugBank [53]. NCBI is the National Center for Biotechnology Information which provides access to bio-medical and genomic information. We gathered the Fasta version of viruses' sequences from NCBI. DrugBank is a freely accessible database that contains information on drugs and their targets; therefore, we collected the SMILES [54] of drugs from it. Thus, we have information on drugs and viruses.

In the next step, we prepared the feature vector of each antiviral and each virus by creating similarity matrices for antivirals and viruses. To compute the drugs' similarity matrix, as Bajusz et al. [55] suggested, we converted the SMILES of antivirals to fingerprints and then applied the following Tanimoto score [56] on each fingerprint pair.

$$Tanimoto(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

Then, the feature vector of each antiviral is its Tanimoto scores with whole antivirals. Consequently, their generated similarity matrix acts as the feature set of antivirals.

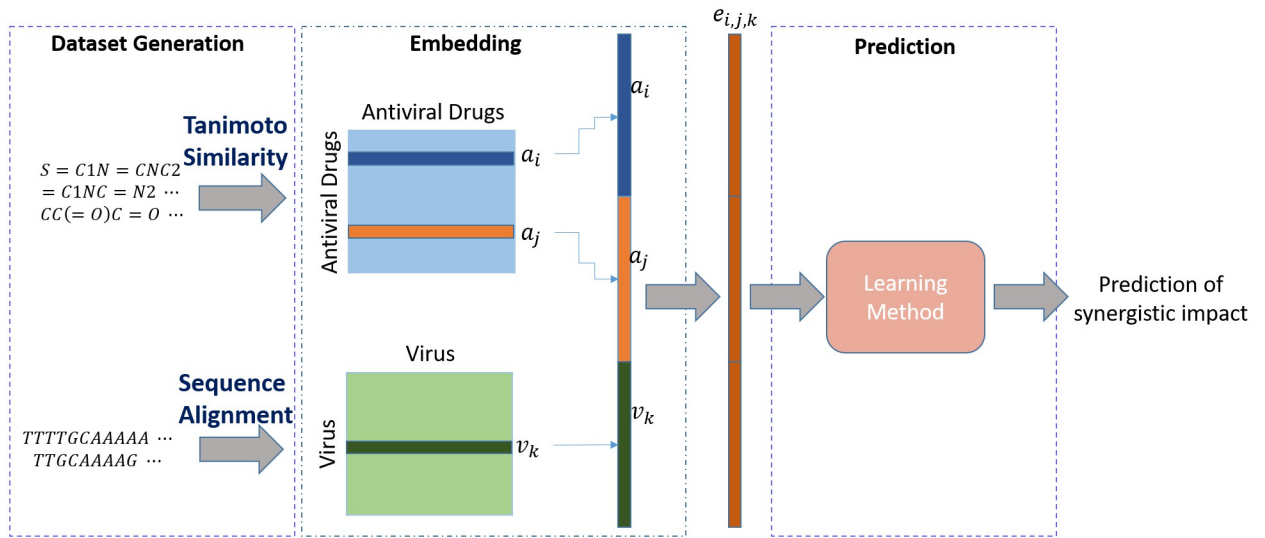
As mentioned earlier, we gathered the Fasta sequences of viruses from NCBI by choosing the complete genome version of the virus or its first row from the RefSeq section. Thus, we gathered the sequences of 44 viruses. Then, to prepare the viruses' feature vectors, we calculated their similarity matrix using sequence alignment [57]. We implemented the Smith-Waterman algorithm [58], a pairwise sequence alignment method, on every pair of sequences using the NUC44 score matrix. This algorithm takes two strings and aligns them to maximize the alignment score. It works as follows.

$$S_{i,j} = \max \begin{cases} S_{i-1,j-1} + s(a_i, b_j) & \nearrow \\ S_{i,j-1} + s(a_i, -) & \leftarrow \\ S_{i-1,j} + s(-, b_j) & \uparrow \end{cases} \quad (2)$$

Where, **a** and **b** are two strings with lengths of *m* and *n*, respectively. The first row of the equation states that when the *i*-th character of **a** and *j*-th of **b** match, the total score increases. When there is no match, the maximum value based on insertion or deletion is computed using the second or third row. The algorithm returns the value of  $S(m, n)$  as its alignment score [59]. These scores are considered as the entries of each virus feature vector. In other words, we compute the sequence alignment scores for each virus and generate the similarity matrix based on them. Then, each row of the similarity matrix is considered as the feature vector of its corresponding virus. Having these two similarity matrices and the list of available combination therapies, we have prepared the CombTVir dataset for further analysis in the next sections.

## Methods

As mentioned in the previous section, the dataset consists of the antiviral feature set *A*, the virus feature set *V*, and the antiviral-antiviral-virus associations *Y*. We consider the latter as labels. Having this, we aim at predicting the synergistic effect of combining two antivirals *i* and *j*, where  $\mathbf{a}_i, \mathbf{a}_j \in A$  on the given virus *k*,  $\mathbf{v}_k \in V$  using support vector machine, random forest, and a deep model which we call DRaW. Fig 1 shows the general framework of the proposed methods. The antiviral set *A* contains *m* antivirals and the virus set *V* contains *n* viruses. To use the machine learning methods, each identity of the problem, i.e., antiviral and virus, needs a corresponding feature vector. As mentioned in the previous section, the feature set is extracted from the similarity vectors of antivirals and viruses. The final feature vector of each



**Fig 1. Drug combination learning framework.** The framework prepares embeddings for each drug and each target based on their similarity information. Then, the corresponding embeddings of each drug-drug-target combination are concatenated, which is the input of the prediction step. The final step uses one of the proposed learning methods, i.e., SVM, RF, and DRaW to predict the interaction of each pair.

