Large-Scale Prediction of Drug-Target Interactions from Deep Representations

Peng-Wei Hu Keith C.C. Chan Zhu-Hong You
Department of Computing
Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
{csphu, cskcchan, csyzhuhong }@comp.polyu.edu.hk

***Abstract*—Identifying drug-target interactions (DTIs) is a major challenge in drug development. Traditionally, similarity-based methods use drug and target similarity matrices to infer the potential drug-target interactions. But these techniques do not handle biochemical data directly. While recent feature-based methods reveal simple patterns of physicochemical properties, efficient method to study large interactive features and precisely predict interactions is still missing. Deep learning has been found to be an appropriate tool for converting high-dimensional features to low-dimensional representations. These deep representations generated from drug-protein pair can serve as training examples for the interaction predictor. In this paper, we propose a promising approach called multi-scale features deep representations inferring interactions (MFDR). We extract the large-scale chemical structure and protein sequence descriptors so as to machine learning model predict if certain human target protein can interact with a specific drug. MFDR use Auto-Encoders as building blocks of deep network for reconstruct drug and protein features to low-dimensional new representations. Then, we make use of support vector machine to infer the potential drug-target interaction from deep representations. The experiment result shows that a deep neural network with Stacked Auto-Encoders exactly output interactive representations for the DTIs prediction task. MFDR is able to predict large-scale drug-target interactions with high accuracy and achieves results better than other feature-based approaches.**

Keywords—autoencoder, drug-target interaction, deep representations.

I. INTRODUCTION

Drug discovery is a comprehensive study of diverse objects and provides detailed descriptions of the biological activity, genomic features and chemical structure to the disease treatment. Lead compound interacting human protein is one of the key procedures responsible for driving important biological actions within the human body cell. The mainly treatment processes within our body are carried out by adjust proteins status that physically interact to form counter effect of disease. Discovery of such drug-target interactions that take place within a human body can suggest new drug target protein and aid the design of new compounds by providing rational drug targets [1]. As the biggest drug database, PubChem collected more than 35 million compounds of which 7000 compounds containing target protein information. During the drug discovery processing, medicine production line often generate results different from the original goal. Such effects may be raised with hidden factors and biological domains in drug target selection and lead compound screening. Instability and no specificity of drug-target interactions have to be addressed appropriately before send them to clinical phase. Apparently, complex procedure of confirm lead compound range from target identification to lead compound optimization is a really long-term work. Even many optimal approaches have been proposed to tackle the problem of drug-target interaction prediction, a new drug discovery still need cost 6-10 years. Therefore, the identification of potential drug-target interactions is a challenging issue at the start stage of drug development process [2].

In order to solve such problems, some interdisciplinary scientists introduce several computational approaches to cope with such problems. Previous attempts are divided broadly into similarity-based approach and feature vector-based techniques [3]. Similarity-based methods are developed to discover potential DTIs through the similarity matrices of drug and protein. Some early works at discover drug-target interactions have been proposed based on their compound and protein sequence similarity [4]-[6]. Feature vector-based methods are regarded as more advanced strategies that face drug and protein features straightforward. They can uncover the description of the hidden knowledge in terms of meaningful features and then generate rules to reproduce experts' decision process. These methods provide meaningful solutions for discovering interest patterns such as single molecule sub-structure influence but they are not able to precisely reflect the molecule substructure and protein subspace interactions. It's also difficult for current techniques to analysis real high-dimensional protein descriptors except apply parallel scheme [7]. Fortunately, it is widely believed such problem can be resolved by deep learning mechanisms [8]-[9]. Compressed representation have been used to assess large volumes of protein attributes and possible uncovers significant hidden relationships exist in the protein [10]-[11].

In this study, we develop a new deep learning-based method for the prediction of drug-protein interactions from protein sequence descriptors and molecule fingerprints with Support Vector Machine aiming at improving the efficiency

978-1-5090-0620-5/16/$31.00

and effectiveness of the classification accuracy. Firstly, we introduce multi-scale local descriptor approach for discover realistic large amino acid sequences descriptors and use chemical fingerprints for represent the chemical space. Secondly, in order to enhance the accuracy and transfer the large descriptors to deep representations, the extracted features of input layer would be automatically learned by an unsupervised Stacked Auto-encoder for output a reconstruct lower dimensional layer. Finally, we focus on use classifier to judge whether one drug interact with one target. Our method constitutes a significant advance because it logically consider coupled representations of protein and molecules that remain unobserved in any interaction. We adopt a popular data standard to test proposed method which includes G protein-coupled receptor, enzyme, ion channel, and nuclear receptor data set [4]. MFDR has been tested with Gold standard data sets that can be a very useful approach to predict the DTIs. The basic steps of MFDR, Figure 1 shows a procedure of the proposed method according to our definition.

