RESEARCH ARTICLE

Antivirals for monkeypox virus: Proposing an effective machine/deep learning framework

Morteza Hashemi[1], Arash Zabihian[2], Masih Hajsaeedi[1], Mohsen Hooshmand[1]*

1 Department of Computer Science, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran,
2 Department of QA, Kimia Zist Parsian Pharmaceutical Company, Zanjan, Iran

☉ These authors contributed equally to this work.
* moshsen.hooshmand@iasbs.ac.ir
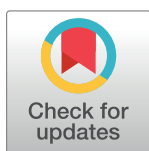
Abstract

Monkeypox (MPXV) is one of the infectious viruses which caused morbidity and mortality problems in these years. Despite its danger to public health, there is no approved drug to stand and handle MPXV. On the other hand, drug repurposing is a promising screening method for the low-cost introduction of approved drugs for emerging diseases and viruses which utilizes computational methods. Therefore, drug repurposing is a promising approach to suggesting approved drugs for the MPXV. This paper proposes a computational framework for MPXV antiviral prediction. To do this, we have generated a new virus-antiviral dataset. Moreover, we applied several machine learning and one deep learning method for virus-antiviral prediction. The suggested drugs by the learning methods have been investigated using docking studies. The target protein structure is modeled using homology modeling and, then, refined and validated. To the best of our knowledge, this work is the first work to study deep learning methods for the prediction of MPXV antivirals. The screening results confirm that Tilorone, Valacyclovir, Ribavirin, Favipiravir, and Baloxavir marboxil are effective drugs for MPXV treatment.

1 Introduction

Monkeypox (MPXV) is a viral zoonosis disease resulting from an enveloped double-stranded DNA virus that belongs to the Poxviridae family and causes an international public health emergency [1]. The initial occurrence of monkeypox in animals and humans was reported in 1958 and 1970, respectively [2]. A global epidemic of MPXV in May 2022 resulted in more than 85,000 cases across regions with no history of transmission [3]. Also, over 30,000 cases of MPXV had been reported in the US as of March 2023 [4]. The virus transmits through contact with skin lesions, respiratory droplets, body fluids, and fomites [2]. Although several antiviral drugs are suggested for treatment against MPXV, the Food and Drug Administration (FDA) has not approved any specific drugs for human monkeypox [5]. On the other hand, drug repurposing is an efficient approach for accelerating the discovery of novel treatments by providing new indications for approved drugs [6]. Currently, modern computational methods are making a significant impact on drug repurposing for the treatment of viral infectious diseases.

For example, drug repurposing has played an important role in research and proposing novel drugs for Covid 19. Tang et al. suggested a non-negative matrix factorization method for suggesting antivirals against SARS-CoV-2 [7]. However, the matrix factorization methods suffer from two problems, lack of generalizability and data leakage [8]. Beck et al. SARS-CoV-2 [9] using deep model. Zeng et al. have the same mission to suggest new drugs for SARS-CoV-2 using deep models [10]. Other studies suggested and showed the use of machine learning and a spectrum of deep learning methods have a higher performance and more trustful results for drug-target prediction of proteins, SARS-CoV-2, and other viruses [9, 11–19]. It is worth mentioning that although deep learning algorithms are the winners in the field of drug repurposing, the machine learning methods can reach a similar performance in most cases with a lower amount of resource consumption [20]. Therefore, this paper tends to check and propose methods from both machine learning and deep learning for monkeypox antiviral prediction.

This study is centered on using computational drug repurposing to identify a novel candidate for targeting monkeypox among approved antiviral drugs. To do this, we have created a virus-antiviral dataset and gathered similar viruses to the MPXV. In addition to that, we apply proper *similarity measures* to create effective features for viruses and antivirals. We, as mentioned above, utilize machine learning and deep learning methods for the prediction. Then, based on the results of the learning methods, we chose the most promising drugs and performed a docking study to approve the results of the machine learning methods. The docking study confirms that the results of the learning methods are promising and proper voices for MPXV treatment. To the best of our knowledge, this is the first work in applying results of deep and machine learning for drug recommendation for the MPXV virus.

The contribution of the paper is 5-fold.

- Generating a virus-antiviral dataset for MPXV prediction.

- Applying machine learning algorithms, i.e., decision tree, SVM, and random forest algorithms.

- Proposing and applying a CNN model to train the prediction model.

- Proposing five approved drugs for MPXV treatment, i.e., *Tilorone, Valacyclovir, Ribavirin, Favipiravir, and Baloxavir marboxil.*

- Validating the proposed drugs by docking study.

The results are promising and the docking study supports the proposed MPXV antivirals with high scores. The structure of the paper is as follows. Section 2 explains the dataset generation and the learning methods for the MPXV prediction. Additionally, it introduces the homology modeling and docking studies applied to the proposed antivirals. Section 3 compares the performance of learning algorithms, introduces the proposed drugs, and reports homology modeling and docking studies on the proposed drugs. Section 4 finalizes the paper.

