International Journal of Applied Sciences: Current and Future Research Trends
(IJASCFRT)
ISSN (Print) , ISSN (Online)

https://ijascfrtjournal.isrra.org/index.php/Applied_Sciences_Journal

A Deep Learning Multi-Class Model for Drug-Target Binding Affinity Prediction

Nassima Aleb*

Computer Science Department, Jubail University College, Jubail Industrial City, Kingdom of Saudi Arabia
Email: alebn@ucj.edu.sa

Abstract

Drug design and discovery is a very challenging and costly process. It involves a crucial phase of drug-target interaction (DTIs) identification. Nevertheless, most existing methods use either binary classification to predict the presence of an interaction in a Drug-Target pair, or regression methods to predict the exact float-value representing the Binding Affinity. These latter methods are more valuable but suffer from unsatisfactory results despite their very sophisticated models and multiple inputs. In this paper, we present a new approach for predicting the strength of drug-target binding, we tackle the question as a Multi-class classification problem. This approach is very rational since the key points, in drug-target interaction, are to have a precise indication about the binding strength and to establish a ranking between drug-target pairs' binding strengths. Our model input being sequences presenting hidden patterns, we use convolutional LSTM networks, since they inherit the ability in discovering patterns from Convolutional networks, and learning from sequential data from recurrent networks. Besides the usual performance metrics, we investigate new interesting performance metrics that have never been explored before. The results show that our approach is very convincing.

Keywords: Deep learning models for drug discovery; Drug repurposing; Drug-target binding affinity; Discretization; Multi-class classification models.

1. Introduction

Detecting the presence of interactions between a drug and a target is decisive in drug discovery and repositioning. Nevertheless, the existing biological assays and other screening methods, remain very expensive, and time-consuming [1]. Consequently, computational methods have been explored to predict potential drug-target interaction with a minimum error rate. These methods are classified into two categories: DTI (Drug-Target Interaction) class, and DTBA (Drug-Target-Binding-affinity) class [2].

---

* Corresponding author.

DTI methods are merely interested in a binary classification model since their objective is limited to the detection of the presence or absence of an eventual interaction between a drug and a target without considering the strength of such interaction. This constitutes an unrealistic simplification of the problem. Furthermore, most used datasets, in this class, contain exclusively positive examples, missing values are interpreted as negative cases which is inaccurate [3]. In drug discovery, a drug-target binding having a weak strength is not suitable. Consequently, models and approaches predicting the strength of the binding between a drug and a target are of great importance. However, to date, DTBA prediction methods did not achieve the ultimate goal, even by using complicated models and multiple input forms representations for the same model. In this paper, we present a novel deep learning based approach for drug-target affinities prediction. MC-DTBA (Multi-Class DTBA) is based on the idea that both DTI and DTBA are too drastic. They are at the two opposite ends. Our idea is to tackle this problem as a multi-class classification problem, where classes represent different levels of strength and are defined based on the binding affinities values. This provides a ranking of drug-target pairs binding affinities, which constitutes the actual need. Our idea is justified by the following arguments:

1) DTI methods are not judicious. Predicting the presence of weak interaction is misleading since in drug discovery such interactions are equivalent to the absence of interactions.
2) The main objective in drug discovery is to identify the level of binding strength between a drug and a protein [4], to extract the highest scoring compounds, called "hits". An in vitro testing is applied to these hits resulting in leads identification. Leads are selected compounds that are filtered for supplementary testing and investigation. The most promising leads will be considered as drug candidates [5]
3) Results found, so far, with DTBA methods are not very convincing even with high complexity systems
4) Datasets, used by almost all the methods of DBTA, have different techniques to compute the affinities values. Furthermore, for the same dataset, there may exist many versions resulting from ad-hoc manipulations. The Kiba dataset [42] is an example of such 'unjustified' manipulations [6]. This makes binding values prone to errors or at the best not very accurate.
5) Given the level of randomness and the huge number of exogenous parameters that intervene in the drug-target binding occurrence and quality, it is more needed to represent the strength with a level or a "ranking" rather than with a very precise float-value.

A. Related Works

As we said before, we are not aware of any previous work tackling the drug-target interaction prediction problem in a multi-class classification fashion. Hence, in the following, we will present some previous work in DTBA, because, their objective is closer to that of our approach. Few methods are focusing on DTBA prediction. These methods are generally data-driven and usually use machine learning techniques for regression rather than binary classification. Machine learning methods such as Random Forest (RF) algorithm have been used as an effective substitute for scoring functions [7-9]. However, RF-score was unsuccessful in virtual screening and docking tests because the oversimplification of the drug-target complex description causes a significant loss of information [10]. SimBoost [11] and KronRLS [12] are two state-of-the-art methods for DTBA prediction. SimBoost is based on features construction, it exploits drugs and targets similarity matrices [11]. KronRLS is based on Regularized Least Squares [12]. These two methods can predict continuous values