The rest of this paper is organized as follows: Section 2 focus on introducing previous advanced studies about the drug-target prediction. Then Section 3 elaborates the details of our novel method for large-scale features deep representations and classifier for predicting drug-target interactions. Section 4 presents the experiments and results as well. Finally, conclusions are given in Section 5.

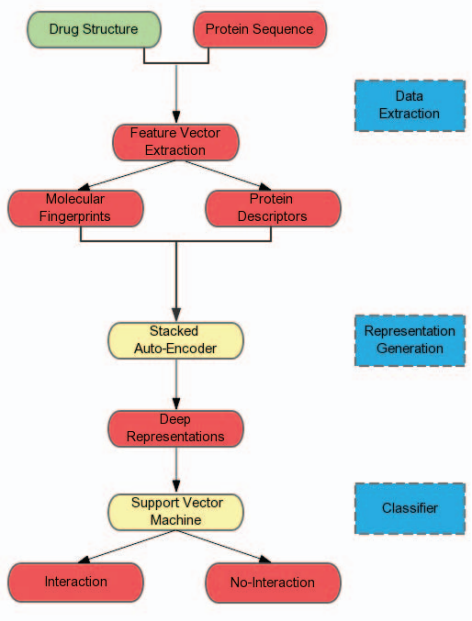

Fig. 1. The procedure of MFDR

II. RELATED WORK

From the perspective of the drug design, the prior drug-target prediction based on the data model is a good place to start. Previously, the main research area in drug-target interaction is similarity-based approaches that use drug and target similarities. The conventional techniques for predicting drug-target interactions are popular in screening potential drug candidates for further drug action verification. Such similarity-based methods derived from drug-drug and target-target similarities exploring. Drug similarity was tested from the molecules of drugs by using SIMCOMP [12]. Target similarity was computed from the protein sequence by using Smith–Waterman score [13]. In [4], they used the Smith-Waterman score to describe a genomic space and used SIMCOMP score to describe pharmaceutical space. And then, they proposed kernel regression approach called bipartite graph learning to predict drug-target interactions. In [14], they used the drug and target similarities as the support SVM kernels and classified interaction twice and merge the experimental results to provide drug target predictions. KBMF2K firstly map the drug and target spaces to low-dimensional spaces similarity for drug–target interaction prediction [15]. In [16], they developed a network-consistency-based prediction method (NetCBP) to predict drug–target interactions, which rely on drug similarity network and the target similarity network integration. Some above previous works enjoyed very high prediction accuracy. However, similarity-based direction has a basic problem that support data is not a direct biological expression. The similarity just represents another dimension of original drug and target properties that may make experiment only achieve high rate of errors in terms of millions candidates. Even more interesting is feature-based methods have been attempts to use classifier to infer drug-target interactions adopt different encoding schemes impose different descriptors on the protein sequences and compounds. A recent study [17] first try to address drug structures and protein sequence as structure–activity relationship. They use SVM as a classifier to predict DTIs can be regarded as a significant direction even if large-scale calculation is time consuming. Bigram-PSSM take advantage of PAAC descriptors for more accurate prediction [18]. Other studies of more recent interest include integrate heterogeneous biological data [19] or capture rare domain knowledge to predict DTIs [20]. Even though such approaches cannot require as high-dimensional descriptors analysis. To overcome these drawbacks, we propose a novel model to extract large-scale drug-target descriptors and classify output information after a deep representation phase.

III. METHODS

Our method undergoes two main computational steps: the compound-protein interaction representations discovery step briefly describe how extract more meaningful protein sequence attributes. At the representation step, we introduce Stacked Auto-encoder for obtain new representations instead of original high-dimensional protein and compound features. New representations are statistically significant solution to the sparse and large data set. Then, in the classification step, we use these discovered new space sets as input for assigning them as DTIs associated information.