<https://doi.org/10.1371/journal.pone.0309733.g001>

combination is the fruit of concatenation of  $i$ -th and  $j$ -th antivirals ( $\mathbf{a}_i$  and  $\mathbf{a}_j$ ) and  $k$ -th virus, or  $\mathbf{e}_{i,j,k} = \mathbf{a}_i \parallel \mathbf{a}_j \parallel \mathbf{v}_k$ . Therefore, the vector  $\mathbf{e}_{i,j,k}$  represents the feature vector of  $i$ -th and  $j$ -th antivirals, and  $k$ -th virus with the aim of predicting the label  $y_{i,j,k} \in Y$  using the mentioned feature vector to minimize the general loss function as follows:

$$\min \mathcal{L} = \min \left( \sum_{i=1}^m \sum_{j=1, j \neq i}^m \sum_{k=1}^n \text{dist}(y_{i,j,k}, \hat{y}_{i,j,k}) \right).$$

Where,  $y_{i,j,k}$  shows the label of the synergistic effect of drugs  $i, j$  on virus  $k$ .  $\hat{y}_{i,j,k}$  shows its predicted version and is computed using an effective learning method. The function  $\text{dist}(\cdot, \cdot)$  is the distance function for the evaluation of the learning methods. As discussed earlier, combination therapy uses several learning methods from the literature. We use SVM [11], random forest [12] for their high performance in different domains of learning [13, 14], and a convolutional deep learning model due to their efficiency, performance, and reliability [60].

### Support vector machine

The support vector machine (SVM), introduced by Vapnik, finds the margin with the maximum length between two classes of data. However, since the two classes are not always linearly separable, various techniques such as the use of kernels and misclassification allowance are employed for optimal classification. Therefore, its loss function is as follows.

$$\min \mathcal{L} = \min_{G, b} \frac{1}{2} \|G\|^2 + C \sum_{i,j,k, i \neq j} \varepsilon_{i,j,k}, \tag{3}$$

$$y_{i,j,k}(G^t \mathbf{e}_{i,j,k} + b) \geq 1 - \varepsilon_{i,j,k}, \forall i, j \in \{1, \dots, m\}, k \in \{1, \dots, n\} \tag{4}$$

Where,  $G$  is support vectors,  $C$  is error regulation parameter,  $\varepsilon$  is the allowed error [61].

## Random forest

Random forest (RF) is an ensemble machine learning method that utilizes several decision trees and each tree randomly chooses several features from the feature sets. Following the learning phase of the trees, the class with the majority vote is chosen as the predicted label. Using several trees and a random selection of features for each tree leads to neutralizing the overfitting effect of decision trees. More importantly, the ensemble of trees yields a reliable prediction result of random forest. This principle makes a random forest a high-performance ML method for classification. In this work, the decision trees use *Gini* and *logloss* functions for score computation in each level of trees [62].

## DRaW—a deep learning method

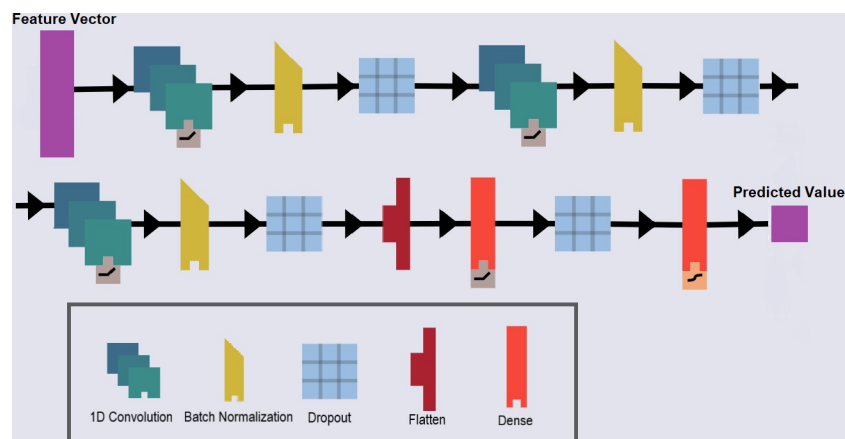
Fig 2 shows the architecture of the proposed deep model, DRaW. It consists of three CNN layers and each CNN layer is a combination of 1D convolution, batch normalization, and dropout layers. After the CNN layers, there are two dense layers with a dropout layer in between. All the internal activation functions are *ReLU* and the last layer activation function is a *sigmoid* function. DRaW accepts the  $\mathbf{e}_{i,j,k}$  as input feature vectors and computes their corresponding  $\hat{y}_{i,j,k}$ .

$$\hat{y}_{i,j,k} = DRaW(\mathbf{e}_{i,j,k})$$

Its loss function is binary cross-entropy on all members of the dataset.

$$\min \mathcal{L} = \min \left( - \sum_{i=1}^m \sum_{j=1, j \neq i}^m \sum_{k=1}^n [y_{i,j,k} \log(\hat{y}_{i,j,k}) + (1 - y_{i,j,k}) \log(1 - \hat{y}_{i,j,k})] \right). \quad (5)$$

Algorithm 1 presents the DRaW algorithm. The inputs are the antiviral-antiviral-virus label set  $Y$ , the antiviral feature matrix  $A$ , and the virus feature matrix  $V$ . Additionally, the algorithm needs the initiation of three more parameters of *ratio*, *folds*, and *epochs*. The *ratio* determines the positive-to-negative (P-to-N) sampling ratio. The *folds* parameter sets the number of folds for k-fold stratification, and finally, *epochs* identifies the number of iterations. The output is



**Fig 2. The DRaW deep model consists of three convolution layers, each containing a convolution, a batch normalization, and a dropout module.** The activation function used in the inner layers is ReLU. Finally, the last layer is the classification module, which uses a sigmoid activation function for classification.

<https://doi.org/10.1371/journal.pone.0309733.g002>

the predicted associations  $\hat{Y}$ . The algorithm in line 1 chooses a sample set of labels based on the *ratio* parameter. It chooses the whole positive samples, and a random number of negative samples using the ratio. For instance, when the sampling ratio is set to 1: 10, the algorithm randomly selects ten negative samples for every positive sample. Then, the algorithm employs a stratified version of the k-fold cross-validation procedure, and line 2 demonstrates the data folding process based on the *folds* parameter. The main section of the algorithm starts from line 3 and goes on. For each fold, the data is split into training and test sets based on the corresponding fold in line 4. The model is trained based on the features and labels of the training set in line 6, and this training is done in several epochs. The DRaW model in Fig 1 serves as the basis for this training. Line 7 predicts the test labels. Line 8 of the algorithm computes the loss function based on the binary cross entropy loss function introduced in Eq 5. After, ending the epochs, the algorithm predicts the test set labels. In the end, the algorithm calculates the performance based on the evaluation metrics presented in Section.