2 Materials and methods

This paper proposes a prediction framework for monkeypox virus-antiviral interaction that uses computational learning methods and we call it *MPXV-Pred*. This section provides the steps of MPXV-Pred. The whole process contains five main steps, i.e., data collection, dataset generation, the learning phase using machine learning and deep learning methods, the voting procedure to propose the most effective antiviral, and finally the docking step. Fig 1 shows the pipeline of the proposed method. The following sections describe these steps.
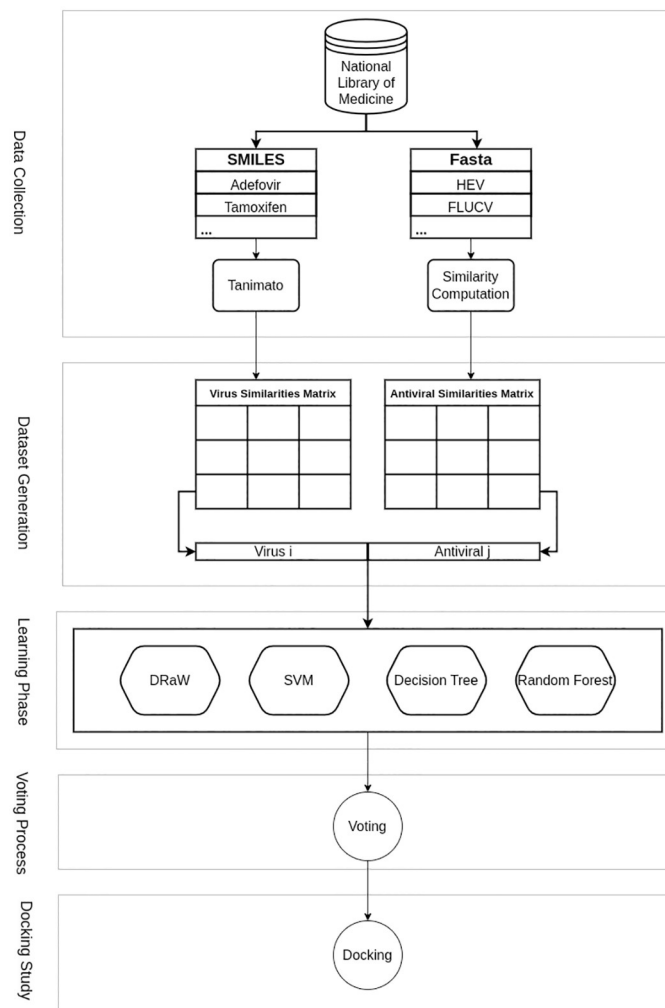
Fig 1. MPXV-Pred framework. It aims to predict novel antivirals for the MPXV. The pipeline contains five steps data collection, dataset generation, machine and deep learning methods, the voting procedure to announce the promising antiviral, and docking study. I) Data collection: the raw and primary information on viruses and antivirals have been collected from the *NIH*. The approved virus-antiviral interactions have also been provided from DrugVirus.info 2 [21]. II) Dataset generation: the output of this step is the generated dataset including the similarity matrix of viruses, similarity matrices of antivirals, and virus-antiviral interaction matrix. The virus similarity matrix has been computed using *sequence alignment* and calculating the similarity percent. Additionally, the Tanimoto score has been used to compute the antiviral similarity matrix. III) The learning phase: this phase employs four different machine learning methods, i.e. CNN, SVM, Decision tree, and Random forest. IV) The voting step uses the prediction results of the ML methods of the previous step and votes among them to suggest the efficient antiviral for MPXV. V) The Docking study investigates and approves the predicted antivirals for MPXV treatment.

https://doi.org/10.1371/journal.pone.0299342.g001

2.1 Data collection

There exists no proper dataset available to contain information on viruses and antivirals that have common properties with the MPXV virus. Therefore, in the first step, it is necessary to prepare a new corresponding dataset. To do that we collected raw data of viruses and antivirals, e.g., smallpox from "The National Library of Medicine" databases [22]. The raw data includes the Fasta format of virus sequences of 100 viruses, the simplified molecular-input line-entry system (SMILES) format of 198 drugs, and the interaction between them. The

collected dataset has 96 approved virus-antiviral interactions, therefore, its sparsity reaches 99.52%.

2.2 Similarity computation and generating the dataset

Using sequences and SMILES, the similarities between each pair of viruses in addition to each pair of drugs were computed correspondingly. Tanimoto [23] is a popular approach to calculating the similarities between drugs. Thus, we applied this algorithm to the collected data to create a similarity matrix of drugs. To calculate the similarity among the viruses we align them in the first step. Therefore, we use the PairWiseAligner tool from the BioPython library [24]. The alignments are performed locally with the Smith-Waterman algorithm [25] and the NUC44 substitution matrix is used to score the alignments. The gap score is set to −10 and the gap extend score is set to −0.5. These two scores are default scores from emboss [26] tools. Fig 1a shows the procedure of collecting the data and inverting it to similarity matrices.

2.3 Learning phase

The learning phase aims to train a model to predict the most effective antivirals for the MPXV. To do this, MPXV-Pred utilizes three machine learning methods, i.e., decision tree, SVM, and random forest [27]. The MPXV-Pred uses the similarity vectors of each virus and antiviral as their feature vector. Therefore, the input vectors of each learning method are the concatenation of feature vectors corresponding to each virus-antiviral pair. The interaction matrix plays the role of the virus-antiviral label. Therefore, we can formulate the problem as follows.

$$\hat{x}_{ij} = LeraningMethod(\mathbf{v}_i \| \mathbf{av}_j) \tag{1}$$

Where, $\mathbf{v}_i$ and $\mathbf{av}_j$ represent the features vectors of i-th virus and j-th antiviral. The $\hat{x}_{ij}$ shows the prediction result of the interaction between the i-th virus and j-th antiviral. Moreover, it uses a convolution-based deep learning model which we call our proposal as **D**rug **R**epurposing-analytic **W**ay (DRaW) [8] to complete the same mission. Fig 2 shows the DRaW framework. DRaW consists of three convolution layers in total. The first layer is the convolution layer with 128 filters with size 3 and batch normalization is applied to improve training stability. And dropout layer is used to randomly drop 50% of the unit during training to reduce overfitting. for the second and third layers, we've used the same pattern but with different sizes of filters. And lastly, a dense layer with 128 units is applied for classification.

DRaW's input vector is the virus-antiviral pair's representation from the concatenation of $\mathbf{v}_i$ and $\mathbf{av}_j$, or $\mathbf{e}_{i,j} = \mathbf{v}_i \| \mathbf{av}_j$.

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} [x_{ij} - DRaW(\mathbf{e}_{i,j}, opt:Adam, loss:Cross-Ent,)] \tag{2}$$

2.4 Voting process

The MPXV-Pred investigates the suggested antivirals of whole methods and reports those that happen in *at least two of the learning methods*. This may help to suggest the most effective potential antivirals for MPXV treatment.

2.5 Homology modeling, refinement and validation

Since the outbreak of the MPXV global epidemic, *tecovirimat* is the only currently available therapeutic agent that has been suggested for use in Europe under "exceptional circumstances"
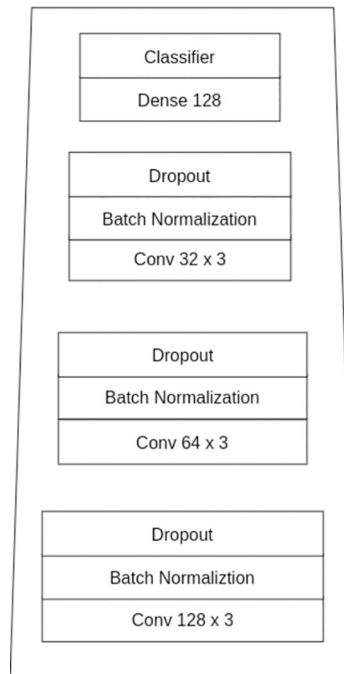
Fig 2. DRaW architecture. It is a CNN-based deep learning method.

https://doi.org/10.1371/journal.pone.0299342.g002

by the European Medicines Agency [28]. In other words, there is no approved drug for MPXV. According to the literature, the envelope receptor F-13 is a potential target for tecovirimat. This protein plays a crucial role in the growth and maturation of the monkeypox virus. The mechanism of F-13-inhibition could be a promising treatment for monkeypox. Therefore, we employed molecular docking simulations to investigate the binding capability of drugs predicted by our proposed model to the F-13 protein as a potential and promising target.