representing binding affinity scores, and binarized values indicating the presence or absence of interaction. However, the drawback of these methods and other similar ones, is that they are either based on similarities or require expert knowledge. Accordingly, this category of methods requires the task of "feature engineering" in their process. This phase is necessary for all machine learning approaches and constitutes one of the supremacy points of Deep Learning (DL) over machine learning techniques. Deep neural networks can automatically extract important features from the raw input data. They are able of detecting hidden patterns, associating low-level features to high-level features, and capturing complex nonlinear relationships in a dataset [13]. Deep learning has witnessed unprecedented growth in popularity in recent years. It has seen tremendous successful applications in image processing and computer vision [15,16], speech recognition, natural language processing [17-19]. Stimulated by their remarkable success, deep learning-based techniques are now being investigated in many other complex domains, including bioinformatics such as in genomics studies [20-22] and quantitative-structure activity relationship (QSAR) studies in drug discovery [23,24]. The most valuable advantage of deep learning approaches is that they don't require a prior feature extraction phase, they can extract automatically latent features of the raw data by non-linear transformations in each layer [25] and thus they excel in learning hidden patterns in the data. Consequently, DL methods overcome the limitations of the previously applied techniques. DL algorithms for DTBA prediction, usually show the best performance compared to other ML and conventional algorithms. We can categorize DL-based algorithms for DTBA prediction, regarding two main aspects. The first is the input data representation, especially, drug features. Some examples are the Simplified Molecular Input Line Entry System: SMILES [26], Ligand Maximum Common Substructure (LMCS), and Extended Connectivity Fingerprint (ECFP) [27]. Some methods use a combination of these features. The second aspect is the architecture of the model that is defined using various neural network (NN) types. Usually, the representation of the input data for the drug and target starts the DTBA prediction, next to this, various NN types are applied to learn independently their features. The obtained features are then merged and fed to a neural network for the prediction task. Stacked auto-encoder models based on Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) were among the earliest deep-learning approaches for DTBA prediction [33]. The main objective of these models was to define real-valued vector forms for chemical and genomic representation. [28-32]. CNN networks have a well-known reputation for pattern recognition, especially when dealing with multi-dimensional structures. Consequently, DTBA prediction methods based on CNN, have also been investigated for 3D learning from drug-target complexes structures [33-35]. DeepDTA, introduced in [33] is based on Convolutional Neural Network (CNN) architecture that includes blocks of convolutional layers that learn latent features of both drug and target, followed by pooling layers reducing the size. DeepDTA performance was assessed on two datasets. In an attempt of enhancing DeepDTA performance, the same authors developed two other models investigating the use of various input representations with different models. The first model is called WideDTA [34]. Like its predecessor, WideDTA is also a CNN-based model. It attempts to improve the performance of DeepDTA by considering more input representations. The WideDTA investigated multiple inputs. For drugs, it used a combination of ligand SMILES and Ligand Maximum Common Substructure (LMCS). Whereas for proteins, it used Amino-acid sequences along with Protein Domains and Motifs (PDM). Furthermore, all the sequences are represented as a set of words instead of residue sequences. The second model is DeepAffinity [35]. It is an approach based on seq2seq model. DeepAffinity uses the SMILES representation for drugs and the structural property sequence (SPS) for protein

representation. Drug SMILES and protein SPS are both encoded into embedding representations and provided to a seq2seq model. PADME [6] is another DL-based approach. It uses drug and target features and fingerprints to predict the binding affinity values. There are two versions of PADME. One called PADME-ECFP uses the Extended-Connectivity Fingerprint [36] as input representation for drugs. The second version called PADME-GraphConv is based on convolutional graphs [37]. The two versions use the same protein descriptors [38]. As mentioned by the authors of the paper presenting PADME, a surprising result is that sequence representations outperformed graph-based ones, which was not expected. The use of sequences for drug-target prediction was also assessed in much other research and has proven to be promising.

B. Contributions

In this paper, we propose a new DL-based approach for drug-target affinities prediction. Our method uses SMILES (Simplified Molecular Input Line Entry System) representation for compounds. and Amino-acid sequence representation for proteins. We use the same datasets as DTBA approaches. First, based on the values distribution of the used datasets, our method uses the Elbow method to define the number of appropriate clusters for each dataset. Then, the found number of clusters NC, is used with the Kmeans clustering algorithm to partition the dataset into NC distinct non-overlapping clusters where each data point belongs to a unique cluster representing a class in the Multiclass model. In this way, our method overcomes the limitations of both DTI and DBTA existing methods by predicting interaction strength levels instead of binary class labels, and instead of a float-value.

We use two CONVLSTM2D blocks that learn drug and protein features and attempt to extract their patterns and sequential relationships. The learned representations are concatenated and fed to a fully connected layer block to predict the drug-target binding strength level. Using CONVLSTM2D layers is justified by the need of both capturing local residue patterns and learning from sequential data while ensuring robustness against the problems of long-term dependency. The first capability is provided by CNNs while the second is a well-known quality of LSTMs. We conducted extensive experiments with usual and new evaluation metrics. The results show that the proposed model is an effective approach for drug target binding affinity prediction. The subsequent sections of this paper are organized as follows. The Material and Method section presents the datasets that we used, compound and protein representations, and the discretization method. It also introduces the network architecture. The Experiment section describes the experiments conducted. It introduces the baseline method, performance metrics, experimental design, and the experimental results and clarifies some implementation and design choices. Finally, the last section concludes the paper.

