

A Deep Learning Framework for Prediction of the Mechanism of Action

Jingyuan Dai
University of Nevada, Las Vegas
4505 S Maryland Pkwy
Las Vegas, NV 89154

Shahram Latifi
University of Nevada, Las Vegas
4505 S Maryland Pkwy
Las Vegas, NV 89154

ABSTRACT

This paper aims to apply deep learning algorithms to advance a new drug's mechanism of action (MoA) prediction. Since one drug can have one or more MoAs, algorithms must be developed to perform multi-label classification problems. This paper puts forward a deep learning framework, MoA Net, which ensembles one residual network and five convolutional neural networks to predict MoA targets. To find optimal parameter sets, the authors implement Bayesian tuning techniques on each sub network of MoA Net. The study uses logarithmic loss function to evaluate the model's performance. Results show successful MoA target prediction in the dataset provided by the LISH and LINCS Connectivity Map.

General Terms

Classification, Deep Learning, Feature Engineering

Keywords

Convolutional Neural Nets, Genetic Expression, Drugs, Mechanism of Actions, Residual Networks

1. INTRODUCTION

Over the decades, with scientists understanding a disease's underlying mechanism of action (MoA), discovering drugs' effects relies heavily on scientific research experiments. However, with the traditional *one drug, one MoA* target archetype, progress in drug discovery and drug repurposing is costly, in both time and economic senses [1]. For example, the average time for the US Food & Drug Administration (FDA) to approve a new drug application ranges from 9 to 12 years, with more than \$1 million in cost each year [2]. With deeper understanding in the development of a drug's MoA, given clinical trails, the traditional *one drug, one MoA* target assumption has evolved to a *multiple drug, multiple MoA* targets model. This model sheds light on drug application development first, because it is known that one drug can target multiple proteins, rather than only one [2]; and second, different drugs may interact with different proteins in addition to their primary targets [3]. As such, researchers are motivated to discover innovative methods to reduce drug production costs.

With the advent of machine learning techniques, MoA prediction has bloomed in recent years. There are many review articles summarizing how machine learning approaches can be helpful in predicting MoAs. [4-8] The state-of-art methods can be categorized as two types: supervised and semi-supervised. Additionally, supervised methods can be classified further as methods that are based on distance, feature, and matrix factorization [4].

Supervised machine learning methods can be applied when both drug and MoA labels are available. In these methods,

when a certain drug and MoA target interact with each other, they are both labeled as ones, while the non-interactive ones are labeled as zeros. Distance-based, or similarity-based, methods are built upon the assumption that similar drugs have a high possibility to target similar MoAs [5]. Based on different measures of defining similarities among drugs and MoAs, different kernel functions are used. The nearest profile method (NN) is one of the most exploratory methods, using Euclidean distance [9]. To further extend, other methods based on chemical, ligand, expression, side effect, and annotation of the drug to MoA similarities have been developed [10]. In one study [11], an algorithm based on the similarities of neighbors was developed. The algorithm takes the weighted average among neighbors to make predictions. Later, the integrated neighbor-based [12] method, which ensembles different neighbor-based models to generate a higher score, thus improve prediction performance, was put forward. In these kinds of ensembling methods, Jaccard similarity, Cosine similarity, and Pearson correlation similarity are commonly used to calculate the similarity score.

Feature-based supervised machine learning methods form a large group of algorithms predicting drug to target MoA. These methods utilize the most state-of-art machine learning models that have been successfully implemented in other fields. Random forest, Support Vector Machine (SVM) and other binary-classification methods are commonly used [13-15]. The logic is for each drug-MoA pair, the model first creates the feature vectors, and then performs the binary classification task [4]. A widely-used feature composition method is representing each drug as a set of 1080 descriptors (both structural and physicochemical), such as amino acid and dipeptide composition descriptors [17]. Then, this 1080-dimensional feature vector for each drug-MoA pair is fed into models such as random forest, etc. To better overcome the overfitting problems in a large-scale dataset, as well as increase accuracy in prediction, later works including DTINet [18], NeoDTI [19], and restricted Boltzmann machine [20], all proved robustness and improved accuracy in identifying binary drug to MoA prediction.