The primary challenge arises from the unavailability of the complete crystal structure of the poxvirus F-13. To address this challenge, the UniprotKB [29] entry Q5IXY0, representing the monkeypox envelope protein F-13 with a length of 372 amino acids, was uploaded to Alpha-Fold2 [30] to generate a predicted 3D protein structure. We also determined the predicted local distance difference test (pLDDT) values for the predicted structure. This modeled structure was optimized by energy minimization using the YASARA server [31], which used the YASARA force field for this purpose.

Before conducting molecular docking and assessing the potential binding of predicted drugs to the F-13 protein, the constructed structure of this protein needs to be refined and validated. The predicted tertiary structure of monkeypox envelope protein F-13 was refined using the Galaxy Refine service [32]. Furthermore, to validate the structure of the refined model, the energy of residue-residue interaction using a distance-based pair potential and the energy was transformed to a score (called z-score) were analyzed using the ProSA-web server [33]. The Stereochemical quality of the modeled F-13 protein was calculated using the Errat [34]. Additionally, Ramachandran plot analysis was employed to assess the consistency of the constructed model [35].

2.6 Molecular docking

Structure-based molecular docking is a computational technique to evaluate how a ligand interacts with a target and addresses three main objectives: virtual screening, posture

prediction, and estimation of binding affinity [36]. Among the top ten drugs predicted by DRaW, we investigated instances that were also included in the predicted drug lists of three other models. To elucidate further, according to the data from Table 4, Tilorone, Valacyclovir, Ribavirin, Baloxavir, and Favipiravir are the drugs that appear most frequently across the predicted lists of all models. Therefore, we decided to do molecular docking studies employing these drugs.

The homology-modeled protein structure was prepared using the Autodock tools (ADT). This procedure involved the incorporation of polar hydrogens and Kollman charges. The 3D-SDF structures of all five candidate drugs were retrieved from NCBI PubChem [37]. The preparation of ligands involved the addition of polar hydrogens and Gasteiger charges. Additionally, root detection and selection of torsions from the torsion tree were conducted to rotate all the rotatable bonds [38]. We identified the F-13 binding site based on the methodology outlined by Li et al. [39]. According to their study, the grid box was fixed at (−6.27702) * (−2.48567) * (−8.66086) in XYZ-coordinates, with a radius of 8.9, and the spacing for the grid box was kept at 0.375 Å. Finally, docking studies were performed by AutoDock 4.2 using the Lamarckian genetic algorithm.

2.7 Complexity analysis

We provide a complexity analysis of all four learning methods. Let m be the number of viruses and n be the number of antivirals. The feature vector is the concatenation of virus and antiviral similarity vectors. We assume its size is d.

Decision tree. Time complexity of the decision tree is $O(m * n * log(m * n) * d)$ where $m * n$ is the number of whole virus-antiviral pairs in the training phase and d is the dimensionality of the data. The runtime complexity is $O(depth)$.

SVM. Train complexity of Support Vector Machine is $O((m * n)^2)$. The runtime complexity is $O(k * d)$ where k is the number of support vectors and d is the dimensionality of data.

Random forest. The time complexity of a random forest depends on the number of trees in the forest and the depth of each tree. So the training time complexity is $O(m * n * log(m * n) * d * t)$. Where t is the number of decision trees. The runtime time complexity is $O(depth * t)$.

Deep learning. We assume the number of epochs in the training phase is e and each epoch time interval is equal to T_t. Then, the training phase time complexity is $O(m * n * d * e * T_t)$. Its runtime complexity is $O(d*T)$, where T is the test phase running time.

3 Results

This section provides the result of the MPXV-Pred using the proposed methods with early stpping. The reported results are based on 10-fold cross-validation.

Table 1 reports the specifications of the SVM, random forest, decision tree, and DRaW.

3.1 Performance evaluations

We evaluate the results of the methods using the following metrics. These metrics are useful for binary classification [40]. The MPXV-Pred dataset is imbalanced, therefore, metrics such as F1-score and AUPR are significant in interpreting the results [41].

$$Specificity = \frac{TN}{TN + FP}, \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{4}$$

Table 1. Learning models' parameters.

ML Algorithms	Parameters	Parameters' Values
DRaW	optimizer	adam
	loss	BinaryCrossEntropy
	epoch	500
	batch size	128
SVM	kernel	rbf
	degree	3
	gamma	scale
	tol	1e-3
Random Forest	criterion	gini
	no. estimators	100sqrt
	max features	sqrt
Decision Tree	criterion	gini
	splitter	best

https://doi.org/10.1371/journal.pone.0299342.t001

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall}, \tag{7}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \tag{8}$$

$$AUC - ROC = \sum_{i=1}^{n} TPR_i \Delta FPR_i, \tag{9}$$

$$AUPR = \sum_{i=1}^{n} precision_i \ \Delta \ recall_i. \tag{10}$$

3.2 Prediction results

Tables 2 and 3 show two different handling of the results. As mentioned above, we applied three machine learning methods and DRaW, a deep convolutional learning method for drug

Table 2. Comparing the performance of machine learning models by different measures. In this experiment, the threshold limit in the classification problem is set equal to 0.5.

Method	Sp	Re(Se)	Pr	Ac	F1	MCC	AUC-ROC	AUPR
Decision Tree	0.9982 ± 0.0006	0.3337 ± 0.0324	0.4953 ± 0.0931	0.9949 ± 0.0006	0.3962 ± 0.0424	0.4028 ± 0.0479	0.6660 ± 0.0162	0.4161 ± 0.0536
SVM	0.9996 ± 0.0004	0.0816 ± 0.0387	0.5600 ± 0.2881	0.9951 ± 0.0003	0.1383 ± 0.0620	0.2062 ± 0.0875	0.8593 ± 0.0730	0.2456 ± 0.1060
Random Forest	1.0 ± 0.0	0.1329 ± 0.0283	1.0 ± 0.0	0.9957 ± 0.0004	0.2337 ± 0.0436	0.3622 ± 0.0379	0.9543 ± 0.0196	0.3890 ± 0.0408
DRaW	0.9999 ± 0.0002	0.1143 ± 0.0557	0.9200 ± 0.1789	0.9955 ± 0.0011	0.1966 ± 0.0842	0.3104 ± 0.0716	0.9641 ± 0.0172	0.4114 ± 0.0534

https://doi.org/10.1371/journal.pone.0299342.t002

Table 3. Comparing the performance of machine learning models by different measures. Since the dataset is imbalanced and only about 0.4% of the samples interact, it is regular to define the threshold limit as floating, unlike the previous test. The result of this experiment is shown in this table.