2. Materials And Methods

A. Datasets

For DTI binary prediction, the most popular benchmark datasets are Yamanishi datasets [39]. However, they cannot be used for DTBA prediction models, since they don’t contain actual binding affinity values that are needed to train DTBA models to predict the strength of the binding between drugs and their targets. Only a few

benchmark datasets have been used to develop in silico DTBA prediction methods [1]. The most used datasets for these models are the three large-scale benchmark datasets: Davis, Metz, and KIBA (Kinase Inhibitor BioActivity). These datasets are introduced in [40-42]. All three datasets are large scale biochemical selectivity assays of the kinase inhibitors [1]. The Davis dataset contains selectivity assays of the kinase protein family and the relevant inhibitors with their respective dissociation constant (Kd) values. It contains interactions of 442 proteins and 68 ligands. In the three datasets, the binding affinity is computed differently. In the Davis dataset, the Kd value is directly provided as a measure of binding affinity. KIBA score represents a continuous value of the binding affinity that was calculated utilizing Kd, Ki, and IC50 scores [42]. The KIBA dataset initially contained 467 targets and 52 498 drugs, it was filtered in [11] to include only drugs and targets with at least 10 interactions. The obtained final version contains a total of 229 unique proteins and 2111 unique drugs. The higher KIBA score indicates a lower binding affinity. The Metz dataset provides the Ki as a measure of binding affinity. A small value of Ki indicates a strong binding affinity between a drug and its target [41]. However, these datasets have been curated more than once [12]. Consequently, there exists more than one version of them. In this paper, we use the same version used in PADME [6]. Table 1 summarizes the statistics for these three benchmark datasets.

B. Discretization

The first step of our method is the discretization process. It is the mapping of the original regression problem into a classification one. In [43,44], it has been clearly shown that "it is possible to obtain excellent predictive results by transforming regression problems into classification ones. Mapping regression into a classification is a kind of pre-processing technique that enables the use of classification algorithms on regression problems." [43,44]. This could be justified by the fact that due to infinite degrees of freedom, continuous features have a reduced chance of correlating with the target variable. Furthermore, discretizing a model has the effect of reducing significantly the impact of noise, which is the small fluctuation in the data. Using discretization requires the creation of a dataset with discrete classes rather than continuous values. Two families of approaches are usually used for performing this task. The first category includes unsupervised methods: Equal-Width; Equal-Frequency and K-Means. the second one contains supervised methods like Decision Trees. In our work, we used the K-Mean discretization method. It consists in applying the K-Means clustering algorithm to the continuous values to divide them into discrete clusters each one having a centroid. Before applying the K-means method, we used the Elbow method to investigate the optimal number of classes for each dataset, we searched for these values in the range (1,15), to ensure a flawless coverage. The graphs representing the inertia based on the clusters number are given in Figure 1 column (a). The optimal number of clusters, for each dataset is determined by selecting the value of k at the "elbow", that is, the point after which the inertia value starts a linear decreasing. We notice that for all the datasets, the optimal value of clusters was 6. The obtained clusters with their centroids are given in Figure 1 column (b).

Table I: Datasets Summary.

Dataset	Drugs	Proteins	Known DTIs
Davis	68	442	30056
Kiba	2116	229	118254
Metz	1412	156	35259

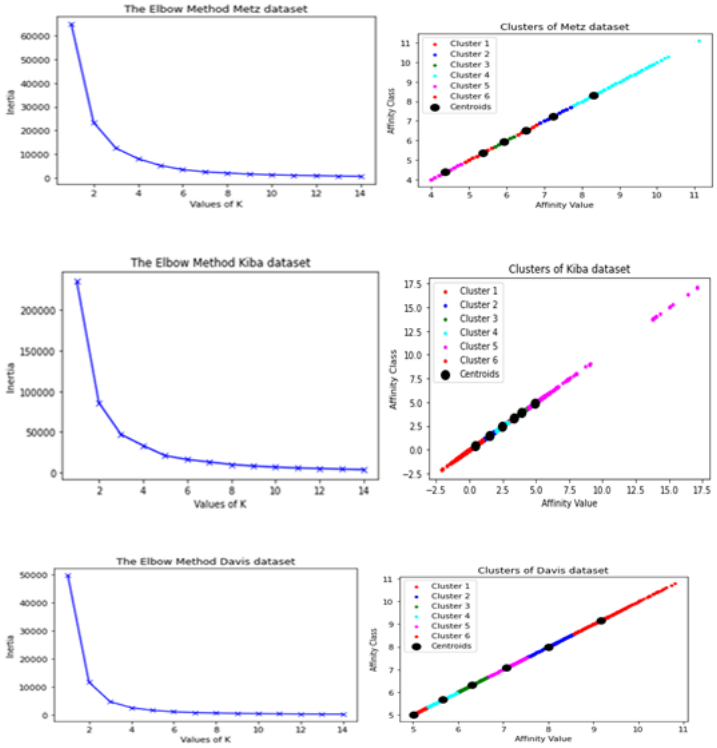

Figure 1: Discretization process: (a): The Elbow (b):Clusters Centroids.

