ProBinDeePred : a web-service for protein - drug binding affinity prediction based on deep learning study

Anıl Utku Ilgın
anil.ilgin@metu.edu.tr
Middle East Technical University Computer Engineering Department
Ankara, Turkey

ABSTRACT

Motivation: Information about interaction between bioactive compounds and target proteins is essential for drug discovery research. Due to the vast size of chemical space and possible interactions between proteins, computational methods are required. Various research has been done to satisfy this requirement, such as MDeePred, DeepDTA, DeeConv-DTI, but with different protein families needing other models or problems of implementing a comprehensive model based on many proteins families are exhausting both in terms of resource and time. And also, since applying deep-learning methods for DTI prediction is a complex process, it's needed to make an easy-to-use web service to let users access bioactivities of given drug-target protein pairs, whether predicted or measured.
Results: In this study, a web service to get bioactivity values of target protein – compound pairs either proven in a bioassay or predicted through a learning model. In the proposed system, MDeePred [13] study is used for predicting target protein – compound pairs with improved performance and refined hyperparameters to produce specific prediction models for each target protein family. MDeePred is a novel multi-channel protein-drug binding affinity prediction method using several features of both proteins and compounds. ECFP4 fingerprints for compound and sequential, structural, evolutionary, and physicochemical properties are used for proteins. To not repeat or predicting existing bioactivities, a comprehensive database named CROssBAR [4] is used as a source of information to provide users existing bioactivities. The trained models are used to predict a bioactivity value for the protein-drug pairs that don't exist in the database. In the end, this comprehensive web service provides end-users easy-to-use, reliable information about given protein-drug pairs bioactivities.

Contact : anil.ilgin@metu.edu.tr [or anil.ilgin@ceng.metu.edu.tr]

KEYWORDS

neural networks, bioinformatics, web service, target proteins drug interaction predictions,

CENG488 '21, July 40–31, 2021, Zoom, Online

ACM ISBN 000-0-0000-0000-00/00/00…$0.00
https://doi.org/00.0000/0000000.0000000

ACM Reference Format:
Anıl Utku Ilgın. 2021. ProBinDeePred : a web-service for protein - drug binding affinity prediction based on deep learning study. In *CENG488 '21: Guided Research Symposium, July 40–31, 2021, Zoom, Online.* ACM, New York, NY, USA, 8 pages. https://doi.org/00.0000/0000000.0000000

1 BACKGROUND

Human body functions are organized by various hormones, enzymes, and proteins in a complex way. When exposed, our bodies may have problems caused by foreign agents such as bacterias, viruses. In addition, problems may also occur without exposition to outsider threats, and there can be internal problems such as autoimmune diseases or cancer. All these problems share a fundamental property; they all make interactions on chemical levels. Therefore, chemical agents can prevent or eliminate these diseases, and these chemical agents are called drugs. Drugs are made of chemical compounds working as agents which bind to specific proteins only, called target proteins, to produce therapeutic effects [15]. Examples can be given as aspirin and other non-steroidal anti-inflammatory drugs; they bind to specific enzymes in the human body and inhibit the production of prostaglandins, which trigger inflammation in the body. [9] The potency of these bindings depends on many variables, such as chemical structures of drugs, and they react to the smallest changes in them, which may result in very faint or nullified bindings. This potency is also called binding affinity and measured with many mathematical models. [14]

2 INTRODUCTION

Identifying this target protein–chemical compound pair with high binding affinity is a keystone for drug discovery. The traditional experimental approach of identification is bio-assay experiments. These bio-assays are quantitative assays required to measure the affinity between compounds and protein pairs. [7] With nearly 19000 approved drugs and 5000 target proteins in the human body, making these experiments time-consuming and expensive. To overcome these limitations, it is necessary to implement computational methods to predict protein–active compound pairs with high binding affinity by analyzing the datasets of target proteins and chemical compounds. Therefore, physical experiments can be conducted in laboratories by using predicted pairs with high binding affinity, reducing the costs of the experiments.
State-of-the-art computational methods for drug-target interaction prediction, or shortly DTI prediction, are docking simulations and machine learning methods. In docking simulations,

3D structures of chemicals and proteins are used in computations. However, since many of the proteins are not crystallized, their 3D structures are unknown; therefore, docking simulations