| Method	Sp	Re(Se)	Pr	Ac	F1	MCC	AUC-ROC	AUPR	Threshold
Decision Tree	0.9982 ± 0.0006	0.3337 ± 0.0324	0.4953 ± 0.0931	0.9949 ± 0.0006	0.3962 ± 0.0424	0.4028 ± 0.0479	0.6660 ± 0.0162	0.4161 ± 0.0536	0.5
SVM	0.9969 ± 0.0028	0.3219 ± 0.2042	0.3096 ± 0.2138	0.9936 ± 0.0019	0.2978 ± 0.1840	0.3037 ± 0.1852	0.8593 ± 0.0730	0.2456 ± 0.1060	0.1807 ± 0.2685
Random Forest	0.9982 ± 0.0020	0.3761 ± 0.1305	0.6582 ± 0.3151	0.9951 ± 0.0014	0.4272 ± 0.0437	0.4657 ± 0.0626	0.9543 ± 0.0196	0.3890 ± 0.0408	0.2840 ± 0.1316
DRaW	0.9970 ± 0.0024	0.5181 ± 0.2220	0.5400 ± 0.1899	0.9946 ± 0.0015	0.4711 ± 0.0362	0.4965 ± 0.0377	0.9641 ± 0.0172	0.4114 ± 0.0534	0.2208 ± 0.1297
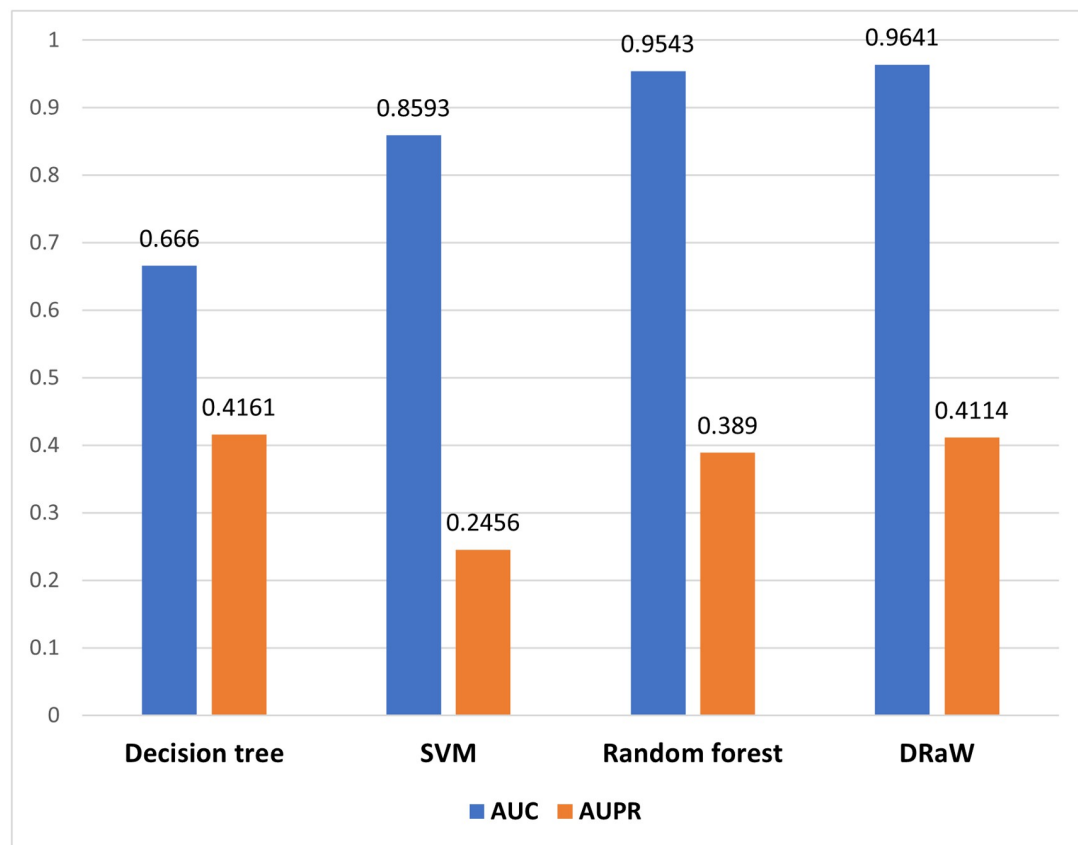
Fig 3. Performance comparison of methods on MPXV-Pred dataset. AUC-ROC and AUPR bar charts.

repurposing, to the dataset. Table 2 shows the result by defining a fixed threshold equal to 0.5. In the majority of metrics DRaW model has the highest score. DRaW has the highest AUC-ROC. Additionally, The highest AUPR score belongs to the decision tree and DRaW. We believe the higher result of the decision tree model in comparison to the random forest and SVM could be due to overfitting [42]. Because decision trees have a high probability of falling into the overfitting problem.

In contrast to the previous calculation of evaluation metrics, we defined a floating threshold to choose the best values of all metrics except AUC-ROC and AUPR (Because these two do not depend on a single threshold). Table 3 shows the results of the floating threshold. The last row shows the threshold of each method. The decision tree model does not need any threshold and, therefore, is left empty. The DRaW has the highest AUC-ROC and the decision tree has the lowest AUC-ROC. The DRaW and decision tree have the highest AUPR. The high value of AUPR for the decision tree, as mentioned before, could be due to overfitting.

Fig 3 shows the bar chart of the AUC-ROC and AUPR of the decision tree, SVM, random forest, and DRaW models. As shown in the figure, random forest and DRaW have the highest AUC-ROC, and the winner is the latter. Additionally, DRaW and decision tree have the highest AUPR.

3.3 Voting results

Table 4 shows the suggested drugs of voting results of whole four methods. As shown in the table, the first rank in all methods belongs to Tilirone. Valacyclovir occupies the second rank

Table 4. Proposed drugs from the voting process.

DRaW	SVM	Random Forest	DecisionTree
Tilorone	Tilorone	Tilorone	Tilorone
Valacyclovir	Amantadine	Favipiravir	Ribavirin
Ribavirin	Baloxavir marboxil	Ribavirin	
Foscarnet	Lamivudine	Itraconazole	
Inosine (Inosine pranobex)	Favipiravir	Baloxavir marboxil	
Baloxavir marboxil	Valacyclovir	Valacyclovir	
Favipiravir	Thymalfasin	Fluvastatin	
Valganciclovir	Rimantadine	Famciclovir	
Lamivudine	Lopinavir	Ilaprazole	
Ganciclovir	Zanamivir	Resveratrol	

https://doi.org/10.1371/journal.pone.0299342.t004

in CNN and has been suggested by three out of four methods. Ribavirin has been suggested by three out of four methods. The same as the latter happened for Favipiravir. Finally, Baloxavir marboxil has been reported by three out of four methods. Notably, the decision tree suggested two drugs and we report those two, i.e., Tilirone and Ribavirin. One important observation is that while the four methods follow different approaches for learning and prediction, they confirm each other and the majority of top-rank antivirals are the same in all of them.