C. Drug and Target Representation

MC-DTBA uses SMILES for drug representation and amino-acid protein sequences for proteins. The Simplified Molecular-Input Line-Entry System (SMILES) is a specification in the form of a line notation that uses printable characters for describing the structure of chemical elements: molecules and reactions [26]. SMILES is a true language, although with a small vocabulary size (atom and bond symbols) and only a few grammar rules. In our model, SMILES are represented by a one-hot encoding. Protein sequences are encoded similarly. The lengths of SMILES and protein sequences are variable, consequently, for each dataset, we decided on a maximal length. The sequences that are shorter than the maximum length are 0-post-padded, while longer sequences are truncated. For all the datasets, we opted for a length Accordingly to the distribution of SMILES and protein sequences lengths, we opted for the maximal length of 100 for SMILES and 1200 for protein sequences.

D. Proposed Model

MC-DTBA is based on convolutional recurrent neural networks. CONVLSTM2D is a special architecture that combines the gating of LSTM with 2D convolutions. In these layers, input transformations and recurrent transformations are both convolutional. Convolutional networks are famous for their ability in discovering patterns and LSTMs are a class of recurrent neural networks that excel in learning from sequential data while avoiding the problems of long-term dependency since they encompass memory blocks allowing efficient long sequences learning. These characteristics justify our use of CONVLSTM2 in DBTA prediction both drugs and targets are represented by sequences presenting patterns. Therefore, we opted for this type of layers. First a one-hot encoding is applied to each sequence. Then, an embedding layer is used to represent sequences characters with high-dimensional (128-dimensional) dense vectors. The resulting sequences are deeply explored by two convolutional recurrent blocks, each one composed of three CONVLSTM2D layers to capture the presence of discriminative features. These layers are directly connected without pooling which allows preserving the entire information, a pooling is performed at the end of the block, to reduce the output size of the previous layers and provide a generalization of the learned features. The output of the convolutional recurrent blocks are concatenated. To conclude, the final prediction is performed, in a standard way, by using fully connected layers after the feature extraction. All the hidden layers are activated by the "relu" activation function. The output layer was activated with the softmax () function since the problem is framed as a multiclass classification problem. The whole neural network model was implemented with Keras [45]. The most powerful feature of CONVLSTM2D models is their ability to capture the local dependencies with the help of filters and at the same time find out dependencies between distant sequence locations. The number and size of filters in a CONVLSTM2D are among the hyperparameters that have a high impact on the model performance. Thus, increasing the number and the size of filters likely increases the ability of the model in patterns recognizing patterns [46]. Stacking many CONVLSTM2D layers allows the automatic detection of more abstracted features. The fully connected block is constituted of three Dense layers. The first two ones have, each one, 1024 nodes the last one contains 512 nodes. To prevent overfitting, we make extensive use of Dropout with various proportions. Finally, the prediction layer contains one node and is activated with the 'sigmoid' activation function. For regularization, we used Dropout and Batch Normalization techniques [47] to avoid overfitting. The activation functions used for fully connected layers are all Rectified Linear Units (ReLU) that have been widely used in deep learning studies [25]. We used categorical cross-entropy as a loss function. We used a mini-batch size of 128 to update the weights of the network. The learning was completed with 100 epochs. We utilize the Adadelta optimizer [47]as the optimization algorithm to train the networks. To have additional control on the learning rate, we use a *Callback* for initializing the learning rate and automatically tuning it during the training and reduce its value based on the monitored parameter, the initial learning rate value was set to 0.01, and the minimum value was set to 0. 0001. The structure of the network is shown in Figure 2. MC-DTBA performance was assessed for the three datasets.

3. Experiments

In this section, we describe the experiments we conducted for our proposed model, which is a deep learning multi-class model for drug-target binding affinity prediction method taking advantage of great performances of

both recurrent and convolutional networks.

E. Baseline Method

There are no previous methods that could be used directly as a baseline since all the previous approaches are either binary classification or regression models. However, to assess the performance of our method, and compare it with existing models, we have selected one state of the art approach which is DeepDTA, for which we have transformed the last layer in such a way to output a set of classes instead of a unique value. We selected DeepDTA as a baseline for many reasons:

1) DeepDTA is a state of the art method that is usually used to assess methods against
2) DeepDTA model has been trained on two datasets that are used in our study, this make the performance comparison more equitable.
3) Contrary to other models, both our approach and DeepDTA are using SMILES and protein sequences as inputs.

DeepDTA uses two distinct stacked CNN blocks, one for protein sequences and one for SMILES to predict affinity between target protein and drug. We used DeepDTA with the same code and optimized hyperparameters provided in their paper [33]. The unique modification we performed, was in the last layer to make it predict a class rather than a float-value. This method is followed by many approaches to perform comparisons.