A. Feature vector extraction

Feature extraction usually influences the quality of training data when we analyze large-scale biological data. Some elaborate protein extraction methods have revealed the valuable representations and also have had many remarkable discoveries [21-24]. In order to fully extract the interaction related features, we adopt an advanced multi-scale protein sequence representation method to extract feature vectors from sequences by using binary coding scheme [24]. Normally, an original polypeptide sequence should contain multiple continuous sequence segments which are composed of residues. For collect specific feature vector of protein sequence, we will take multi-scale descriptors to calculate and concatenate each continuous local region by introduced decomposition technique. Based on the actual situation, this approach is able to transform the protein sequences into multi-scale feature vectors which can span several length levels.

Molecular fingerprints are descriptions of drug chemical sub-structures originally introduced to assist in chemical database searching [25]. Our chemical fingerprints set are generated by the PubChem System for encode the 3D structure of a molecule for our computing method. These fingerprints are used by PubChem for similarity neighboring and similarity searching to idealize 3D chemical structure. A fingerprint is an ordered list of binary (1/0) 881 bits in length. Fingerprints property is "CACTVS_SUBGRAPHKEYS" in PubChem and Base64 encoded to provide a textual representation of the binary data.

Each drug is represented by a chemical feature vector $D^{(chem)} = (d_1, \ldots, d_q)^T$, where each element encodes for the presence or absence of each substructure by 1 or 0 and q is the number of fingerprints. Each target protein is represented by a sequence feature vector $T^{(protein)} = (t_1, \ldots, t_p)^T$, where each element encodes for the value of each descriptor range from 0 to 1 and p is the number of descriptors.

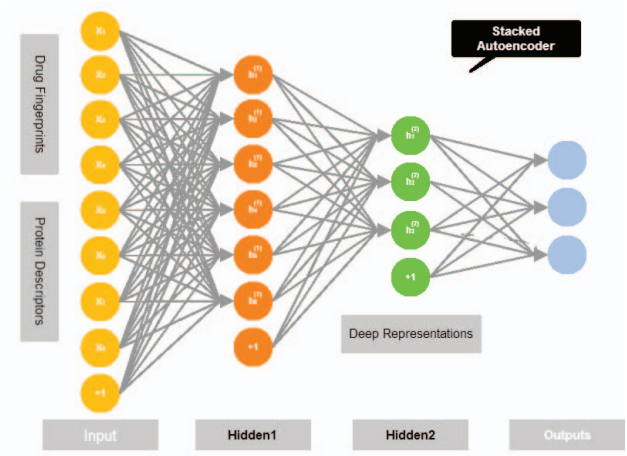

Fig. 2. A Stacked Auto-Encoder composed by two visible layers and two hidden layers

B. Deep representations inferring interactions

After the multi-scale feature of drug and protein collected, known drug–target interactions will be represented by many factors includes the properties of drug compounds and the properties of target protein. Some biochemical effect may be contained in these properties that including structure shape, amino acid composition, hydrophobicity, van der Waals force, Hydrogen bond, Water effects, Metal-ligand interactions and so on. In our method, each drug-target interaction sample will be represented by more than one thousand dimensional vectors. Each drug-target interaction features is made up of chemical substructures and multi-scale protein representations. That is to say, each drug-target interaction can be represented as $DT = (dt_1, dt_1, dt_1, \ldots dt_s)$, where dt_x is the x_{th} drug-target interaction feature combine with d_q and t_p. As with the original large-scale interactions, however, there are sparsity and imbalance issues if we have to deal with new representations directly. Dimensionality reduction has proven to be a useful method deal with large-scale data. The only problem is dimensionality reduction usually loss some important information of input data. Drug discovery is directly influence human body, so any information about the medicine should keep as more as possible. Fortunately, deep learning will keep valuable information after execute the training process. According to [26], deep learning builds multi-layer architecture neural networks and trained with the greedy layer wise unsupervised pre-training algorithms. DNN is about applying the greedy layer-wise unsupervised pre-training mechanism that can reconstruct the original raw data set. We can learn valuable features with deep representation instead of traditional features filtering method. Then, we can use classifier and obtain higher accuracy with better generalization from the learned features. In addition, the risk of fall in a local minimum rather than global minimum problem in traditional training method has been solved by deep network that greedily trained up hidden layer with Auto-encoder at a time. Because of the feature type of our drug data are real numbers and sparse distribution, we choose to stack Sparse Auto-Encoder for build deep architecture of the neural network model.