3.4 Results of homology modeling, refinement and validation

We utilized AlphaFold tools to predict the three-dimensional structure of the monkeypox envelope protein F-13, which has a length of 372 amino acids. The evaluation of structure prediction relied on the predicted local distance difference test (pLDDT) score [29]. The highest pLDDT score among the five predicted envelope protein F-13 structures was 91.04. The additional pLDDT scores are shown in Table 5.

Fig 4 Illustrates the homology model confidence score with predicted LDDT and predicted aligned error (Rank 1). Given that a pLDDT score greater than 70 is indicative of a high-quality structure, the results from the pLDDT analysis further confirm the high accuracy of the predicted structures.

The ProSA was used to validate the three-dimensional model of the F-13 protein. This tool employs the z-score and a plot of the residue energies of the input structure as two quality metrics. The z-score of our refined model was −8.09 kcal/mol. As illustrated in Fig 5A, the model structure of F-13 follows the known proteins whose structures have already been determined through X-ray crystallography (light blue) or NMR spectroscopy (dark blue) studies. Also, the energy plot, shown in Fig 5B, depicts the local model's quality by showing energies as a function of amino acid sequence position. As a rule, positive numbers often indicate inaccurate portions of a model. The "overall quality factor" for non-bonded atomic interactions, or ERRAT, is a

Table 5. pLDDT scores.

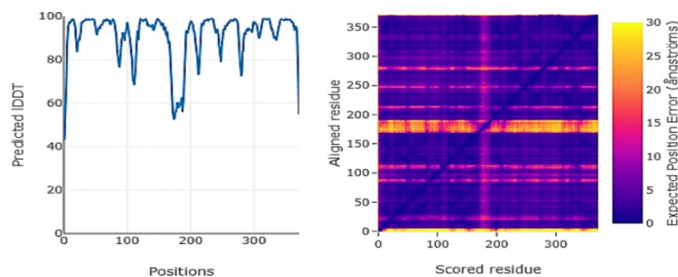| Model Rank	Mean pLDDT	Max PAE	pTM
1	91.04	30.95	0.92
2	90.67	30.77	0.92
3	90.26	30.86	0.92
4	90.21	31.13	0.91
5	89.41	30.81	0.91

https://doi.org/10.1371/journal.pone.0299342.t005

Fig 4. The confidence score of the homology model, along with the predicted LDDT and the predicted aligned error.
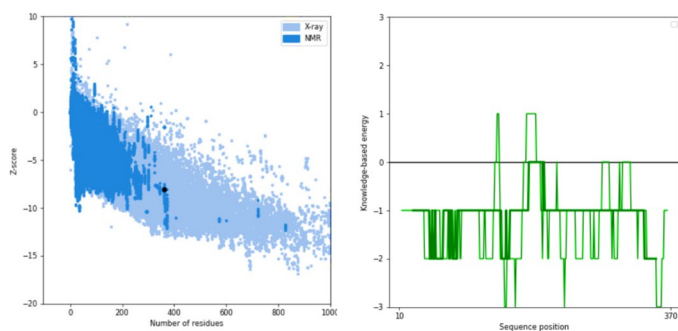
https://doi.org/10.1371/journal.pone.0299342.g004

Fig 5. ProSA analysis. (A) Z-score plot of F-13, the z-scores of F-13 highlighted as a large dot; (B) The energy profile of F-13 protein.

https://doi.org/10.1371/journal.pone.0299342.g005

score that represents the quality of a model. The generally accepted threshold for a high-quality model is above 50 [43]. In this study, the ERRAT score for the modeled F-13 protein was 89.73, shown in Fig 6, indicating that the model quality is significant and acceptable.

In Fig 7, the Ramachandran plot analysis revealed that 87.3% of the residues were in the favored regions (A, B, and L), while 11.8% were in additionally allowed regions (a, b, l, and p). Only 0.3% of the residues were in generously allowed regions, and 0.6% were in the unflavored regions. These results indicate that the generated structures of the F-13 protein are highly valid and reliable.

3.5 Docking result

Molecular docking studies were conducted to assess the potential interactions between the F-13 protein and a set of five chosen antiviral drugs. We chose the best ligand conformations by
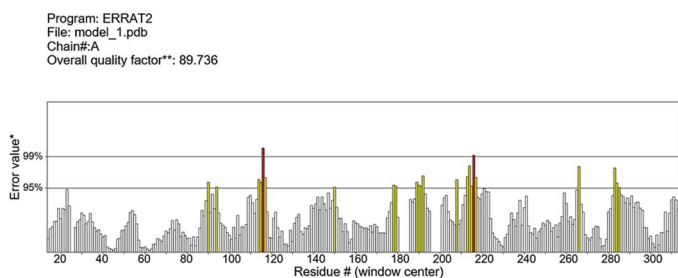

Fig 6. ERRAT result depicting the overall quality factor of modeled protein.

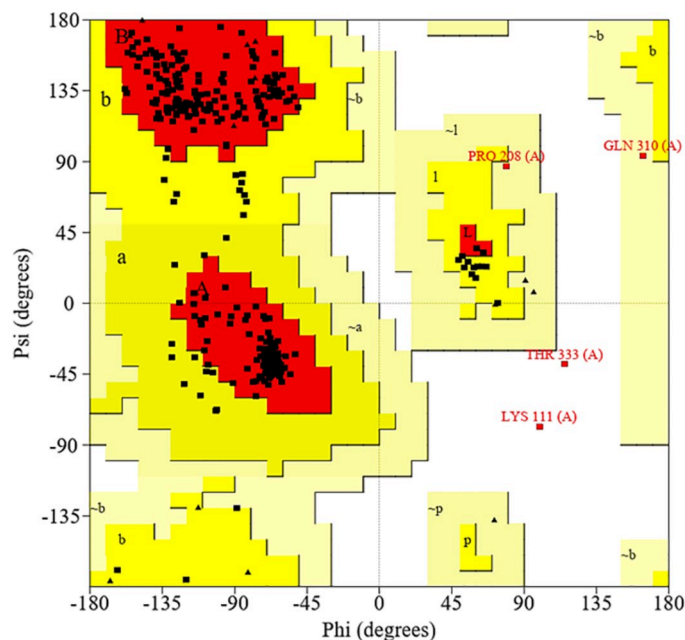https://doi.org/10.1371/journal.pone.0299342.g006

Fig 7. Ramachandran plot for the modeled structure of F-13.

clustering them based on both RMSD and binding affinity [44]. To explore the intermolecular interaction forces, the results of the docking were visualized using Biovia Discovery Studio Visualizer [45]. Table 6 displays the five selected antiviral drugs along with their respective docking scores with the homology-modeled F-13 protein structure. Based on the docking scores, Baloxavir exhibited the most favorable binding energy of −8.32 kcal/mol and formed three hydrogen bonds with LYS 88, ASN 90, and SER 58. Figs 8 to 12 illustrate both the 3D and 2D representations of each drug-F-13 protein complex.