F. Evaluation Metrics

Categorical Accuracy is the usual performance metric in a multi-class classification problem. However, many additional metrics might be of great interest for the study of drug-target bindings. These metrics have not been used in the previous papers because of their problem framing. In fact, for the models that specify the problem as regression the natural metrics are various forms of

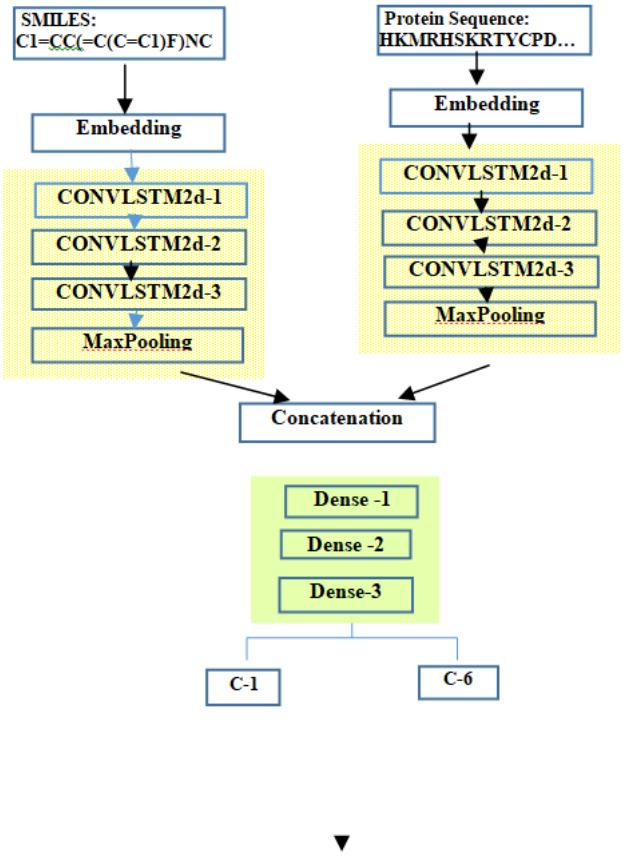

Figure 2: Model Architecture.

errors: mean square error, mean average error, and others. On the other side, the models where the problem is framed as binary classification rely on the usual metrics in this domain which are: accuracy, precision, recall, AUC and other similar metrics. To assess the performance of our model, we have selected four metrics, the first being the usual categorical accuracy. We used three other metrics that are suitable for our study: concordance index [48], TopK-Categorical-Accuracy, and Hinge-Categorical metric [49]. In the subsequent, we will introduce briefly each of these metrics and explain its appropriateness for our work, and when necessary, how we implemented it in our model.

1) Concordance index

The concordance index or C-index (or CI) can be considered as a generalization of the area under the ROC curve (AUC) that is usually used in binary classification. It represents the global assessment of the model discrimination power: this is the model's ability to correctly provide a reliable ranking of the drug-target pairs. The concordance index of a set of data considers each data pair and calculates the probability that the order of the predicted label values is the same as the order of the true label values. As shown in Figure 3, the prediction f_i of the higher affinity y_i is greater than the prediction f_j of the smaller affinity value y_j[12]. When the main objective is the prediction of the labels relative order rather than their values, the CI measure is a suitable performance metric. For example, in DTI investigations, it is desirable to have an estimation of the interaction

likelihood of a given drug (or target) with all the targets and to rank accordingly the targets (or drugs) [12]. The CI is the proportion of the concordant pairs over the total number of pairs. Its formula is given in Figure 3

$$\frac{1}{Z}\sum_{r_i>r_j} h(fi - fj)$$

$$h(x) = \begin{cases} 1 & \text{if } x>0 \\ 0.5 & \text{if } x=0 \\ x=0 & \text{if } x<0 \end{cases}$$

Figure 3: Concordance Index.

Z is a normalization constant representing the number of pairs with different labels, and h(u) is the step function. The values of the CI range between 0.5 and 1.0 [12] Ties in the predictions are counted as half concordant pairs. Random predictions have an average concordance index of 0.5. Hence, a model without any learning skills has a mean concordance index of 0.5. The maximal value of CI is 1.0, it corresponds to the perfect order of predictions regardless of the quality of the prediction accuracy, as the concordance index is interested exclusively in the order of the predictions, it does not give any insight on the quality of the predictions themselves. Thus, for instance, it will be equal to the maximal value 1 if the order is correct even if the predictions are inaccurate. CI is typically used with continuous values; this is not the case of our predictions since they are categorical classes. However, our classes were computed from continuous affinities values in such a way that preserves the order of affinity strengths. This has made it possible to evaluate CI by using the values corresponding to the obtained predictions.

2) Categorical Hinge metric: maximum-margin classification

Categorical Hinge, or multiclass Hinge, was introduced by researchers Weston and Watkins [49]. It helps to generate decision boundaries between classes. It attempts to maximize the decision boundaries between groups that must be discriminated against in machine learning problems. In that way, it is comparable to how Support Vector Machines work, since both are interested in maximizing the margin between classes. During the training phase, the Hinge loss finds the minimal boundaries that allow discriminating samples of the same class from samples of different ones. In other words, it finds the classification boundaries that is the minimum of the quantity in Figure 4. Several different variations of multiclass Hinge have been proposed [49]. We used the version implemented by Keras which is Weston and Watkins version. it is described by the formula in Figure 3.

$$\ell(y) = \sum_{y \neq t} \max(0, 1 + \mathbf{w}_y \mathbf{x} - \mathbf{w}_t \mathbf{x})$$

Figure 4: Categorical Hinge metric.

Where t is the target class, y is the predicted class, w_t and w_y are the model parameters and x is the input. It is an appropriate metric for our work because it gives an idea about how to discriminate drug-target pairs interactions; which might help to understand the subtlety behind the drug-target association.