Matrix factorization methods have been proven to manage the missing values in a drug-MoA matrix. They decompose the drug-MoA matrix into one feature vectors of the drugs and one of the corresponding MoAs, such that a rank reduction is performed [4], [25]. Pioneering works include Kernelized Bayesian Matrix Factorization [21]. This method utilizes a fully conjugate probabilistic model and a deterministic variational approximation method to project drugs and MoAs into one unified subspace. Another method, namely the probabilistic matrix factorization method [22], has also been put forward. Based on a collaborative filtering algorithm, this method first defines a conditional probability for each observed drug to MoA interaction, a zero-mean spherical

Gaussian prior on the two matrices, then computes the log-likelihood of the two matrices to get the underlying drug to target MoA score. Recent works include the Neighborhood Regularized-based Logistic Matrix Factorization method [23]. This method models the probability of a drug interacting with a MoA target, based on logistic matrix factorization. A Dual-Network integrated Logistic Matrix Factorization method [24] extends the Neighborhood Regularized-based Logistic Matrix Factorization method by calculating probabilities for unknown drug-MoA pairs, then computes the interaction score by maximizing an objective function using the Hadamard product.

In order to address an imbalanced data problem, semi-supervised machine learning methods are put forward. These methods achieve promising prediction performance in inferring unlabeled data by the labeled data, where unlabeled data means no MoA or an unknown MoA, and labeled data means a known MoA. In this case, the unlabeled data provides information into the prediction process [25]. Due to the use of large amount of unlabeled data, semi-supervised methods achieve good performance. A Laplacian regularized least square, based on the BLM model, has been put forward with no use of negative samples [26]. Later, methods that combine drug-MoA interaction and chemical and genomic space information to form a new kernel were also studied.

Although the models mentioned above generate promising results, they have limits. Distance-based methods stand out due to their similarity function simplicity. However, when constructing similar neighbors, most algorithms are only constructed based on first-order similarity, and similarity transitivity is not considered [16]. Feature-based models outperform distance-based methods in increasing model robustness and prediction accuracy; however, models such as SVM and tree-based models tend to have greater bias for the majority class, resulting in an imbalanced data issue, especially when the amount of labeled MoAs are relatively small, compared to drug features. In contrast, by integrating labeled and unlabeled data, semi-supervised learning methods generate more reliable results, given such an imbalanced case; however, these are costly in both time and computation perspectives to implement on large-scale datasets. The main drawback of such methods is that they are put forward based on binary classification machine learning methods. Given the fact that multiple drugs may trigger one or more MoAs, a more efficient family of machine learning methods will be studied. One popular track is to borrow deep learning networks, such as convolutional neural networks, in the computer vision area.

This paper puts forward a framework that utilizes fully convolutional and residual neural networks, originally from the computer vision field, and performs model ensembling to achieve better prediction results.

2. METHODOLOGY

This paper proposes a deep learning scheme, MoA Net, to predict the MoAs of specific drugs. MoA Net contains a principle component analysis (PCA) part to extract features from the given feature set, then uses multiple ensembled convolutional neural networks and residual neural networks to perform the final multi-label MoA classification prediction. To find optimal parameter sets, MoA Net uses Bayesian tuning techniques for each convolutional and residual neural network. The structure of MoA Net is shown below in Figure 1:

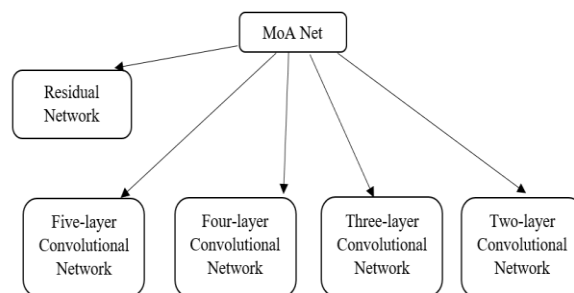


Fig 1: Structure of MoA Net

Half of MoA Net's structure is composed of multiple sub convolutional neural networks that range from two to five layers. The essential block of one convolutional neural network is composed of one batch normalization layer, followed by one dropout layer to avoid overfitting, one weight normalization layer, and ending with a dense convolutional layer to generate output to the next block. The structure of one block of convolutional neural network is showed in Figure 2:

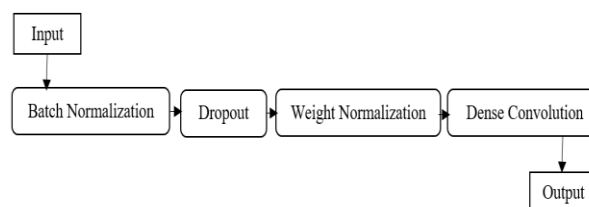


Fig 2: Structure of one block of convolutional neural network

Figure 4 shows an example of a three-layer network. It contains one input layer, three hidden layers, and one output layer. The input layer takes in selected features from PCA, then a first stage batch normalization[27] is applied to scale different features. In the following three hidden layers, one dropout layer [28] is used to avoid overfitting, and then weight normalization [29] is used to accelerate the convergence of stochastic gradient descent optimization.

The other structure of MoA Net is its deep residue neural network. This is inspired by He K et al.'s work [30]; the residual structure is shown in Figure 4. The initial input layer, with features selected by PCA, is first batch-normalized, and then one dropout layer and one dense convolutional layer are applied. One batch-normalization layer, one dropout layer, and one dense convolutional layer is referred as one single block. This block is then used as the residue, which is taken as a performance average, before the final batch normalization, dense layer, and the output layer.

In order to find the optimal parameter set for each sub-network in MoA Net, the Bayesian tuning technique [31] is applied for each of the convolutional neural networks and residual networks. Bayesian tuning incorporates prior belief about the objective function $f(x)$, and then updates the prior with sample points drawn from $f(x)$ to get a posterior that better approximates it. It is proven to be efficient to tune parameters in algorithms that are expensive for the evaluation and computation cost [31]. To address the framework of this paper, a flowchart is provided in Figure 3 below.