Stacked Auto-Encoder is a stacked architecture network that applies Auto-Encoder in each layer [27]. In a neural network, each “neuron" in one layer is a computational unit that could be regarded as input vector $X = (x_1; x_2, \ldots, x_n)$ (and a+1 intercept term), and outputs $h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b)$ where a nonlinear function $f : \Re \to \Re$. is activation function. The connections among different neurons in the network can be taken as a weight matrix W. In usual cases of neural network, sigmoid function is normally using $f(z) = \frac{1}{1+exp(-z)}$. In fact, a conventional Auto-Encoder would endeavor to learn a function $h_{W,b}(x) \approx x$ which means it is discovering an approximation to the identity function, so as to output an approximate outcome $\hat{x}$. The identity function seems a typically trivial function trying to learn but by placing constraints on the network. For deep

represent DTIs information, we can discovery useful structure of drug-target interactions data from limiting hidden units. Take our multi-scale features as examples, $DT = (dt_1, dt_1, dt_1, \ldots dt_s)$ is defined as input vector X. Suppose the original feature representations are collected from a 1448-dimensional feature space, i.e. $x \in \Re^{1448}$ which means there are 1448 visible input units. If we set that there are 600 hidden units in the hidden layer1, according to the requirement $h_{W,b}(x) \approx x$, the next layer need to learn a compressed representation of the input. This also means that hidden layer will start to reconstruct the 1448-dimensional input x by given vector of hidden unit activations $a^{(2)} \in \Re^{600}$. Access to the multi-scale DTIs data could then be transformed to deep representations through the reconstruction process, instead of the high-dimensional and noise visible units. Because there is an interesting structure hide in the input data like two input features have relationship or such as in the feature space of drug chemical structure and target protein interactions. Otherwise, this reconstructive function wouldn't work if the inputs features were completely random, i.e., each x_i is independent of the other features. The overall cost function of non-sparse Auto-Encoder can be defined as:

$$J(W,b) = \left[\frac{1}{m}\sum_{i=1}^{m} J(W,b;x^{(i)},x^{(i)})\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2 \quad (1)$$

$$= \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}||h_{W,b}(x^{(i)}) - x^{(i)}||^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^{(l)})^2$$

The first term in $J(W,b)$ is an average sum-of-squares error term, where m is the training samples number. The second term is a regularization for prevent over-fitting, where λ be