3) Top K Categorical Accuracy

The Top-K-categorical accuracy is used in multiclass classification models, to compute how often targets are in the top *K* predictions. *K* is a given parameter that represents the number of top elements to look at for computing accuracy. Top *K* Categorical Accuracy calculates the percentage of data points for which the targets are in the top *K* predictions. For each data, predictions are ranked in the descending order of probability values. If the rank of predictions corresponding to target is less than or equal to *K*, it is considered accurate. The division of the number of records accurately predicted by the total number of records is used to compute the Top*K* Categorical Accuracy. It is a suitable metric for our study since it allows us to check if the target label is one of the *K* top predictions. In our paper, we used the value K=*2*, which corresponds to the second most precise model accuracy, the first one being the categorical-accuracy. Thus, in this work, we utilized Categorical accuracy, Concordance Index, TopK-categorical accuracy, and Hinge-Categorical to measure the performance of the proposed model and compared it with the current state-of-art methods that we chose as our baseline, namely, DeepDTA [33].

G. Experimental Design

To assess MC-DTBA performance, we used cross-validation (CV), which is the conventional approach in recent researches. This allows a more comprehensive coverage of the complete datasets, which provides a more accurate evaluation of the model's performance. As it is the usual tradition in previous related works, the number of folds was set to 5. CV parameters can have an impact on accuracy and make the evaluation results imprecise. In [12], the authors introduced three different CV scenarios that make the performance evaluation more accurate and realistic. One can split the input drug-target pairs data by taking into consideration various constraints. Each resulting split defines a given scenario. We followed the same idea, by considering three main ways to split the input data:

1) **Scenario 1 (Sc1)**

Random drug-target pair. This corresponds to the standard k-fold CV that splits the data into k-folds randomly and keeps one of these folds for testing.

2) **Scenario 2 (Sc2)**

This is the case where drugs in the test fold are absent from the training fold. This corresponds to considering a new drug since the drug is missing from all the training data. This case is also called: cold-drug-splitting in some previous studies.

3) **Scenario 3 (Sc3)**

This is related to the situation where all the targets in the test fold are absent from the training fold. This corresponds to considering a new target, which means the target is missing from all the training data. It is also called cold-target-splitting.

For every dataset, we performed the three types of CV splitting and for every CV splitting scheme, we evaluated the prediction performance. Deep learning models' performance depends highly on various hyperparameters like learning rate, activation functions, optimizer, and other parameters. As explained before, for the learning rate, we opted for dynamic tuning by using a Keras Callback that starts with an initial value of the learning rate, and reduces it dynamically during the training when the monitored loss value is stagnant, meaning that the model is not improving. Additionally, we considered two hyper-parameters the number and the size of the filters for each layer in the convolutional LSTM block. The hyperparameters combination providing the best validation accuracy was selected as the best combination. Overfitting is a common problematic behavior of deep learning models that have been targeted by many researchers. This has led to the apparition of various techniques, among which dropout and batch-normalization are the most used methods. Dropout consists of hiding randomly a given proportion of nodes during the training process. This allows a better generalization of the model by preventing it to be too dependent on training data. In our model, we used dropout with various percentages in almost all the layers. We included a Batch-normalization in many positions of our model since it is usually recommended before and after the convolutional process. We also penalized the loss function with regularization L2-norm [47] Finally, we updated the weights using the Adadelta optimizer with a penalized loss to give a generalized prediction for the model. Embedding vector values are randomly initialized by the glorot initializer, which imposes normal distribution of weights and variance of output following variance of input. The model was constructed using tensorflow. keras [50].

H. Results

This section presents our experimental results. We have assessed our model for the three datasets, in three possible scenarios: Sc1, Sc2, and Sc3. In Sc1, both drugs and targets have been seen in the training. For Sc2, the only drug has been seen in training. Whereas, for Sc3, the only target has been seen in the training. The tables from II to IV summarize our results and those of DeepDTA. We note that MC-DTBA results are very convincing and it greatly outperforms DeepDTA across all datasets and scenarios and for all evaluation metrics. The high value of Topkcategorical accuracy with the selected k=2 for all the experiments, shows clearly the high performance of the model since this value indicates the accuracy of the model for the top two predicted

classes. Hinge values (the lesser the better) reflect also the supremacy of MC-DTBA on DeepDTA. Additionally, from Tables II to IV we notice that for both methods, globally the results for the Kiba dataset are better than other datasets and scenario Sc2 is also generally better than scenario Sc3. The performance of the first scenario, corresponding to the learning of all drugs and targets, is typically better than that of scenarios 2 and 3, which describe respectively, testing on new drugs and new targets. This could be easily justified and, was in some way predictable.

Table Ii: Results For Davis Dataset.

Dataset	Scenario	Metrics	MC-DTBA	DeepDTA
DAVIS	1	Accuracy	0.8594	0.7675
		TopKCategorical Accuracy	0.9375	0.8972
		CI	0.8588	0.6654
		Hinge	0.3463	0.4634
	2	Accuracy	0.8575	0.7622
		TopKCategorical Accuracy	0.9269	0.8533
		CI	0.7410	0.6473
		Hinge	0.3902	0.4774
	3	Accuracy	0.8379	0.7679
		TopKCategorical Accuracy	0.9208	0.8465
		CI	0.7354	0.6379
		Hinge	0.4103	0.4803

Table Iii: Results For Kiba Dataset.