Fig 8 shows Tilorone's 3D and 2D representations. In this figure, the docking results reveal hydrogen bonds with ASP109 and ILE108 residues, as well as van der Waals interactions with several other residues. Additionally, the binding of tilorone to the F-13 protein indicates unfavorable acceptor-acceptor interactions with HIS338.

Fig 9 shows Valacyclovir's 3D and 2D representations. The interaction between Valacyclovir and F-13 encompasses various non-covalent interactions, including hydrogen bonds, pi-alkyl interactions, and van der Waals forces.

Fig 10 shows Ribavirin's 3D and 2D representations. This figure illustrates undesirable donor-donor bonding involving TYR285 and VAL284 in the interaction between ribavirin and F-13. Nevertheless, the overall interaction pattern involves hydrogen bonds with TRP279,

Table 6. Docking scores of selected drugs.

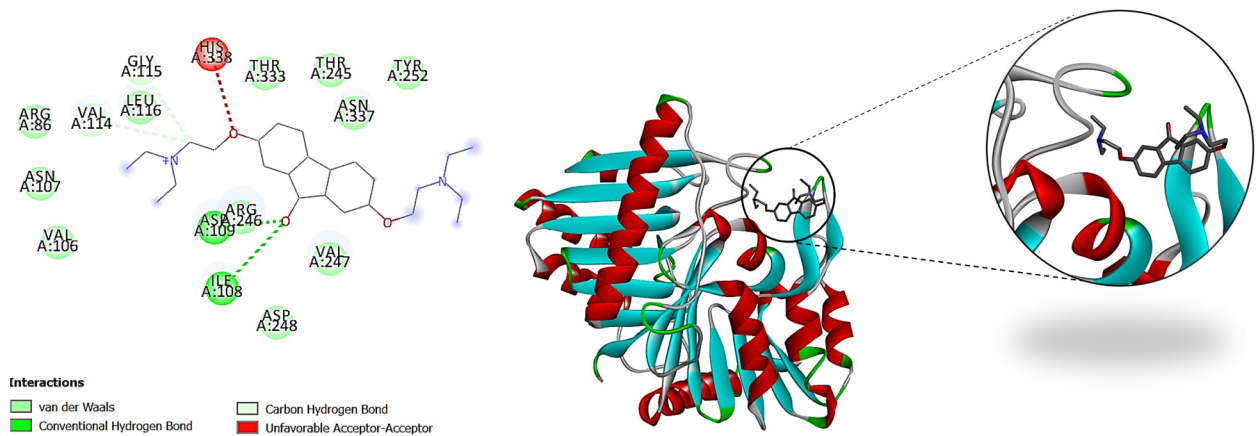| Drug Name	PUBCHEM CID	Molecular Formula	Docking Score (kcal/mol)
Tilorone	5475	$C_{25}H_{34}N_2O_3$	−7.10
Valacyclovir	135398742	$C_{13}H_{20}N_6O_4$	−7.19
Ribavirin	37542	$C_8H_{12}N_4O_5$	−7.40
Baloxavir	124081876	$C_{24}H_{19}F_2N_3O4S$	−8.32
Favipiravir	492405	$C_5H_4FN_3O_2$	−5.13