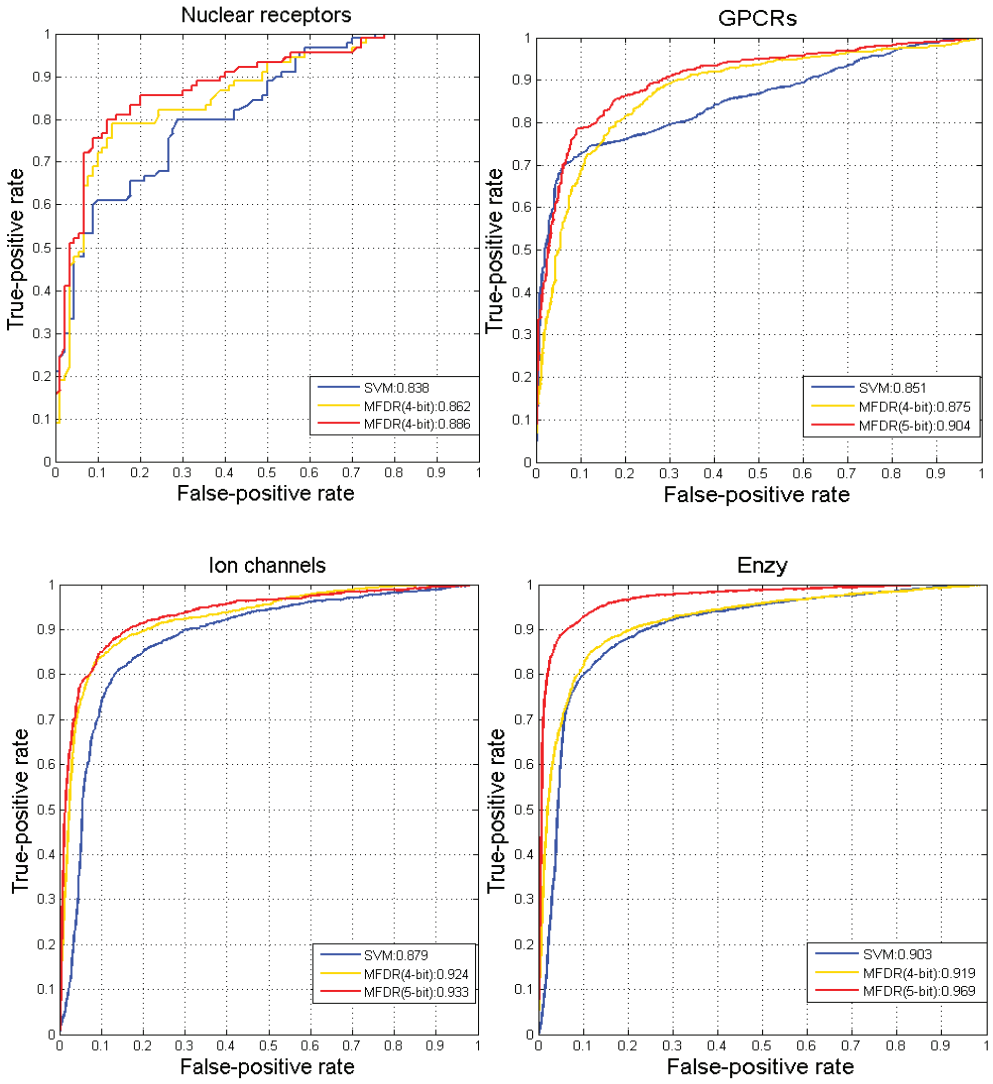

Fig. 3. ROC curves of four different drug-target interaction predictions

supposed to control the relative importance of the two terms. Normally, Auto-Encoder is aim to minimize Equation (1) for that output $h_{W,b}(x) \approx x^{(i)}$ can approximate the raw data $x^{(i)}$ as much as possible. Further, large hidden units still could be used to discover valuable information if we impose a sparsity constraint on the hidden unit [28]. Sparse Auto-Encoder try to keep the output mean value of hidden layer to 0 which means most neurons are considered to be inactive. The overall cost function of Sparse Auto-Encoder can be defining as:

$$J_{sparse}(W,b) = J(W,b) + \beta \sum_{j=1}^{s_2} KL(\rho||\hat{\rho}_j) \qquad (2)$$

Where

$$\hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m}[a_j^{(2)}(x^{(i)})] \qquad (3)$$

$$KL(\rho||\hat{\rho}_j) = \rho \log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j} \qquad (4)$$

where ρ is a sparsity parameter, s_2 is the number of the hidden neurons and β controls the weight of the sparsity penalty term. Eq. (3) is average activation of hidden unit j and we need to enforce the constraint $\hat{\rho}_j = \rho$. Eq. (4) is the Kullback-Leibler divergence between a Bernoulli random variable with mean ρ and a Bernoulli random variable with mean $\hat{\rho}_j$.

Specifically, we should stack Sparse Auto-Encoders layer by layer to a whole deep network. A typical two hidden layers Stacked Auto-Encoder structure diagram is shown in Figure 2 describes the main procedure of proposed model. As in Fig.2, the input layer is visible layer that take original data set. In every hidden layer, the neurons receive data from previous layer, and then they compute the received data through an Auto-Encoder. In the end of each hidden layer, the neurons output the computed new features to the next hidden layer or visible layer. After a deep network processing, the original data set will represent by deeper feature spaces layer by layer. Therefore, Stacked Sparse Auto-Encoders is able to learn enriched representations from the large-scale original data sets.

TABLE I
STACKED AUTO-ENCODER PARAMETERS

Parameter	Value
Neurons in hidden layer 1	**600**
Neurons in hidden layer 2	**200**
Beta (weight of sparsity penalty term)	**5**
Sparsity (desired average activation of the hidden units)	**0.05**

C. Prediction model

After we bring in Stacked Auto-Encoder as an unsupervised learning model to get new representations, one effective classifier will be used to predict whether a given drug-target interaction is positive or not according to the gold standard dataset. SVM (Support Vector Machine) will be used as the classifier for build our predicting model. SVM is a popular classification algorithm originally developed by Vapnik et.al and it has been proved extremely effective in chemical and biological classifications [7] [29]. The basic procedure of utilizing SVM model for DTIs prediction can be described briefly as follows. Firstly, we use the open source package from LIBSVM [30] to implement SVM. Then, SVM maps the inputted drug and target original representations space X into a high dimensional feature space F with a linear algorithm due to the linear relations exist in training data. After that, the SVM model will find out an optimized linear division within the feature F. According to our test and previous experiences, Radial Basis Functions (RBF) kernel is the best kernel selection of the traditional kernel, especially it's appropriate for the high-dimensional data sets and has better boundary response.