Dataset	Scenario	Metrics	MC-DTBA	DeepDTA
KIBA	1	Accuracy	0.8694	0.8308
		TopKCategorical Accuracy	0.9928	0.9483
		CI	0.8467	0.8136
		Hinge	0.3692	0.4870
	2	Accuracy	0.8575	0.8174
		TopKCategorical Accuracy	0.9731	0.9420
		CI	0.7510	0.7246
		Hinge	0.3902	0.4979
	3	Accuracy	0.8370	0.7959
		TopKCategorical Accuracy	0.9508	0.903
		CI	0.7476	0.7263
		Hinge	0.4022	0.4547

Table Iv: Results For Metz Dataset.

Dataset	Scenario	Metrics	MC-DTBA	DeepDTA
METZ	1	Accuracy	0.8491	0.7632
		TopKCategorical Accuracy	0.9104	0.8650
		CI	0.7773	0.6404
		Hinge	0.3632	0.4694
	2	Accuracy	0.8434	0.7675
		TopKCategorical Accuracy	0.9069	0.8542
		CI	0.7011	0.6530
		Hinge	0.4230	0.4802
	3	Accuracy	0.8379	0.7468
		TopKCategorical Accuracy	0.8963	0.8297
		CI	0.7005	0.6174
		Hinge	0.4176	0.5030

4. Conclusion

To tackle the drug-target binding affinity problem more effectively, we propose a deep learning-based approach for multi-class DTBA prediction. First, we performed a discretization procedure that starts by applying the Elbow method to define the appropriate number of clusters, then the Kmeans algorithm is used to divide the datasets in clusters representing target classes. We represent both drugs and proteins with sequences. Then we use convolutional recurrent Neural Networks (CONVLSTM2D) to learn representations from the drug and target sequences. We compare the performance of the proposed model with DeepDTA a state of the art method in DTBA prediction. We perform our experiments on three commonly used datasets, Davis, Kiba, and Metz. Our results showed the relevance of our method. Furthermore, the use of stacked CONVLSTM2D to learn representations of proteins and drug sequences is appropriate. This could be an indication that compound and amino-acid sequences require a structure that can handle simultaneously, their ordered relationships along with their hidden patterns, which the CONVLSTM architecture accomplished successfully. Our results showed also that the use of sophisticated forms for the input is not sufficient to describe drug-protein interactions. Hence, using sequence information as the input makes the model simple and generally applicable. Predicting one class of activity among several classes instead of a float-value or a binary classification also makes it desirable for identifying the level of binding strength between a drug and a protein, which is the main objective for drug discovery, especially in the candidate selection stage.

References

[1] Thafar M, Raies AB, Albaradei S, Essack M and Bajic VB (2019) Comparison Study of Computational Prediction Tools for Drug-Target Binding Affinities. Front. Chem. 7:782. doi: 10.3389/fchem.2019.00782

[2] Hu, P. Chan, KCC. You, Z. Large-scale prediction of drug– target interactions from deep representations. International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada. IEEE, 2016,pp. 1236–1243

[3] Hamanaka,M. Kei Taneishi Hiroaki Iwata Jun Ye Jianguo Pei Jinlong Hou Yasushi Okuno. Cgbvs-dnn: prediction of compound– protein interactions based on deep learning. Mol. Inform., 36. doi: 10.1002/minf.201600045.

[4] Truchon JF, Bayly CI. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. J. Chem. Inf. Model 47(2), 2007, pp.488–508

[5] Carpenter, K. A., Cohen, D. S., Jarrell, J. T., & Huang, X. Deep learning and virtual drug screening. Future medicinal chemistry, 10(21), 2018, pp.2557–2567. https://doi.org/10.4155/fmc-2018-0314

[6] Feng, Q. (2019). PADME: A Deep Learning-based Framework for Drug-Target Interaction Prediction (Master thesis), Simon Fraser University, Burnaby, BC, Canada.

[7] Ballester,P.J. and Mitchell,J.B. A machine learning approach t predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics, 26, 1169–1175. 2010.

[8] F Li,H. et al. Low-quality structural and interaction data improves binding affinity prediction via random forest. Molecules, 20, 10947–10962. ; 2015.

[9] Shar,P.A. et al. Pred-binding: large-scale protein–ligand binding affinity prediction. J. Enzyme Inhib. Med. Chem., 31, 1443–1450. 2016.

[10] Gabel,J. et al. Beware of machine learning-based scoring functions on the danger of developing black boxes. J. Chem. Inf. Model., 54, 2807–2815. 2014.

[11] He,T. et al. Simboost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. J. Cheminform., 9, 24. 2017.

[12] Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J., et al Toward more realistic drug-target interaction predictions. Brief. Bioinformatics 2015.

[13] Schmidhuber J. Deep learning in neural networks: An overview. Neural networks, 61:85{117, 2015.

[14] Ciregan,D. et al. Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island. IEEE, pp. 3642–3649. and computer vision (Ciregan et al., 2012.

[15] Donahue,J. et al. Decaf: a deep convolutional activation feature for generic visual recognition. In: ICML, Beijing, China, 2014, pp. 647–655.