**Algorithm 1** Proposed Deep Model (DRaW)

**Input:**  $A, V, Y, ratio, folds, epochs$

**Output:**  $\hat{Y}$

```

1: data split(Y, ratio)
2: k-Fold ← stratified-k-Fold(folds)
3: for each fold in k-Fold do
4:   divide data into train and test
5:   For each epoch in epochs do
6:     Model = Training( $A_{tr}, V_{tr}, Y_{tr}$ )
7:      $\hat{Y} = Compute(A_{ts}, V_{ts}, Y_{ts})$ 
8:     Loss computation using Eq 5
9:   end for
10: end for
11: Performance evaluation

```

## Complexity analysis

Assuming a dataset with  $m$  antivirals and  $n$  viruses, the complexity analysis is divided into two parts: dataset preparation and generation of feature vectors for antivirals and viruses. As mentioned earlier, we used the Tanimoto score and *sequence alignment* score to create the similarity matrices. The Tanimoto score is used to measure the similarity between two sets, while the sequence alignment score is used to measure the similarity between two sequences. The Tanimoto score complexity is  $cn^2$ , where  $c$  is a small constant. This means that the runtime is fast, as a result, the entire procedure can be completed in just a fraction of a second. Performing pairwise sequence alignment for all viruses is a task that takes a considerable amount of time. The complexity of this algorithm is  $Cn^2$ , where  $C$  is a huge constant, therefore, it is a time-consuming computation. SVM training time complexity is in the range  $O(m^2n^2)$  and  $O(m^3n^3)$  depending on the  $C$  hyperparameter and its runtime is  $O(|G|mn)$ , where  $|G|$  is the number of support vectors [63]. The time complexity of random forest uses  $N$  trees each with at most  $V$  sampled features [64]. Therefore, its training time complexity is  $O(NVmn(\log m + \log n))$ . Its runtime is  $O(Nd)$ , where  $d$  is the depth of the tree. The DRaW runs for  $E$  epochs of each  $T$  long. Therefore, its time complexity is  $O(mnET)$  asymptotically.

## Results

This section provides the results of the proposed methods of virus-antiviral combination therapy. We performed 10-fold stratified cross-validation on a system with Ubuntu 22.04 LTS operating system. The system runs on an Intel Xeon Processor E5 v4 family with 4 CPU threads, 16 GB of RAM, and 20 GB of storage capacity.

The performance evaluation metrics are as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (6)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (7)$$

$$\text{AUC - ROC} = \sum_{i=1}^n \text{TPR}_i \Delta \text{FPR}_i, \quad (8)$$

$$\text{AUPR} = \sum_{i=1}^n \text{precision}_i \Delta \text{recall}_i. \quad (9)$$

Moreover, we conducted a t-test on the predicted results [65] to evaluate the domain applicability and statistical analysis. We assume the null hypothesis H0 results from a lack of correlation between the original and predicted labels. The alternative hypothesis states the existence of a correlation between the two sets. Large values of p-value confirm the H0 and small values reject it.

We conducted the simulation for several P-to-N sampling ratios, i.e., 1:3, 1:5, 1:10, 1:100, and 1:500. For the lower sampling ratios—1:10 and lower ones—the performance of all methods is almost equal and close to perfect. Therefore, we report the results for the sampling ratios of 1:100 and 1:500. In our study, we performed a grid search on various configurations of SVM and random forest to identify the optimal performance of these ML methods. For SVM, we analyzed three different kernels (Linear, Poly, and RBF) and evaluated three different values of C for each kernel. The results of this analysis are provided in S4-S6 Tables in [S1 File](#) for sampling ratios 1:10, 1:100, and 1:500, respectively. As the results show, the SVM with specifications “poly” kernel and C = 10 has the best performance. Therefore, we use this model of SVM for comparison with other learning models. Additionally, we evaluated the random forest model using two criteria, *Gini* and *logloss*. We also tested two different values for the maximum number of features for each criterion. The results of these analyses are presented in S9-S11 Tables in [S1 File](#) for sampling ratios 1:10, 1:100, and 1:500, respectively, in the [S1 File](#). The results confirm that the random forest with the *logloss* criteria and a maximum number of features for  $\log(n)$  has been chosen for comparison with other learning models. These configurations for RF and SVM were then used for general comparison purposes.

[Table 2](#) shows the metric scores for DRaW, SVM, and random forest for the P-to-N sampling ratio of 1:100. While all methods have the same accuracy, The SVM has the highest AUC-ROC and the random forest has the highest AUPR.

[Table 3](#) reports the results for the P-to-N sampling ratio of 1:500. The same pattern similar to [Table 2](#) happens for this ratio as well.

Visual comparison of AUC-ROC and AUPR for different methods is presented in [Fig 3](#). Results are reported for three P-to-N sampling ratios: 1:10, 1:100, and 1:500. In this study, we

**Table 2. Results for positive-to-negative sampling ratio of 1:100.**

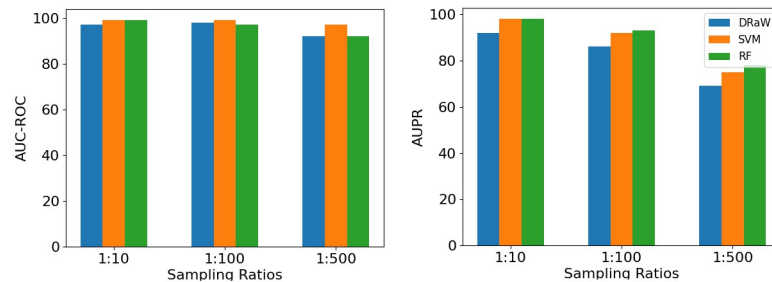
Method	DRaW	SVM	RF
Acc	<b>0.99 ± 0.0</b>	<b>0.99 ± 0.0</b>	<b>0.99 ± 0.0</b>
MCC	0.87 ± 0.02	0.85 ± 0.03	<b>0.92 ± 0.02</b>
AUC-ROC	0.98 ± 0.01	<b>0.99 ± 0.0</b>	0.97 ± 0.02
AUPR	0.86 ± 0.04	0.92 ± 0.03	<b>0.93 ± 0.04</b>