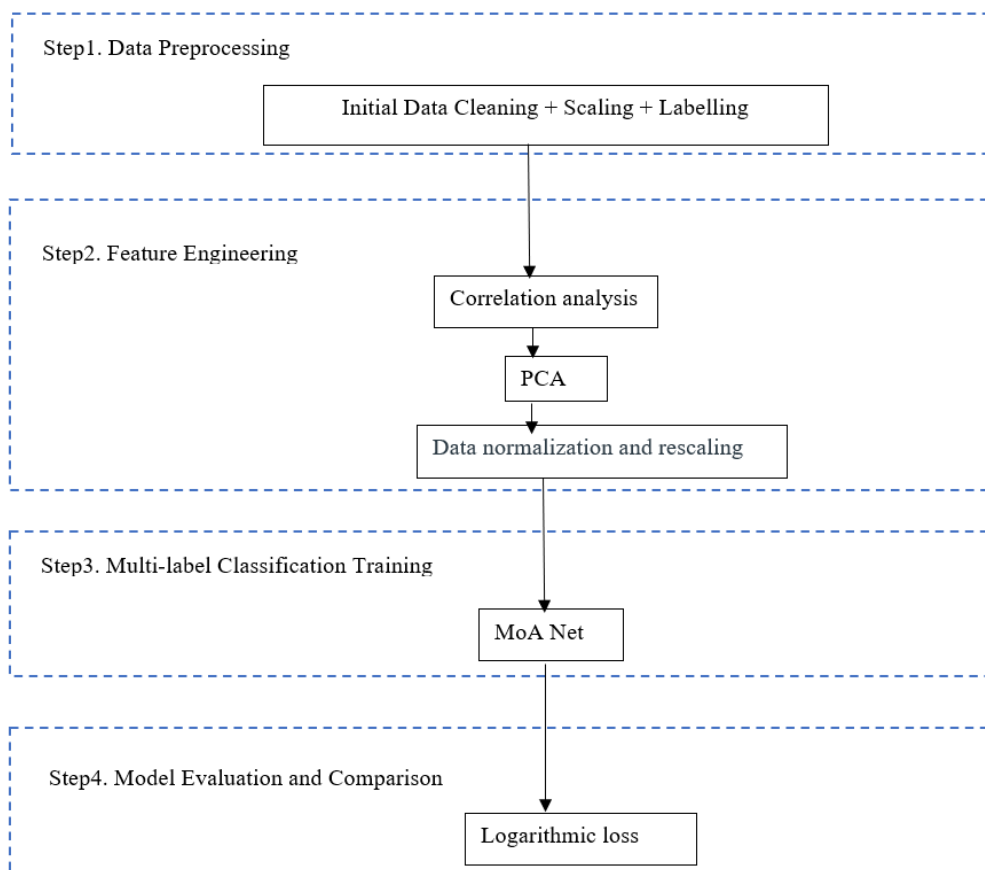


Fig 3: Flowchart of the proposed framework

The first step is to perform data preprocessing, including cleaning, scaling, and labelling the initial features. The second step is feature engineering, where PCA selects features out of the original feature map, and then construct features that will be used as input to the MoA Net. Data normalization and rescaling are also performed to make sure they are in the right format for prediction. The third step is to perform the multi-label target learning of MoA prediction using MoA Net. MoA Net ensembles multiple convolutional neural networks with different layers, as well as a residual network. This framework generalization is achieved through the flexibility of ensembling convolutional neural network and residual network combinations. The Bayesian tuning technique is applied to each convolutional neural network and residual network to achieve optimal performance of each sub-network. The final step is to evaluate the performance of this frame by using ten-fold cross validation.

3. DATA DESCRIPTION AND FEATURE ENGINEERING

The datasets used in this paper are those released from the challenge from the Laboratory for Innovation Science at Harvard (LISH), and the NIH Common Funds Library of Integrated Network-Based Cellular Signatures (LINCS) to advance drug development through improvements to MoA prediction algorithms [32].

The dataset contains 872 features of both genetic expression and cell viability data, with 206 MoA targets to predict. Among these 872 features, 772 genes are measured and provided as genetic expression data, labelled as g-1 to g-771. One hundred cell lines are extracted to form cell viability,

labelled as c-1 to c-71, which is defined as the ratio of the number of visible cells over total number of cells, after incubating certain drugs and the mixture of cells. For each incubation experiment, a 24, 48 and 72-hour incubating window is used, respectively. The dataset provides 206 MoA targets for each drug, labelled as 0 for No-MoA and 1 for MoA. Since one drug can have one or more than one MoA, this is a multi-label classification problem.

In summary, the total number of 872 features are collected from genetic expression and cell viability data extracted from 100 cell lines treated with a certain drug under three incubating windows (24, 48 and 72 hour). The total number of target MoAs is 206.

The first step of feature engineering is to do a correlation analysis by dropping features that have high correlation. In order to prevent information loss at the initial stage, the correlation threshold is set to be 0.9. One genetic expression is dropped and 28 cell viability features are dropped after this analysis. PCA is conducted on both genetic expression and cell viability features after performing MinMaxScaling to the numerical features. A total number of 1069 features are fed as input into MoA Net, among which there are 399 out of 772 genetic expression features and a total number of 100 out of 100 cell viability features.

4. MODEL IMPLEMENTATION AND TUNNING

MoA Net is a deep learning ensembled model that consists of one residual network and five convolutional neural networks with different layers. Here, the average ensemble is used to reduce model variance, and thus, achieves better performance.

A Bayesian tuning technique is applied to find the optimal parameter set of each sub-network.

For the residual network, each of its convolutional blocks consists of a batch normalization operation of the input, followed by a dropout operation with a different dropout rate, then a fully convolutional layer producing output to the next block. The activation function used for each block is the ReLU [33] function, and the loss function used to optimize back propagation of the weights is the binary cross entropy [34].

For the five convolutional neural networks, one to four hidden layers are implemented, respectively. Similar to the residual network, there is one batch normalization operation to scale the input, and then a dropout operation with different dropout rate to reduce model overfitting, followed by a fully connected convolutional layer for each block.

Figure 4 shows the structure of the residual network and a three-layer convolutional neural network of MoA Net.

The range used for the dropout rate is from 0.1 to 0.5 continuously, and the number of neurons is from 32 to 1024 discretely. Table 1 shows the optimal parameter set for each sub model of MoA Net, selected after the Bayesian tuning technique.