[16] Simonyan,K. and Zisserman,A. Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR), Hilton San Diego Resort & Spa, May 7–9, 2015.

[17] Dahl,G.E. et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans. Audio Speech Lang. Process., 20, 30–42.Speech recognition, natural language processing (Dahl et al., 2012.

[18] Graves,A. et al. Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing, Vancouver, Canada. IEEE, 2013, pp. 6645–6649.

[19] Hinton,G. et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of

four research groups. IEEE 2012, Signal Process. Mag., 29, 82–97.

[20] Leung,M.K. et al. Deep learning of the tissue-regulated splicing code. Bioinformatics, 30, 2014, i121–i129.

[21] Xiong,H.Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. Science, 347, 2015, 1254806.

[22] Ma,J. et al. Deep neural nets as a method for quantitative structure–activity relationships. J. Chem. Inf. Model., 55, 2015, pp.263–274.

[23] Jing, Y., Bian, Y., Hu, Z.,Wang, L., and Xie, X.-Q. S. (2018). Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. AAPS J. 2018, 20:58. doi: 10.1208/s12248-018-0210-0.

[24] Ekins, S., Puhl, A. C., Zorn, K.M., Lane, T. R., Russo, D. P., Klein, J. Exploiting machine learning for end-to-end drug discovery and development Nat. Mater. 18, 2019, pp.435–441. doi: 10.1038/s41563-019-0338.

[25] LeCun, Y. Bengio, and G. Hinton. Deep learning. nature, 521(7553):436-444 (2015).

[26] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Model. 28, 1988, pp.31–36. doi: 10.1021/ci00057a005.

[27] Krig, S. Feature learning and deep learning architecture survey, in Computer Vision Metrics (Cham: Springer), 2016, pp.375–514. doi: 10.1007/978-3-319-33762-3.

[28] Gomez-Bombarelli,R. et al.Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent. Sci., 4, 2018,pp.268–276.

[29] Jastrzkeski,S. et al. Learning to smile (s). arXiv preprint arXiv: 1602.06289. (2016).

[30] Gomes,J. et al. (2017) Atomic convolutional networks for predicting protein–ligand binding affinity. arXiv preprint arXiv: 1703.10603.

[31] Ragoza,M. et al. Protein–ligand scoring with convolutional neural networks. J. Chem. Inf. Model., 57, 2017, pp.942–957.

[32] Wallach,I. et al. Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv preprint arXiv: 2015, 1510.02855. [35]

[33] Öztürk, H., Özgür, A., and Ozkirimli, E. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics 34 2018, i821–i829. doi: 10.1093/bioinformatics

[34] Öztürk, H., Ozkirimli, E., and Özgür, A. (2019). WideDTA: prediction of drugtarget binding affinity. 2019, arXiv:1902.04166. Available online at: https://arxiv.org.

[35] Karimi, M., Wu, D., Wang, Z., and Shen, Y. DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. Bioinformatics 35, 2019, pp.3329–3338. doi: 10.1093/bioinformatics/btz111.

[36] Rogers, D., and Hahn, M. Extended-connectivity fingerprints. J. Chem. Inf. Model. 50, 2010, pp.742–754. doi: 10.1021/ci100050.

[37] Liu, K., Sun, X., Jia, L.,Ma, J., Xing, H.,Wu, J., et al. Chemi-Net: a molecular graph convolutional network for accurate drug property prediction. Int. J.Mol Sci.2019, 20:3389. doi: 10.3390/ijms20143389.

[38] Gromiha, M. Protein Bioinformatics: From Sequence to Function. New Delhi: Academic Press. 2011.

[39] Yamanishi,Y. et al. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. Bioinformatics, 2008, 24, i232–i240.

[40] Davis,M.I. et al. Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol., 29, 1046–1051.

[41] Metz, J. T., Johnson, E. F., Soni, N. B., Merta, P. J., Kifle, L., and Hajduk, P. J. Navigating the kinome. Nat. Chem. Biol. 7, 2011, 200–202doi: 10.1038/nchembio.530.

[42] Tang,J. et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. J. Chem. Inf. Model., 54, 2014, pp.735–743.

[43] Weiss, S.,.Rule-base Regression. In Proceedings of the 13th International Joint Conference on Artificial Intelligence 1993,, pp. 1072-1078.

[44] Indurkhya, N.Rule-based Machine Learning Methods for Functional Prediction. In Journal Of Artificial Intelligence Research (JAIR), volume 3, 1995, pp.383-403.

[45] Chollet,F. et al. (2015) Keras. https://github.com/fchollet/keras.

[46] Kang,L. et al. Convolutional neural networks for no-reference image quality assessment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 2014, pp. 1733–1740.

[47] Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, http://www.deeplearningbook.org 2016.

[48] Gonen,M. and Heller,G. Concordance probability and discriminatory power in proportional hazards regression. Biometrika, 2005, 92, 965–970.

[49] Weston, J., Chris W., Support Vector Machines for Multi-Class Pattern Recognition. European Symposium on Artificial Neural Networks 1999.

[50] d Abadi,M. et al. (2016) Tensorflow: a system for large- learning. In: OSDI scale machine, Vol. 16, pp. 265–283.