<https://doi.org/10.1371/journal.pone.0309733.t002>

**Table 3. Results for P-to-N sampling ratio of 1:500.**

Method	DRaW	SVM	RF
Acc	<b>0.99 ± 0.00</b>	<b>0.99 ± 0.0</b>	<b>0.99 ± 0.0</b>
MCC	0.77 ± 0.04	0.72. ± 0.06	<b>0.78 ± 0.05</b>
AUC-ROC	0.92 ± 0.03	<b>0.97 ± 0.01</b>	0.92 ± 0.02
AUPR	0.69 ± 0.06	0.75 ± 0.07	<b>0.78 ± 0.05</b>

<https://doi.org/10.1371/journal.pone.0309733.t003>



**Fig 3. AUC-ROC and AUPR values of different methods.** The x-axis displays P-to-N sampling ratios while the y-axis represents AUC-ROC and AUPR values for the left and right plots. (A) The AUC-ROC value of SVM remains stable and almost constant even when the sampling ratio increases. (B) In contrast, the right plot shows a decrease in the AUPR value of all methods. For higher sampling ratios, the AUPR value and overall performance of the random forest remain higher than other methods.

<https://doi.org/10.1371/journal.pone.0309733.g003>

compared the changes in AUC-ROC when varying the sampling ratio. Fig 3A shows that the AUC-ROC scores for 1:10 and 1:100 remain almost unchanged, regardless of whether DRaW, SVM, or RF are used. The ML methods outperform the deep model regarding the mentioned evaluation metric. Among the ML methods, the SVM has the highest AUC-ROC. However, all methods show a decrease in performance when increasing the sampling ratio to 1:500. In the figure labeled as Fig 3B, we can see the AUPR (Area Under the Precision-Recall Curve) of different methods for different P-to-N sampling ratios. As the P-to-N sampling ratio increases, there is a decrease in the AUPR scores of all methods. It is observed that DRaW has a lower score compared to ML (Machine Learning) methods for the whole sampling ratios. Random forest is the top performer based on AUPR for all sampling ratios.

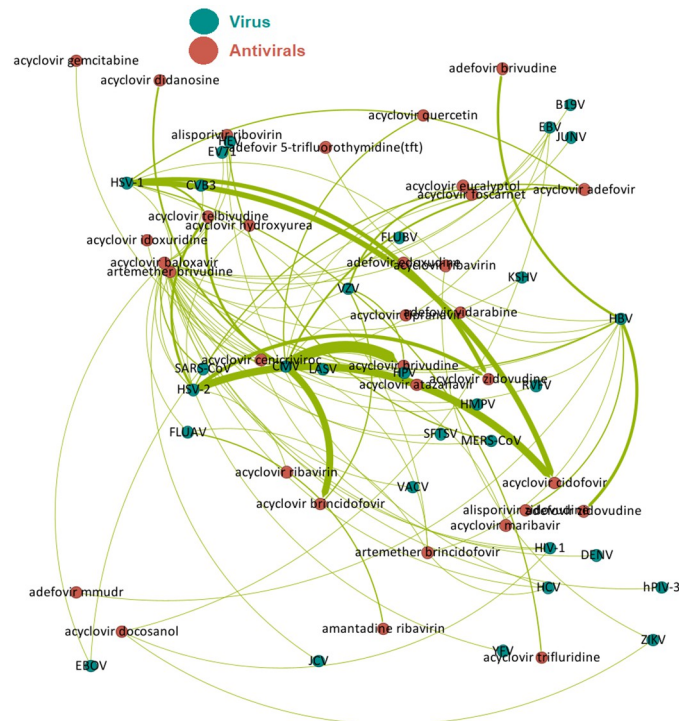
The validation of the proposed model is crucial for generalization and checking the suggested combinations. Therefore, we conducted a t-test statistical analysis to validate the prediction models. Table 4 shows the t-test results of the predicted values. It reports the significance for sampling ratios of 1:10, 1:100, and 1:500 for all methods, i.e., DRaW, SVM, and random forest. We set the threshold to 0.05. All predicted values have p-values below the threshold and reject the null hypothesis.

The results demonstrate that the proposed methods effectively predict synergistic combinations of antiviral drugs. Therefore, we present the predicted combinations of antiviral drugs

**Table 4. t-test validation.**

Ratio	SVM	RF	DRaW
1:10	5.71E-07	1.91E-11	3.5E-156
1:100	1.0E-100	4.59E-41	1.09E-81
1:500	1.0E-100	6.57E-94	9E-146

<https://doi.org/10.1371/journal.pone.0309733.t004>



**Fig 4. Predicted synergistically combined antivirals on viruses.**

<https://doi.org/10.1371/journal.pone.0309733.g004>

that are effective against previously unknown viruses. Fig 4 illustrates a schematic graph of the proposed antiviral drug combinations.

In order to validate the results, we conducted a literature search to identify antiviral drug combinations that have individually demonstrated effectiveness in treating specific viruses. For instance, while acyclovir and brincidofovir have shown treatment efficacy for CMV, our model suggests that combining the two could produce a synergistic effect. However, this proposed effect will need to be confirmed by future experimental studies.

Another prediction of the proposed model is the synergistic effect of acyclovir and cidofovir on HSV-1. Both of these medications are individually effective treatments for the mentioned virus. The literature also indicates that the combination of acyclovir and zidovudine has an additive effect on HSV-1, which the model also predicted to have a synergistic effect. Acyclovir and foscarnet have an additive effect on VZV, where our proposed machine learning models predict their synergistic treatment [66]. The additive combination of acyclovir and maribavir on CMV is predicted to have a synergistic treatment [67]. Additionally, it is predicted that acyclovir in combination with trifluridine and adefovir has a synergistic effect on treatments for HSV-1, and in combination with brincidofovir and brivudine has a synergistic effect on VZV. The model predicts that alisporivir and ribavirin have a synergistic effect on HCV and their additive effect has been confirmed experimentally. Clinical trials are necessary for the validation of these new combinations. Table 5 reports those predictions which at least have an additive treatment for viruses. Additionally, S13 and S14 Tables in S1 File report the complete list of unknown synergistic combination therapies against viruses predicted with proposed methods. The frequency shows the number of predictions in test sets.

More importantly, one of the predicted combinations, i.e., the synergistic effect of acyclovir and ribavirin against the herpes simplex type-1 virus (HSV-1) has been confirmed experimentally [68].