IV. RESULTS

A. Data Preparation

In this study, the data which are used to predict DTIs come from [4]. There is a drug-target interactions gold standard dataset formed by four types of DTIs, which includes enzymes, ion channels, GPCRs, and nuclear receptors. After interactions collection, the final number of positive drug-target interactions in the gold standard data set are 2926, 1476, 635 and 90, respectively. Each category being further organized in drugs are 445, 210 223and 54, respectively and the protein numbers are 664, 204, 95, 26 with four categories. Table 2 shows the statistics of used data set.

Suppose that we have a set of n drugs with biological profiles of m target proteins. To encode drugs features, chemical structures of drug compounds are extracted from PubChem database which use a fingerprint corresponding to t

TABLE II
DRUG-TARGET DATA STATISTIC

Type		Ion channel	Enzyme	GPCR	Nuclear receptor
Drugs		210	445	223	54
881 bits					
Target proteins		204	664	95	26
567 Descriptors	1449 Descriptors				
Positive Drug–target Interactions		1476	2926	635	90

TABLE III
COMPARSION FOR THE AUROC OF THE MFDR
VERSUS OTHERS ON THE FOUR DATASET

Data set	Feature-based			Similarity-based		
	MFDR	Cao, D.S. et al.	Bigram-PSSM	Bipartite Graph Learning	KBMF2K	NetCBP
Nuclear receptors	**0.886**	0.882	0.869	0.692	0.824	0.839
GPCRs	**0.904**	0.890	0.872	0.811	0.857	0.823
Ion channels	0.933	0.942	0.889	0.692	0.799	0.803
Enzymes	**0.969**	0.948	0.948	0.821	0.832	0.825

881 chemical substructures. Each drug was represented by an 881 dimensional feature vector $D^{(chem)} = (d_1, \ldots, d_{881})^T$, where each element encodes for the presence or absence of each substructure by 1 or 0, respectively. To encode protein features, protein sequence are extracted from multi-scale local descriptor feature representation scheme. In terms of test larger descriptors influence, all descriptors calculated in 4-bit and 5-bit are concatenated. For 4-bit binary form, each sequence was represented by a 567 dimensional vector $T^{(protein)} = (t_1, \ldots, t_{567})^T$. For 5-bit binary form, a total 1449 dimensional vector $T^{(protein)} = (t_1, \ldots, t_{1449})^T$ has been built to represent the protein sequence. Each element encodes for the value of each descriptor range from 0 to 1.

On present understanding, known negative DTIs samples are generally much larger than the positive DTIs samples. Because the size of non-interactions is not comparable with the size of positive interactions, some works may take a high true negative result by the significant larger negative samples. To avoid such problem, previous feature-based approaches have randomly selected negative samples from the non-interactions until the ratio hitting one-to-one scale. We considered all real positive drug–target interactions and randomly selected same negative sample as many as the positive samples like [17]-[18]. In summary, the original features number can be extracted from MFDR of 4-bit is 1448

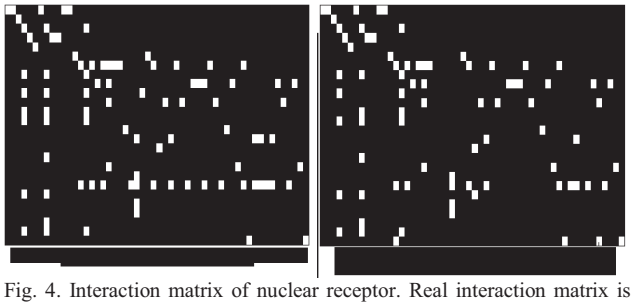

Fig. 4. Interaction matrix of nuclear receptor. Real interaction matrix is known drug-target interactions of nuclear receptor, predicted interaction matrix is generated by MFDR. White pixels represent the positive interactions, whereas Black pixels represent the negative interactions.

that comprise 881 chemical substructures and 567 protein descriptors. And so on, each of the drug-target interaction has 2330 features which extracted from MFDR of 5-bit.