Fig 8. 2D and 3D representations of the docked pose for the predicted interaction between Tilorone. The green dashed lines represent hydrogen bonds and F-13 protein.
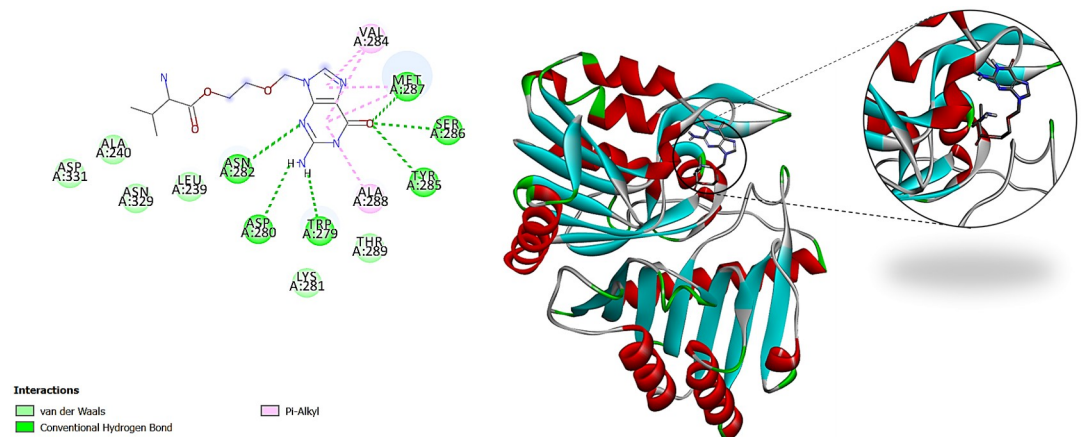
Fig 9. 2D and 3D representations of the docked pose for the predicted interaction between Valacyclovir. The green dashed lines represent hydrogen bonds and F-13 protein.
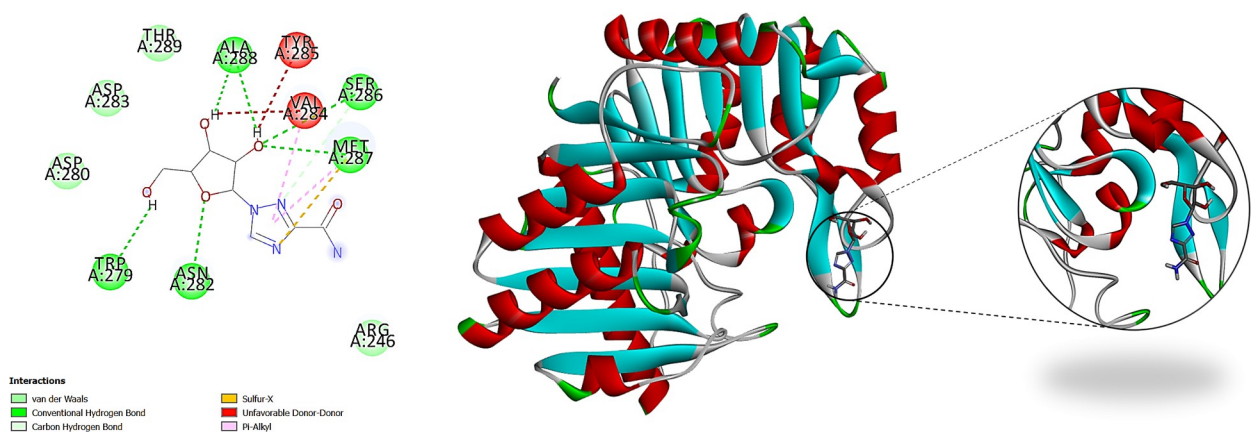
Fig 10. 2D and 3D representations of the docked pose for the predicted interaction between Ribavirin. The green dashed lines represent hydrogen bonds and F-13 protein.
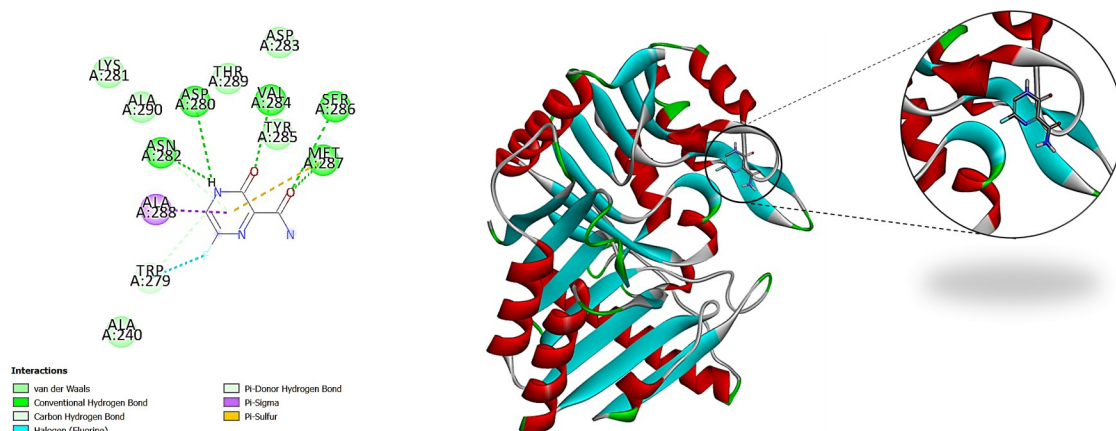
Fig 11. 2D and 3D representations of the docked pose for the predicted interaction between Favipiravir. The green dashed lines represent hydrogen bonds and F-13 protein.
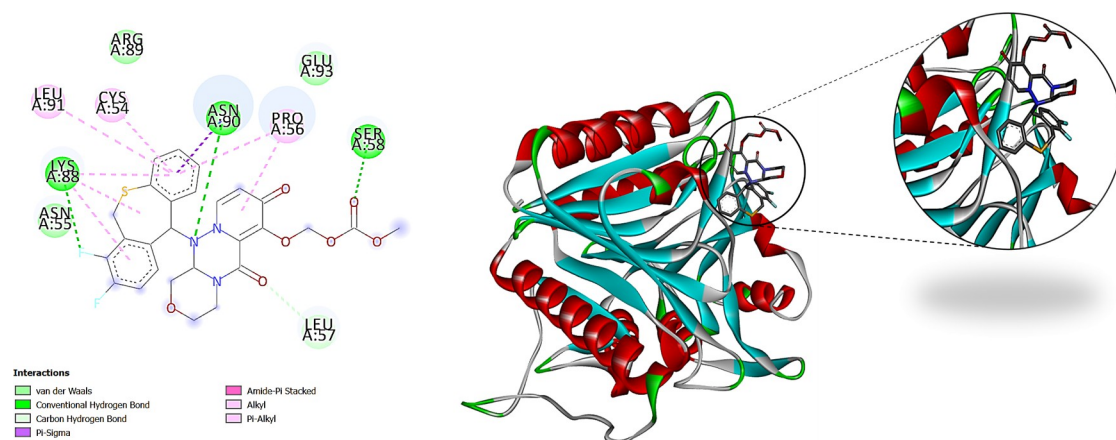
Fig 12. 2D and 3D representations of the docked pose for the predicted interaction between Baloxavir marboxil. The green dashed lines represent hydrogen bonds and F-13 protein.

ASN282, MET287, and SER286 residues, along with van der Waals interactions with other residues.

Fig 11 shows Favipiravir's 3D and 2D representations. For the interaction of favipiravir with F-13 (Fig 11), conventional hydrogen bonds, pi-alkyl interactions, and weak van der Waals bonds dominate.

As the last suggested drug, Fig 12 shows Baloxavir marboxil's 3D and 2D representations. Baloxavir binds to the F-13 protein by forming hydrogen bonds with ASP280, VAL284, SER286, ASN282, and MET287 residues, as well as other interactions such as pi-sigma, pi-sulfur, and van der Waals forces.

4 Conclusion

While the monkeypox virus has caused a global epidemic in recent years, there are no approved drugs for its treatment. This work proposed a computational drug repurposing approach to suggest several drugs to deal with MPXV. To do this, we prepared a dataset from

viruses and antivirals to apply computational learning methods for drug suggestions. We have applied SVM, random forest, decision tree, and convolutional deep model (DRaW) to the virus and antiviral features to learn virus-antiviral interactions. The performance analysis shows that DRaW has the highest performance and the random forest comes after that. Then, using voting, the highly promising antivirals, i.e., Tilorone, Valacyclovir, Ribavirin, Baloxavir marboxil, and Favipiravir, suggested by the learning methods have been chosen for further analysis. We did the homology modeling and molecular docking study on the proposed drugs. The homology modeling using AlphaFold makes it clear that envelope receptor F-13 can be the target of the antivirals. Therefore, we applied the docking studies on the suggested drugs of the computational modeling and it approved the hypothesis with high binding affinity. Generally, to our best knowledge, this is the first work that proposes antivirals for treating the MPXV. These screening results on Tilorone, Valacyclovir, Ribavirin, Baloxavir marboxil, and Faviapiravir can be further analyzed using laboratory analysis. The work approves the high potential of computational drug repurposing for the screening phase of drug discovery which causes lower time and costs.

Author Contributions

Conceptualization: Arash Zabihian, Mohsen Hooshmand.

Methodology: Mohsen Hooshmand.

Project administration: Mohsen Hooshmand.

Resources: Morteza Hashemi, Arash Zabihian, Mohsen Hooshmand.

Software: Morteza Hashemi, Masih Hajsaeedi.