**Table 5. Predicted synergistic combinations of antivirals.** Each citation reports the efficacy of its corresponding antiviral against the virus. The complete list of predicted combinations is available in the [S1 File](#). Note that the synergistic effect of Acyclovir and Ribavirin against HSV-1 has been confirmed.

Antiviral1	Antiviral2	Virus
Acyclovir [69]	Ribavirin [70]	HSV-1 [68]
Acyclovir [71]	Cidofovir [72]	HSV-2
Acyclovir [66]	Brincidofovir [73]	CMV
Acyclovir [69]	Cidofovir [72]	HSV-1
Acyclovir [69]	Zidovudine [74]	HSV-1
Acyclovir [66]	Foscarnet [75]	VZV
Acyclovir [69]	Trifluridine [76]	HSV-1
Acyclovir [66]	Brincidofovir [77]	VZV
Acyclovir [66]	Brivudine [78]	VZV
Acyclovir [69]	Adefovir [79]	HSV-1
Alisporivir [80]	Ribavirin [81]	HCV
Acyclovir [66]	Maribavir [67]	CMV
Acyclovir [82]	Foscarnet [75]	EBV

<https://doi.org/10.1371/journal.pone.0309733.t005>

## Conclusion

This paper proposes machine learning models to predict the synergistic effects of antiviral combinations on viruses. While synergistic combination therapy has a rich history of research, to the best knowledge of the authors there is no research on computational combination therapy for viruses. Therefore, in this paper, we have proposed a first dataset for the virus synergistic combination therapy. Moreover, we conducted several learning methods including random forest, SVM, and a deep model for efficient prediction of the synergistic effect of combined antivirals on the virus. The results confirm the high performance of all proposed methods. The results show the high performance of the random forest model. Increasing the sampling ratios notably resulted in the random forest having the best performance. In the future, using attention-based learning methods to model synergistic viruses can improve results. Additionally, the feature vectors are similarity vectors of antivirals and viruses. The similarity vectors are based on linear operators like cosine similarity. Therefore, using the similarity vector can impact and decrease the effect of learning models. Therefore, the direct feeding of SMILES of antivirals can improve the performance of learning models. Combining the self-attention methods with different ways of preparing the input features is another area for further research.

This paper confirms the results by applying a t-test to the predicted results and rejecting the null hypothesis. Experimental analysis is required to validate proposed drug combinations and determine if their effects are additive or synergistic. One combination (not in the dataset), acyclovir and ribavirin, was successfully predicted and approved in the literature against HSV-1. It is worth mentioning that acyclovir shows up in most of the predictions. This is due to its frequent presence in most approved synergistic combination actions.

## Supporting information

**S1 File.** The supplementary material contains comprehensive information on machine learning hyperparameter optimization, comprehensive results, and predicted combinations.

(PDF)

## Acknowledgments

The authors would like to thank Fatemeh Nasiri for helping with the dataset collection, Masih Hajsaeedi for conducting the sequence alignment, and Javad Asghari for helping with implementation.

## Author Contributions

**Conceptualization:** Arash Zabihian, Mohsen Hooshmand.

**Methodology:** Shayan Majidifar, Arash Zabihian, Mohsen Hooshmand.

**Software:** Shayan Majidifar.

**Supervision:** Mohsen Hooshmand.

**Writing – original draft:** Shayan Majidifar, Arash Zabihian, Mohsen Hooshmand.

**Writing – review & editing:** Arash Zabihian, Mohsen Hooshmand.

## References

1. Bayat A. Science, medicine, and the future: Bioinformatics. *BMJ: British Medical Journal*. 2002; 324(7344):1018. <https://doi.org/10.1136/bmj.324.7344.1018> PMID: 11976246
2. Xia X. Bioinformatics and drug discovery. *Current topics in medicinal chemistry*. 2017; 17(15):1709–1726. <https://doi.org/10.2174/1568026617666161116143440> PMID: 27848897
3. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Molecular pharmaceutics*. 2016; 13(7):2524–2530. <https://doi.org/10.1021/acs.molpharmaceut.6b00248> PMID: 27200455
4. Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. *International journal of infectious diseases*. 2020; 94:44–48. <https://doi.org/10.1016/j.ijid.2020.03.004> PMID: 32171952
5. Rabaan AA, Al-Ahmed SH, Haque S, Sah R, Tiwari R, Malik YS, et al. SARS-CoV-2, SARS-CoV, and MERS-COV: a comparative overview. *Infez Med*. 2020; 28(2):174–184. PMID: 32275259
6. Rizk JG, Lippi G, Henry BM, Forthal DN, Rizk Y. Prevention and treatment of monkeypox. *Drugs*. 2022; 82(9):957–963. <https://doi.org/10.1007/s40265-022-01742-y> PMID: 35763248
7. Mitjà O, Ogoina D, Titanji BK, Galvan C, Muyembe JJ, Marks M, et al. Monkeypox. *The Lancet*. 2023; 401(10370):60–74. [https://doi.org/10.1016/S0140-6736\(22\)02075-X](https://doi.org/10.1016/S0140-6736(22)02075-X)
8. Kumar V, Dogra N. A comprehensive review on deep synergistic drug prediction techniques for cancer. *Archives of Computational Methods in Engineering*. 2022; 29(3):1443–1461. <https://doi.org/10.1007/s11831-021-09617-3>
9. Wang Y, Yang Y, Chen S, Wang J. DeepDRK: a deep learning framework for drug repurposing through kernel-based multi-omics integration. *Briefings in Bioinformatics*. 2021; 22(5):bbab048. <https://doi.org/10.1093/bib/bbab048> PMID: 33822890
10. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Molecular systems biology*. 2011; 7(1):496. <https://doi.org/10.1038/msb.2011.26> PMID: 21654673
11. Keum J, Nam H. SELF-BLM: Prediction of drug-target interactions via self-training SVM. *PLOS ONE*. 2017; 12(2):1–16. <https://doi.org/10.1371/journal.pone.0171839> PMID: 28192537
12. Shi H, Liu S, Chen J, Li X, Ma Q, Yu B. Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics*. 2019; 111(6):1839–1852. <https://doi.org/10.1016/j.ygeno.2018.12.007> PMID: 30550813
13. Jarada TN, Rokne JG, Alhadj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *Journal of cheminformatics*. 2020; 12(1):1–23. <https://doi.org/10.1186/s13321-020-00450-7>
14. Pranckevičius T, Marcinkevičius V. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*. 2017; 5(2):221.