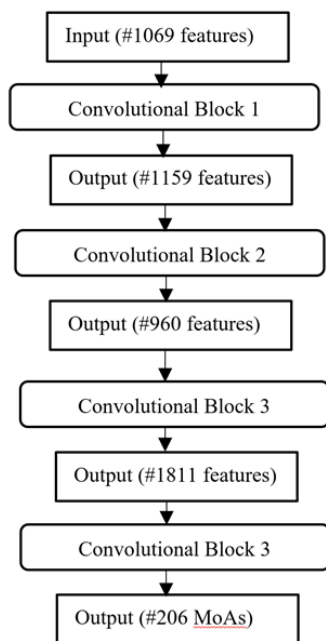


Fig 4. A: Structure of a three-layer convolutional neural network of MoA Net

Table 1. Optimal parameter set for MoA Net

Name	Drop out Rate	Neurons	Learning Rate
Five-layer	0.47689	2560	0.0015
Convolutional	0.49140	2048	
Network	0.46783	1524	
	0.39886	1012	
	0.38091	780	
Four-layer	0.47689	1340	0.0020
Convolutional	0.38000	2000	
Network	0.42001	2555	
	0.38900	910	
Three-layer	0.49140	1159	0.0015
Convolutional	0.18818	960	
Network	0.12452	1811	
Two-layer	0.26886	1292	0.0015
Convolutional	0.45982	983	
Network			
Residual	0.28876	512	
		256	
Network	0.30147	512	
		512	
		256	
		256	

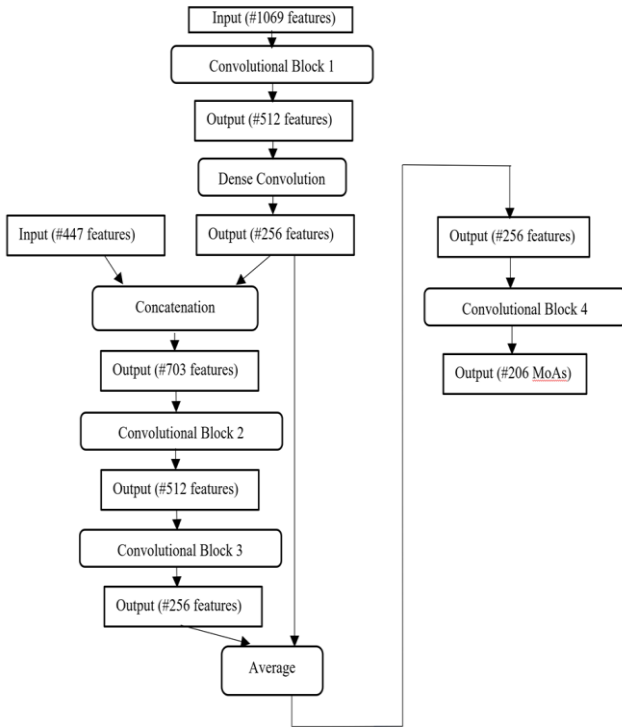


Fig 4. B: Structure of a three-layer convolutional network of MoA Net

5. EVALUATION

This model uses the logarithmic loss function to evaluate the prediction performance on each drug to MoA pair [32]. Logarithmic loss is the log of the likelihood function, and it can be defined as follows:

$$\text{loss} = -\frac{1}{M} \sum_{m=1}^M \frac{1}{N} \sum_{i=1}^N [y_{i,m} \log(\widehat{y}_{i,m}) + (1 - y_{i,m}) \log(1 - \widehat{y}_{i,m})]$$

where N is the number of observations; M is the number of labelled and scored MoA targets;

$\widehat{y}_{i,m}$ is the predicted probability of a positive response for a given observation; and $y_{i,m}$ is the labelled ground truth, where 1 stands for a positive response, and 0 stands for negative.

A ten-fold stratified cross validation is used to evaluate the performance for each sub model in MoA Net. From the results, one can see each sub model is able to converge to a low logarithmic loss, and is able to predict a drug and MoA pair successfully. In particular, the residual network starts initially from the lowest loss among the five sub-models, which shows the capability to predict the target well, but it then slows down until it converges after 20 epochs, while the four convolutional neural networks start at a higher loss, yet converge quicker, after 7 to 15 epochs. The final ensemble model achieves better performance both in its capability to predict the drug and MoA target, and to converge at a faster speed. The detailed logarithmic loss for each sub-model is listed in Table 2. A comparison from different public networks released from the challenge is shown in Table 3. Among the available public networks, performance of traditional machine learning algorithms such as Ridge

regression model, gradient boosting models, and bagging models hits an average of a 0.01957 logarithmic loss. Due to the nature of deep learning algorithms to solve multi-label classification problems, models such as ResNet and TableNet achieve an average logarithmic loss of 0.01840, which is a 5.98% decrease compared with traditional machine learning algorithms. The Box and Whisker plot is showed in Figure 5.