Supervision: Mohsen Hooshmand.

Validation: Arash Zabihian.

Writing – original draft: Arash Zabihian, Mohsen Hooshmand.

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

References

1. Jarrell L, Perryman K. Mpox (monkeypox): Diagnosis, prevention, and management in adults. The Nurse Practitioner. 2023; 48(4):13–20. https://doi.org/10.1097/01.NPR.0000000000000025 PMID: 36975744

2. Saraswat Y, Shah K. Mini Review on Clinical Aspects of Monkeypox. Current Pharmaceutical Biotechnology. 2024;. https://doi.org/10.2174/1389201025666230914094444 PMID: 37711132

3. Lim CK, McKenzie C, Deerain J, Chow EP, Towns J, Chen MY, et al. Correlation between monkeypox viral load and infectious virus in clinical specimens. Journal of Clinical Virology. 2023; 161:105421. https://doi.org/10.1016/j.jcv.2023.105421 PMID: 36893717

4. cdc. 2022 U.S. Map & Case Count | Mpox | Poxvirus | CDC—cdc.gov; 2024. https://www.cdc.gov/poxvirus/mpox/response/2022/us-map.html.

5. O'Laughlin K, Tobolowsky FA, Elmor R, Overton R, O'Connor SM, Damon IK, et al. Clinical use of tecovirimat (Tpoxx) for treatment of monkeypox under an investigational new drug protocol—United States, May–August 2022. Morbidity and Mortality Weekly Report. 2022; 71(37):1190. https://doi.org/10.15585/mmwr.mm7137e1 PMID: 36107794

6. Hudu SA, Alshrari AS, Al Qtaitat A, Imran M. VP37 protein inhibitors for mpox treatment: Highlights on recent advances, patent literature, and future directions. Biomedicines. 2023; 11(4):1106. https://doi.org/10.3390/biomedicines11041106 PMID: 37189724

7. Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator Regularized Non-Negative Matrix Factorization Method-Based Drug Repurposing for COVID-19. Frontiers in Immunology. 2021; 11. https://doi.org/10.3389/fimmu.2020.603615 PMID: 33584672

8. 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

9. 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

10. 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

11. Dotolo S, Marabotti A, Facchiano A, Tagliaferri R. A review on drug repurposing applicable to COVID-19. Briefings in bioinformatics. 2021; 22(2):726–741. https://doi.org/10.1093/bib/bbaa288 PMID: 33147623

12. 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

13. Chen R, Liu X, Jin S, Lin J, Liu J. Machine learning for drug-target interaction prediction. Molecules. 2018; 23(9):2208. https://doi.org/10.3390/molecules23092208 PMID: 30200333

14. 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

15. 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

16. 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

17. 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

18. 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.compbiomed.2021.105119 PMID: 34902608

19. 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

20. Zabihian A, Asghari J, Hooshmand M, Gharaghani S. A Comparative Analysis of Computational Drug Repurposing Approaches–Proposing a Novel Tensor-Matrix-Tensor Factorization Method. submitted. 2024;.

21. Ianevski A, Simonsen RM, Myhre V, Tenson T, Oksenych 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

22. of Medicine (US) NL. The National Library of Medicine. US Department of Health, Education, and Welfare, Public Health Service; 1972.

23. 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

24. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009; 25(11):1422–1423. https://doi.org/10.1093/bioinformatics/btp163 PMID: 19304878

25. 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 PMID: 7265238

26. Ison J, Rice P, Bleasby A. emboss tools; 2009. Available from: http://emboss.open-bio.org/.

27. 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.

28. Agency EM. Tecovirimat SIGA; 2021.

29. Patel CN, Mall R, Bensmail H. AI-driven drug repurposing and binding pose meta dynamics identifies novel targets for monkeypox virus. Journal of Infection and Public Health. 2023; 16(5):799–807. https://doi.org/10.1016/j.jiph.2023.03.007 PMID: 36966703

30. Cramer P. AlphaFold2 and the future of structural biology. Nature structural & molecular biology. 2021; 28(9):704–705. https://doi.org/10.1038/s41594-021-00650-1 PMID: 34376855

31. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, et al. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins: Structure, Function, and Bioinformatics. 2009; 77(S9):114–122. https://doi.org/10.1002/prot.22570 PMID: 19768677

32. Heo L, Park H, Seok C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. Nucleic acids research. 2013; 41(W1):W384–W388. https://doi.org/10.1093/nar/gkt458 PMID: 23737448

33. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007; 35(suppl_2):W407–W410. https://doi.org/10.1093/nar/gkm290 PMID: 17517781

34. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein science. 1993; 2(9):1511–1519. https://doi.org/10.1002/pro.5560020916 PMID: 8401235

35. Ramachandran Gt, Sasisekharan V. Conformation of polypeptides and proteins. Advances in protein chemistry. 1968; 23:283–437. https://doi.org/10.1016/S0065-3233(08)60402-7 PMID: 4882249

36. Asiamah I, Obiri SA, Tamekloe W, Armah FA, Borquaye LS. Applications of molecular docking in natural products-based drug discovery. Scientific African. 2023; p. e01593. https://doi.org/10.1016/j.sciaf.2023.e01593

37. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic acids research. 2021; 49(D1):D1388–D1395. https://doi.org/10.1093/nar/gkaa971 PMID: 33151290

38. Zabihian A, Sayyad FZ, Hashemi SM, Shami Tanha R, Hooshmand M, Gharaghani S. DEDTI versus IEDTI: 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

39. Li D, Liu Y, Li K, Zhang L. Targeting F13 from monkeypox virus and variola virus by tecovirimat: Molecular simulation analysis. Journal of Infection. 2022; 85(4):e99–e101. https://doi.org/10.1016/j.jinf.2022.07.001 PMID: 35810941

40. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020; 21(1):6. https://doi.org/10.1186/s12864-019-6413-7 PMID: 31898477

41. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. Canadian Journal of Emergency Medicine. 2006; 8(1):19–20. https://doi.org/10.1017/S1481803500013336 PMID: 17175625

42. Bramer M. Avoiding overfitting of decision trees. Principles of data mining. 2007; p. 119–134.

43. Laskowski RA, Chistyakov VV, Thornton JM. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic acids research. 2005; 33(suppl_1):D266–D268. https://doi.org/10.1093/nar/gki001 PMID: 15608193

44. Pakpahan M, Rusmerryani M, Kawaguchi K, Saito H, Nagao H. Evaluation of scoring functions for protein-ligand docking. In: AIP Conference Proceedings. vol. 1518-1. American Institute of Physics; 2013. p. 645–648.

45. Biovia DS. Discovery Studio Visualizer v21. 1.0. 20298. San Diego: Dassault Systèmes. 2021;.