15. Playe B, Stoven V. Evaluation of deep and shallow learning methods in chemogenomics for the prediction of drugs specificity. *Journal of cheminformatics*. 2020; 12(1):11. <https://doi.org/10.1186/s13321-020-0413-0> PMID: 33431042
16. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature*. 2020; 577(7792):706–710. <https://doi.org/10.1038/s41586-019-1923-7> PMID: 31942072
17. Luo Y, Zhao X, Zhou J, Yang J, Zhang Y, Kuang W, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nature communications*. 2017; 8(1):1–13. <https://doi.org/10.1038/s41467-017-00680-8> PMID: 28924171
18. Wan F, Hong L, Xiao A, Jiang T, Zeng J. NeoDTI: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics*. 2019; 35(1):104–111. <https://doi.org/10.1093/bioinformatics/bty543> PMID: 30561548
19. Soh J, Park S, Lee H. HIDTI: integration of heterogeneous information to predict drug-target interactions. *Scientific reports*. 2022; 12(1):1–12. <https://doi.org/10.1038/s41598-022-07608-3> PMID: 35260608
20. Huang K, Xiao C, Glass LM, Sun J. MolTrans: Molecular Interaction Transformer for drug-target interaction prediction. *Bioinformatics*. 2021; 37(6):830–836. <https://doi.org/10.1093/bioinformatics/btaa880> PMID: 33070179
21. Kalakoti Y, Yadav S, Sundar D. TransDTI: Transformer-Based Language Models for Estimating DTIs and Building a Drug Recommendation Workflow. *ACS Omega*. 2022; 7(3):2706–2717. <https://doi.org/10.1021/acsomega.1c05203> PMID: 35097268
22. Sadegh S, Matschinske J, Blumenthal DB, Galindez G, Kacprowski T, List M, et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. *Nature communications*. 2020; 11(1):1–9. <https://doi.org/10.1038/s41467-020-17189-2> PMID: 32665542
23. Beck BR, Shin B, Choi Y, Park S, Kang K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational and structural biotechnology journal*. 2020; 18:784–790. <https://doi.org/10.1016/j.csbj.2020.03.025> PMID: 32280433
24. Zeng X, Song X, Ma T, Pan X, Zhou Y, Hou Y, et al. Repurpose open data to discover therapeutics for COVID-19 using deep learning. *Journal of proteome research*. 2020; 19(11):4624–4636. <https://doi.org/10.1021/acs.jproteome.0c00316> PMID: 32654489
25. Tian X, Shen L, Gao P, Huang L, Liu G, Zhou L, et al. Discovery of Potential Therapeutic Drugs for COVID-19 Through Logistic Matrix Factorization With Kernel Diffusion. *Frontiers in microbiology*. 2022; 13. <https://doi.org/10.3389/fmicb.2022.740382> PMID: 35295301
26. Shen L, Liu F, Huang L, Liu G, Zhou L, Peng L. VDA-RWLRLS: An anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares. *Computers in biology and medicine*. 2022; 140:105119. <https://doi.org/10.1016/j.combiomed.2021.105119> PMID: 34902608
27. Hashemi SM, Zabihian A, Hooshmand M, Gharaghani S. DRaW: prediction of COVID-19 antivirals by deep learning—an objection on using matrix factorization. *BMC bioinformatics*. 2023; 24(1):52. <https://doi.org/10.1186/s12859-023-05181-8> PMID: 36793010
28. White NJ, Pongtavornpinyo W, Maude RJ, Saralamba S, Aguas R, Stepniewska K, et al. Hyperparasitaemia and low dosing are an important source of anti-malarial drug resistance. *Malaria journal*. 2009; 8:1–18. <https://doi.org/10.1186/1475-2875-8-253> PMID: 19906307
29. Ahmad I, Huang L, Hao H, Sanders P, Yuan Z, et al. Application of PK/PD modeling in veterinary field: dose optimization and drug resistance prediction. *BioMed research international*. 2016; 2016. <https://doi.org/10.1155/2016/5465678> PMID: 26989688
30. Mokhtari RB, Homayouni TS, Baluch N, Morgatskaya E, Kumar S, Das B, et al. Combination therapy in combating cancer. *Oncotarget*. 2017; 8(23):38022. <https://doi.org/10.18632/oncotarget.16723>
31. Cokol M, Chua HN, Tasan M, Mutlu B, Weinstein ZB, Suzuki Y, et al. Systematic exploration of synergistic drug pairs. *Molecular systems biology*. 2011; 7(1):544. <https://doi.org/10.1038/msb.2011.71> PMID: 22068327
32. Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination therapy in the quest for clinical efficacy. *Nature chemical biology*. 2006; 2(9):458–466. <https://doi.org/10.1038/nchembio817> PMID: 16921358
33. Abdulla A, Wang B, Qian F, Kee T, Blasiak A, Ong YH, et al. Project IDentif. AI: harnessing artificial intelligence to rapidly optimize combination therapy development for infectious disease intervention. *Advanced therapeutics*. 2020; 3(7):2000034. <https://doi.org/10.1002/adtp.202000034> PMID: 32838027