Table 2. MoA Net Performance

Name	Logarithmic loss	Convergence after
Five-layer Convolutional Network	0.016282	7
Four-layer Convolutional Network	0.016194	10
Three-layer Convolutional Network	0.016196	11
Two-layer Convolutional Network	0.016289	15
Residual Network	0.016375	20
MoA Net	0.0162672	3

Table 3. Model Performance Comparison

Name	Logarithmic loss
Ridge	0.02000
Light GBM	0.01950
XGBoost	0.01920
ResNet	0.01850
TableNet	0.01830
MoA Net	0.01626

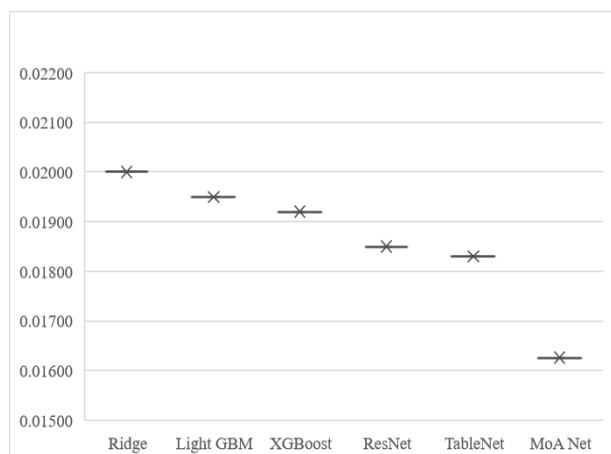


Fig 5. Box and Whisker plot of machine learning models and deep learning models

6. CONCLUSION

This paper proposes a deep learning framework, MoA Net, to advance the prediction of MoA targets from the potential genetic expression and cell viability data of specific drugs. Since the drug to MoA relation is not a one to one relation, rather one drug may have one or more MoA targets, this is a multi-label classification problem. Most machine learning algorithms use transformation to convert this type of problem into a multiple binary classification problem, yet performance is not ideal regarding computation time and accuracy. Therefore, this paper uses a deep learning framework that ensembles one residual network and four convolutional neural networks to improve on the prediction of drug to MoA relation. The five sub models from MoA Net ensure the capability to solve this multi-label classification prediction without converting to a binary classification problem, while the ensembling reduces model variance, so it is able to generalize in the test dataset. To test MoA Net performance, the challenge dataset released from the Connectivity Map of LISH and LINCS was used, and a logarithmic loss was used to evaluate the prediction ability of MoA Net. Results compared from shallow machine learning algorithms showed MoA Net's success, and hence, this network provides insights in advancing the drug to MoA prediction process.

7. ACKNOWLEDGMENTS

The authors want to thank Meagan Madariaga-Hopkins for proof-reading this work.