B. Performance Evaluation

Inasmuch as our method like evaluates deep representations based on their profiles with protein and molecules, which in turn refer to specific sub actions, we use a 2-larger Stacked Sparse Auto-encoder model to rebuild the drug and target features. The first hidden layer of our model is composed of 600 hidden units while the second hidden layer is composed of 200 hidden units. Which mean, there are 200 deep representations come out from thousands original features. Table.1 shows the parameter configuration of the Stacked Sparse Auto-Encoder model. We performed the fivefold cross-validation to split gold standard data into five subsets of equal size. Each subset was then taken in turn as a test set, and we performed the training on the remaining four sets. We used the grid search to select the best regularization parameter C and the kernel parameter γ for the radial basis function (RBF) based on the overall accuracy. In this study, the performance of MFDR was mainly evaluated by using ROC. The ROC (receiver operating characteristic curve) demonstrates the true-positive rate and false-positive rate of the experimental result, where true-positives are the number of correctly predicted drug-target interactions while the false-positives are the number of not correctly predicted drug-target interactions. AUC refers to the area under ROC curve which is an important measure which can be used for evaluating the classification accuracy.

In view of the larger number of protein sequence descriptors are able to reflecting structures more real [22], we followed two different regions outlined by 4-bit and 5-bit extraction. The resulting AUROC scores of MFDR(4-bit) for enzymes, ion channels, GPCRs, and nuclear receptors are 0.919, 0.924, 0.875 and 0.862 respectively. The resulting AUROC scores of classical SVM without deep representation are 0.903, 0.879, 0.851 and 0.838. As the Figure 3 shows, our prediction accuracy are higher than the classical SVM that proved the multi-scale feature deep representations can

improve the DTIs prediction while reduce high dimensional features. We also giving the AUROC scores of MFDR(5-bit) are 0.969, 0.933, 0.904 and 0.886 respectively. It proved that MFDR have chance to improve performance under much larger range of descriptors even project to same low-dimensional representations. In addition, we believe MFDR should get a significantly better performance in terms of higher bit binary coding.

To evaluate the performance of the method in comparison to previous work, we considered three important studies in this similarity-based area and two feature-based studies. We compared the best AUROC scores of the MFDR with these approaches including KBMF2K [15], NetCBP [16], Bipartite Graph Learning [4], Bigram-PSSM [18] and the proposed method by Cao, D.S. et al. [17]. Table 3 shows the AUROC scores of MFDR and others divided by four interaction types. As the results look like those shown in above, the prediction accuracy of the MFDR is superior in comparison with most methods. We go further compared the predicted interactions of the nuclear receptor to real interactions as Figure 4 shown. As there is a high coincidence of the bright pixels to the predicted matrix and the real one, we can claim the deep representations successfully kept values from original large descriptors.

V. CONCLUSION

In this work, a new prediction model was developed to inferring drug-target interactions. We adopt the multi-scale optimization theory to extract the drug and protein details from limited biological information. Deep representation approach also introduced in our method for retaining the realistic biological properties and reducing the high-dimensional features. This is the first time that deep learning was used to predict drug-target interactions. The key aspects of the DTIs prediction model reflect a feasible way in mapping the large-scale drug-target descriptors to lower-dimensional features rationally. The proposed Stacked Auto-encoder model is able to generate representations of the multi-scale data layer by layer. In the last step, our model successfully reconstructs the representative features from the stacked hidden layers and builds a SVM as the final classifier. We gathered several kinds of DTIs datasets that we used to train deep representation model. The experimental result shows that MFDR is able to handle the large scale pharmacological data effectively and improve the performances of drug-target interaction prediction model. Our work can provide some important mechanisms of drug-target interaction to make the drug discovery navigation simpler. In addition, deep learning successfully transfer high dimensional data to a relatively lower dimensional coupled description which make our model more sensitively reflect real actions. It has been proved in a large-scale drug-target interactions and it should have ability to solve other large biological data problem in the future.

REFERENCES

[1] Nagamine N, Shirakawa T, Minato Y, et al. “Integrating statistical predictions and experimental verifications for enhancing protein-chemical interaction predictions in virtual screening, ” PLoS Comput Biol2009;5(6):e1000397

[2] Parsons, Ainslie B., et al. "Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways," Nature biotechnology 22.1 (2004): 62-69.

[3] Hao Ding, Ichigaku Takigawa, Hiroshi Mamitsuka, and Shanfeng Zhu. “Similarity-based machine learning methods for predicting drug–target interactions: a brief review,” Briefings in Bioinformatics, page bbt056, 2013.

[4] Yamanishi Y, Araki M, Gutteridge A,et al. “Prediction of drug-target interaction networks from the integration of chemical and genomic spaces,” Bioinformatics2008;24(13):i232–40.

[5] Xia Z, Wu LY, Zhou X, et al. “Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces,” BMC Syst Biol2010;4(Suppl 2):S6.