34. O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, et al. An unbiased oncology compound screen to identify novel combination strategies. *Molecular cancer therapeutics*. 2016; 15(6):1155–1162. <https://doi.org/10.1158/1535-7163.MCT-15-0843> PMID: 26983881
35. Xia F, Shukla M, Brettin T, Garcia-Cardona C, Cohn J, Allen JE, et al. Predicting tumor cell line response to drug pairs with deep learning. *BMC bioinformatics*. 2018; 19:71–79. <https://doi.org/10.1186/s12859-018-2509-3> PMID: 30577754
36. Preuer K, Lewis RP, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*. 2018; 34(9):1538–1546. <https://doi.org/10.1093/bioinformatics/btx806> PMID: 29253077
37. Zhang H, Feng J, Zeng A, Payne P, Li F. Predicting tumor cell response to synergistic drug combinations using a novel simplified deep learning model. In: *AMIA Annual Symposium Proceedings*. vol. 2020. American Medical Informatics Association; 2020. p. 1364.
38. Zhang T, Zhang L, Payne PR, Li F. Synergistic drug combination prediction by integrating multiomics data in deep learning models. *Translational bioinformatics for therapeutic development*. 2021; p. 223–238. [https://doi.org/10.1007/978-1-0716-0849-4\\_12](https://doi.org/10.1007/978-1-0716-0849-4_12) PMID: 32926369
39. Holbeck SL, Camalier R, Crowell JA, Govindharajulu JP, Hollingshead M, Anderson LW, et al. The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer research*. 2017; 77(13):3564–3576. <https://doi.org/10.1158/0008-5472.CAN-17-0489> PMID: 28446463
40. Zagidullin B, Aldahdooh J, Zheng S, Wang W, Wang Y, Saad J, et al. DrugComb: an integrative cancer drug combination data portal. *Nucleic acids research*. 2019; 47(W1):W43–W51. <https://doi.org/10.1093/nar/gkz337> PMID: 31066443
41. Kuru HI, Tastan O, Cicek AE. MatchMaker: a deep learning framework for drug synergy prediction. *IEEE/ACM transactions on computational biology and bioinformatics*. 2021; 19(4):2334–2344. <https://doi.org/10.1109/TCBB.2021.3086702>
42. Julkunen H, Cichonska A, Gautam P, Szedmak S, Douat J, Pahikkala T, et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature communications*. 2020; 11(1):6136. <https://doi.org/10.1038/s41467-020-19950-z> PMID: 33262326
43. Meng F, Li F, Liu JX, Shang J, Liu X, Li Y. NEXGB: a network embedding framework for anticancer drug combination prediction. *International Journal of Molecular Sciences*. 2022; 23(17):9838. <https://doi.org/10.3390/ijms23179838> PMID: 36077236
44. Tan X, Hu L, Luquette LJ III, Gao G, Liu Y, Qu H, et al. Systematic identification of synergistic drug pairs targeting HIV. *Nature biotechnology*. 2012; 30(11):1125–1130. <https://doi.org/10.1038/nbt.2391> PMID: 23064238
45. Jin W, Stokes JM, Eastman RT, Itkin Z, Zakharov AV, Collins JJ, et al. Deep learning identifies synergistic drug combinations for treating COVID-19. *Proceedings of the National Academy of Sciences*. 2021; 118(39):e2105070118. <https://doi.org/10.1073/pnas.2105070118> PMID: 34526388
46. Myhre V. Drug Combinations for Treatment of Emerging and Re-emerging Viral Infections; 2022.
47. Ianevski A, Yao R, Simonsen RM, Myhre V, Ravlo E, Kaynova GD, et al. Mono-and combinational drug therapies for global viral pandemic preparedness. *Iscience*. 2022; 25(4). <https://doi.org/10.1016/j.isci.2022.104112> PMID: 35402870
48. Ianevski A, Simonsen RM, Myhre V, Tenson T, Oksenyich V, Bjørås M, et al. DrugVirus. info 2.0: an integrative data portal for broad-spectrum antivirals (BSA) and BSA-containing drug combinations (BCCs). *Nucleic acids research*. 2022; 50(W1):W272–W275. <https://doi.org/10.1093/nar/gkac348> PMID: 35610052
49. Gorander S, Svennerholm B, Liljeqvist JA. Secreted portion of glycoprotein G of herpes simplex virus type 2 is a novel antigen for type-discriminating serology. *Journal of Clinical Microbiology*. 2003; 41(8):3681–3686. <https://doi.org/10.1128/JCM.41.8.3681-3686.2003> PMID: 12904375
50. Schinazi RF, Nahmias A. Different in vitro effects of dual combinations of anti-herpes simplex virus compounds. *The American Journal of Medicine*. 1982; 73(1):40–48. [https://doi.org/10.1016/0002-9343\(82\)90061-4](https://doi.org/10.1016/0002-9343(82)90061-4) PMID: 6285731
51. ZG GA, Tanir G, Kaman A, et al. Acyclovir Unresponsive Herpes Simplex Encephalitis in a child successfully treated with the addition of Foscarnet: Case report. *Archivos Argentinos de Pediatría*. 2019; 117(1):e47–e51.
52. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*. 2005; 33(suppl\_1):D501–D504. <https://doi.org/10.1093/nar/gki025> PMID: 15608248

53. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*. 2008; 36(suppl\_1):D901–D906. <https://doi.org/10.1093/nar/gkm958> PMID: 18048412
54. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*. 1988; 28(1):31–36. <https://doi.org/10.1021/ci00057a005>
55. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of cheminformatics*. 2015; 7:1–13. <https://doi.org/10.1186/s13321-015-0069-3> PMID: 26052348
56. Rogers DJ, Tanimoto TT. A Computer Program for Classifying Plants: The computer is programmed to simulate the taxonomic process of comparing each case with every other case. *Science*. 1960; 132(3434):1115–1118. <https://doi.org/10.1126/science.132.3434.1115> PMID: 17790723
57. Mullan L. Pairwise sequence alignment—it's all about us! *Briefings in bioinformatics*. 2006; 7(1):113–115. <https://doi.org/10.1093/bib/bbk008> PMID: 16761368
58. Smith TF, Waterman MS, et al. Identification of common molecular subsequences. *Journal of molecular biology*. 1981; 147(1):195–197. [https://doi.org/10.1016/0022-2836\(77\)90031-6](https://doi.org/10.1016/0022-2836(77)90031-6) PMID: 7265238
59. Durbin R, Eddy SR, Krogh A, Mitchison G. *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press; 1998.
60. Zabihian A, Sayyad FZ, Hashemi SM, Shami Tanha R, Hooshmand M, Gharaghani S. DEDT1 versus IEDT1: efficient and predictive models of drug-target interactions. *Scientific Reports*. 2023; 13(1):9238. <https://doi.org/10.1038/s41598-023-36438-0> PMID: 37286613
61. Evgeniou T, Pontil M. Support vector machines: Theory and applications. In: *Advanced course on artificial intelligence*. Springer; 1999. p. 249–257.
62. Xie X, Yuan MJ, Bai X, Gao W, Zhou ZH. On the Gini-impurity Preservation For Privacy Random Forests. In: Oh A, Neumann T, Globerson A, Saenko K, Hardt M, Levine S, editors. *Advances in Neural Information Processing Systems*. vol. 36. Curran Associates, Inc.; 2023. p. 45055–45082.
63. Bottou L, Chapelle O, DeCoste D, Weston J. In: *Support Vector Machine Solvers*; 2007. p. 1–27.
64. Breiman L. Random forests. *Machine learning*. 2001; 45:5–32. <https://doi.org/10.1023/A:1017934522171>
65. Kim TK. T test as a parametric statistic. *Korean journal of anesthesiology*. 2015; 68(6):540. <https://doi.org/10.4097/kjae.2015.68.6.540> PMID: 26634076
66. Wagstaff AJ, Faulds D, Goa KL. Aciclovir: a reappraisal of its antiviral activity, pharmacokinetic properties and therapeutic efficacy. *Drugs*. 1994; 47:153–205. <https://doi.org/10.2165/00003495-199447010-00009> PMID: 7510619
67. Avery RK, Alain S, Alexander BD, Blumberg EA, Chemaly RF, Cordonnier C, et al. Maribavir for refractory cytomegalovirus infections with or without resistance post-transplant: results from a phase 3 randomized clinical trial. *Clinical Infectious Diseases*. 2022; 75(4):690–701. <https://doi.org/10.1093/cid/ciab988> PMID: 34864943
68. Shishkov S, Pancheva S. The synergistic antiviral effect of acyclovir and ribavirin against the herpes simplex type-1 virus and the pseudorabies virus in vitro. *Acta Microbiologica Bulgarica*. 1990; 25:69–75. PMID: 2166426
69. Mugwanya K, Baeten JM, Mugo NR, Irungu E, Ngunjiri K, Celum C. High-dose valacyclovir HSV-2 suppression results in greater reduction in plasma HIV-1 levels compared with standard dose acyclovir among HIV-1/HSV-2 coinfecting persons: a randomized, crossover trial. *The Journal of infectious diseases*. 2011; 204(12):1912–1917. <https://doi.org/10.1093/infdis/jir649> PMID: 21998479
70. Palmieri G, Ambrosi G, Ferraro G, Agrati A, Palazzini E. Clinical and immunological evaluation of oral ribavirin administration in recurrent herpes simplex infections. *Journal of international medical research*. 1987; 15(5):264–275. <https://doi.org/10.1177/030006058701500502> PMID: 3315775
71. Johnston C, Saracino M, Kuntz S, Magaret A, Selke S, Huang MI, et al. Standard-dose and high-dose daily antiviral therapy for short episodes of genital HSV-2 reactivation: three randomised, open-label, cross-over trials. *The Lancet*. 2012; 379(9816):641–647. [https://doi.org/10.1016/S0140-6736\(11\)61750-9](https://doi.org/10.1016/S0140-6736(11)61750-9) PMID: 22225814
72. Sharma D, Sharma S, Akojwar N, Dondulkar A, Yenorkar N, Pandita D, et al. An insight into current treatment strategies, their limitations, and ongoing developments in vaccine technologies against herpes simplex infections. *Vaccines*. 2023; 11(2):206. <https://doi.org/10.3390/vaccines11020206> PMID: 36851084
73. Papanicolaou GA, Lee YJ, Young JW, Seshan SV, Boruchov AM, Chittick G, et al. Brincidofovir for polyomavirus-associated nephropathy after allogeneic hematopoietic stem cell transplantation. *American*

- Journal of Kidney Diseases. 2015; 65(5):780–784. <https://doi.org/10.1053/j.ajkd.2014.11.020> PMID: 25600489
74. Samiei RN, Ebrahimi S, Fani M, Ghafari S, Makvandi M, Khodadad N, et al. In vitro effect of some nucleoside reverse transcriptase inhibitors against HSV-1 replication. *European Review for Medical and Pharmacological Sciences*. 2020; 24(3):1454–1459.
  75. Chemaly RF, Hill JA, Voigt S, Peggs KS. In vitro comparison of currently available and investigational antiviral agents against pathogenic human double-stranded DNA viruses: A systematic literature review. *Antiviral Research*. 2019; 163:50–58. <https://doi.org/10.1016/j.antiviral.2019.01.008> PMID: 30677427
  76. Carmine A, Brogden R, Heel R, Speight T, Avery G. Trifluridine: a review of its antiviral activity and therapeutic use in the topical treatment of viral eye infections. *Drugs*. 1982; 23:329–353. <https://doi.org/10.2165/00003495-198223050-00001> PMID: 6284470
  77. Lee YJ, Neofytos D, Kim SJ, Cheteyan L, Huang YT, Papadopoulos EB, et al. Efficacy of brincidofovir as prophylaxis against HSV and VZV in hematopoietic cell transplant recipients. *Transplant Infectious Disease*. 2018; 20(6):e12977. <https://doi.org/10.1111/tid.12977> PMID: 30120866
  78. Kearn SJ, Chapman TM, Figgitt DP. Brivudin (bromovinyl deoxyuridine). *Drugs*. 2004; 64:2091–2097. <https://doi.org/10.2165/00003495-200464180-00011> PMID: 15341504
  79. De Clercq E, Holý A, Rosenberg I. Efficacy of phosphonylmethoxyalkyl derivatives of adenine in experimental herpes simplex virus and vaccinia virus infections in vivo. *Antimicrobial agents and chemotherapy*. 1989; 33(2):185–191. <https://doi.org/10.1128/aac.33.2.185> PMID: 2719463
  80. Han J, Lee MK, Jang Y, Cho WJ, Kim M. Repurposing of cyclophilin A inhibitors as broad-spectrum antiviral agents. *Drug Discovery Today*. 2022; 27(7):1895–1912. <https://doi.org/10.1016/j.drudis.2022.05.016> PMID: 35609743
  81. Martin P, Jensen DM. Ribavirin in the treatment of chronic hepatitis C. *Journal of gastroenterology and hepatology*. 2008; 23(6):844–855. <https://doi.org/10.1111/j.1440-1746.2008.05398.x> PMID: 18565019
  82. Pagano JS, Whitehurst CB, Andrei G. Antiviral drugs for EBV. *Cancers*. 2018; 10(6):197. <https://doi.org/10.3390/cancers10060197> PMID: 29899236