8. REFERENCES

- [1] Masoudi-Nejad A, Mousavian Z, Bozorgmehr JH. "Drug-target and disease networks: polypharmacology in the post-genomic era." *In Silico Pharmacol.* 2013; 1(1):17.
- [2] Dickson M, Gagnon JP. "Key factors in the rising cost of new drug discovery and development". *Nat Rev Drug Discov.* 2004; 3(5):417-29.
- [3] Paul SM, Mytelka DS, Dunwiddie CT, et al. "How to improve R&D productivity: the pharmaceutical industry's grand challenge". *Nat Rev Drug Discov.* 2010; 9(3):203-14.
- [4] Maryam B, Elyas S, Kai W, et al. "Machine learning approaches and databases for prediction of drug-target interaction: a survey paper". *Briefings in Bioinformatics.* 2021; 1(22); 247–269.
- [5] Ding H, Takigawa I, Mamitsuka H, et al. "Similarity-based machine learning methods for predicting drug-target interactions: a brief review." *Brief Bioinform* 2014;15:734–47.
- [6] Cheng T, Hao M, Takeda T, et al. "Large-scale prediction of drug-target interaction: a data-centric review." *AAPS J.* 2017;19:1264–75.
- [7] Ezzat A, Wu M, Li XL, et al. "Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey." *Brief Bioinform* 2019;20:1337–1357.
- [8] Chen X, Yan CC, Zhang X, et al. "Drug-target interaction prediction: databases, web servers and computational models." *Brief Bioinform* 2016;17:696–712.
- [9] Yamanishi Y, Araki M, Gutteridge A, et al. "Prediction of drug-target interaction networks from the integration of chemical and genomic spaces." *Bioinformatics.* 2008; 24(13):i232-40.
- [10] Perlman L, Gottlieb A, Atias N, et al. "Combining drug and gene similarity measures for drug-target elucidation." *J Comput Biol* 2011;18(2):133–45.
- [11] Zhang W, Zou H, Luo L, et al. "Predicting potential side effects of drugs by recommender methods and ensemble learning." *Neurocomputing.* 2016;173:979–987
- [12] Zhang W, Chen Y, Liu F, et al. "Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data." *BMC Bioinformatics.* 2017; 18(1):18.
- [13] Wassermann AM, Geppert H, Bajorath J. "Ligand prediction for orphan targets using support vector machines and various target-ligand kernels is dominated by nearest neighbor effects". *J Chem Inf Model* 2009;49(10):2155–67.
- [14] Niu YQ. "Supervised prediction of drug-target interactions by ensemble learning". *J Chem Pharm Res* 2014;6:1991–9.
- [15] Shang Z, Jin L, Jiang Y, et al. "A method of drug target prediction based on SVM and its application". *Prog Modern Biomed* 2012;20.
- [16] Zhang P, Wang F, Hu J, et al. "Label Propagation Prediction of Drug-Drug Interactions Based on Clinical Side Effects". *Sci Rep.* 2015 Jul 21; 5(0):12339.
- [17] Li ZR, Lin HH, Han LY, et al. "PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence". *Nucleic Acids Res.* 2006;34(Web Server issue):W32-W37. doi:10.1093/nar/gkl1305
- [18] Luo Y, Zhao X, Zhou J, et al. "A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information". *Nat Commun.* 2017; 8(1):573.
- [19] Wan F., Hong L., Xiao A., et al. "Neural integration of neighbor information from a heterogeneous network for discovering new drug-target interactions". *Bioinformatics.* 2018
- [20] Wang Y, Zeng J. "Predicting drug-target interactions using restricted Boltzmann machines." *Bioinformatics.* 2013; 29(13):i126-34.
- [21] Gönen M. "Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization". *Bioinformatics.* 2012; 28(18):2304-10.

- [22] Cobanoglu MC, Liu C, Hu F et al. "Predicting drug-target interactions using probabilistic matrix factorization". 2013; 53(12):3399-409
- [23] Liu Y, Wu M, Miao C et al. "Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction". *PLoS Comput Biol.* 2016; 12(2):e1004760.
- [24] Hao M, Bryant SH, Wang Y. "Predicting drug-target interactions by dual-network integrated logistic matrix factorization". *Sci Rep.* 2017; 7():40376.
- [25] Zhou L, Li Z, Yang J, et al. "Revealing Drug-Target Interactions with Computational Models and Algorithms". *Molecules.* 2019;24(9):1714.
- [26] Xia Z, Wu LY, Zhou X et al. "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces". *BMC Syst Biol.* 2010; 4 Suppl 2():S6.
- [27] Sergey S and Christian S, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". arXiv:1502.03167
- [28] Srivastava N, Hinton G, Krizhevsky A, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15.1 (2014): 1929-1958.
- [29] Tim S and Diederik P. K, "Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks". 30th Conference on Neural Information Processing Systems (NIPS) 2016
- [30] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [31] Snoek J, Hugo L, and Ryan P. A. "Practical bayesian optimization of machine learning algorithms." arXiv preprint arXiv:1206.2944 (2012).
- [32] Mechanisms of Action (MoA) Prediction, Kaggle, <https://www.kaggle.com/c/lish-moa/overview>.
- [33] Agarap, Abien F. "Deep learning using rectified linear units (relu)." arXiv preprint arXiv: 1803.08375 (2018).