[6] Van Laarhoven T, Nabuurs SB, Marchiori E. “Gaussian interaction profile kernels for predicting drug-target interaction,” Bioinformatics 2011;27 (21):3036–43.

[7] Zhu-Hong You et.al "A MapReduce based parallel SVM for large-scale predicting protein–protein interactions," Neurocomputing, Volume 145, 5 December 2014, Pages 37–43

[8] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, 2006.

[9] Bengio, Yoshua, Aaron Courville, and Pierre Vincent. "Representation learning: A review and new perspectives," Pattern Analysis and Machine Intelligence, IEEE Transactions on 35.8 (2013): 1798-1828.

[10] Spencer, Matt, Jesse Eickholt, and Jianlin Cheng. "A Deep Learning Network Approach to ab initio Protein Secondary Structure Prediction," Computational Biology and Bioinformatics, IEEE/ACM Transactions on 12.1 (2015): 103-112.

[11] Nguyen, Son P., Yi Shang, and Dong Xu. "DL-PRO: A novel deep learning method for protein model quality assessment," In Neural Networks (IJCNN), 2014 International Joint Conference on, pp. 2071-2078. IEEE, 2014.

[12] Masahiro Hattori, Yasushi Okuno, Susumu Goto, and Minoru Kanehisa. “Heuristics for chemical compound matching,” Genome Informatics Series, pages 144–153, 2003.

[13] Smith, Temple F., and Michael S. Waterman. "Identification of common molecular subsequences," Journal of molecular biology 147, no. 1 (1981): 195-197.

[14] Yu, Hua, et al. "A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data," PLoS One 7.5 (2012).

[15] Gönen, Mehmet. "Predicting drug–target interactions from chemical and genomic kernels using Bayesian matrix factorization," Bioinformatics 28, no. 18 (2012): 2304-2310.

[16] Chen, Hailin, and Zuping Zhang. "A semi-supervised method for drug-target interaction prediction with consistency in networks," PLoS One. (2013): e62975.

[17] Cao, Dong-Sheng, et al. "Large-scale prediction of drug–target interactions using protein sequences and drug topological structures," Analytica chimica acta 752 (2012): 1-10.

[18] Mousavian, Zaynab, et al. "Drug–target interaction prediction from PSSM based evolutionary information." Journal of pharmacological and toxicological methods 78 (2016): 42-51.

[19] Harvard Yamanishi, Yoshihiro, Masaaki Kotera, Minoru Kanehisa, and Susumu Goto. "Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework," Bioinformatics 26, no. 12 (2010): i246-i254.

[20] Luo, Weimin, and Keith CC Chan. "Discovering patterns in drug-protein interactions based on their fingerprints," BMC bioinformatics 13.Suppl 9 (2012): S4.

[21] Liu B., Wang S. & Wang X. “DNA binding protein identification by combining pseudo amino acid composition and profile-based protein representation,” Scientific reports 5, 15479 (2015).

[22] Liu, Bin, Deyuan Zhang, Ruifeng Xu, Jinghao Xu, Xiaolong Wang, Qingcai Chen, Qiwen Dong, and Kuo-Chen Chou. "Combining evolutionary information extracted from frequency profiles with sequence-based kernels for protein remote homology detection," Bioinformatics 30, no. 4 (2014): 472-479.

[23] Jiancang Zeng, Dapeng Li, Yunfeng Wu, Quan Zou, Xiangrong Liu. “An empirical study of features fusion techniques for protein-protein interaction prediction,” Current Bioinformatics. 11.1 (2016).

[24] You, Z. H., Chan, K. C., & Hu, P. "Predicting Protein-Protein Interactions from Primary Protein Sequences Using a Novel Multi-Scale Local Feature Representation Scheme and the Random Forest," PLoS One. (2015): e0125811.

[25] Kauvar LM, “Predicting ligand binding to proteins by affinity fingerprinting,” Chem Biol. 1995 Feb; 2(2):107-18.

[26] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, 2006.

[27] Bengio, Yoshua, et al. "Greedy layer-wise training of deep networks,"Advances in neural information processing systems 19 (2007): 153.

[28] Ng, Andrew. "Sparse autoencoder." CS294A Lecture notes 72 (2011).

[29] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks," Machine learning 20, no. 3 (1995): 273-297.

[30] Harvard Chang, Chih-Chung; Lin, Chih-Jen "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology 2 (3). (2011).