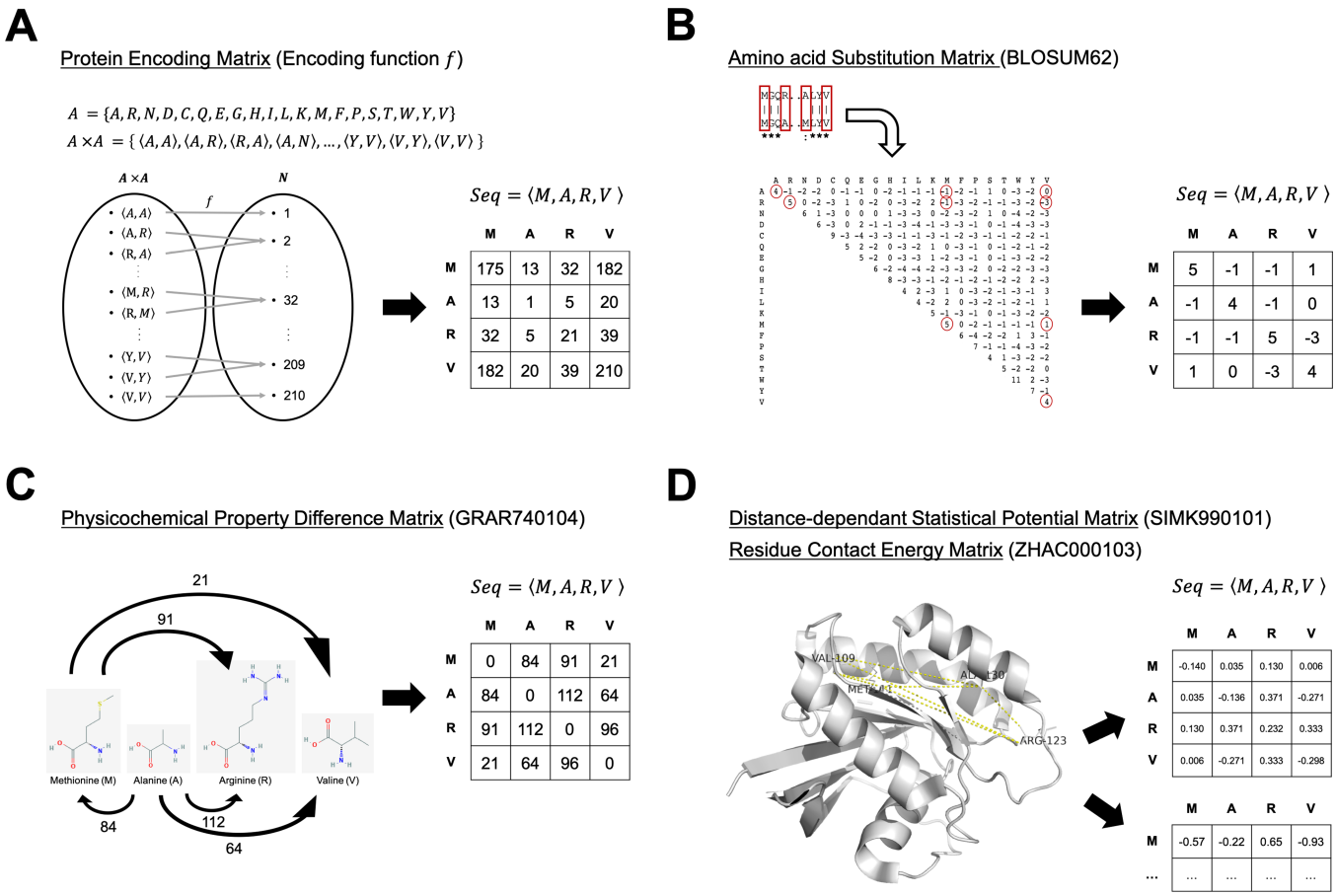

Figure 1: Protein feature matrices with examples of Seq = <M,A,R,V>. [13]

are mostly inapplicable. So, machine learning methods are used more widely in this area. Using deep learning for DTI is becoming more popular among the machine learning methods with its performance in many areas. [2] There are many different studies for this purpose, such as DeepDTA, DeepCONV-DTI, and MoleculeNet [13]. Most deep learning DTI prediction methods are made of two steps: extracting feature vectors of datum and applying deep learning. Generally, for featurization, the biological, structural, and physicochemical properties of drugs, target proteins are used. [2] For deep learning, many network models such as convolutional neural networks (Mdeepred), graph neural networks [5] ,multiple layer perceptrons [2] are used. However, the complexity of these deep learning methods is making it difficult to apply them without professional assistance. In addition, the unavailability of many research source codes or datasets also hinders the DTI researches requiring computational methods.

In this study, a web service to deliver drug-target interactions with ease is proposed to overcome this problem. In addition to measured DTIs taken from databases [4], pre-trained deep learning models for individual target protein families are generated to provide predicted interactions also. The pre-trained models are developed with the MDeePred study, a hybrid deep-neural network study to predict affinity values of protein-compound pairs. The method uses the proteo-chemometric approach, featuring proteins based on their structural (probabilistic residue contact energies, statistical potentials), evolutionary (substitution probabilities), physicochemical (compositional, polarity, molecular volume) and sequence (protein-encoding sequences) properties; and compounds based on their chemical structures (string representations of their structures). After this, protein and compound features are fed into separate neural networks: proteins into a convolutional neural network and compounds into a feed-forward neural network. The result is a prediction for the binding affinity value of a given protein-compound target in a single neuron. [13]

3 MATERIALS AND METHODS

3.1 MDeePred Feature Matrices

MDeePred study is used to prepare a deep learning model for the prediction of protein and drug interactions. In the MDeePred study,

protein features are represented as several NxN-sized 2-D matrices where N is the protein sequence length, and each matrix is a different numerical representation of protein properties. Those properties are protein encodings, substitution properties of amino acids, property differences of amino acids, their distance-dependent

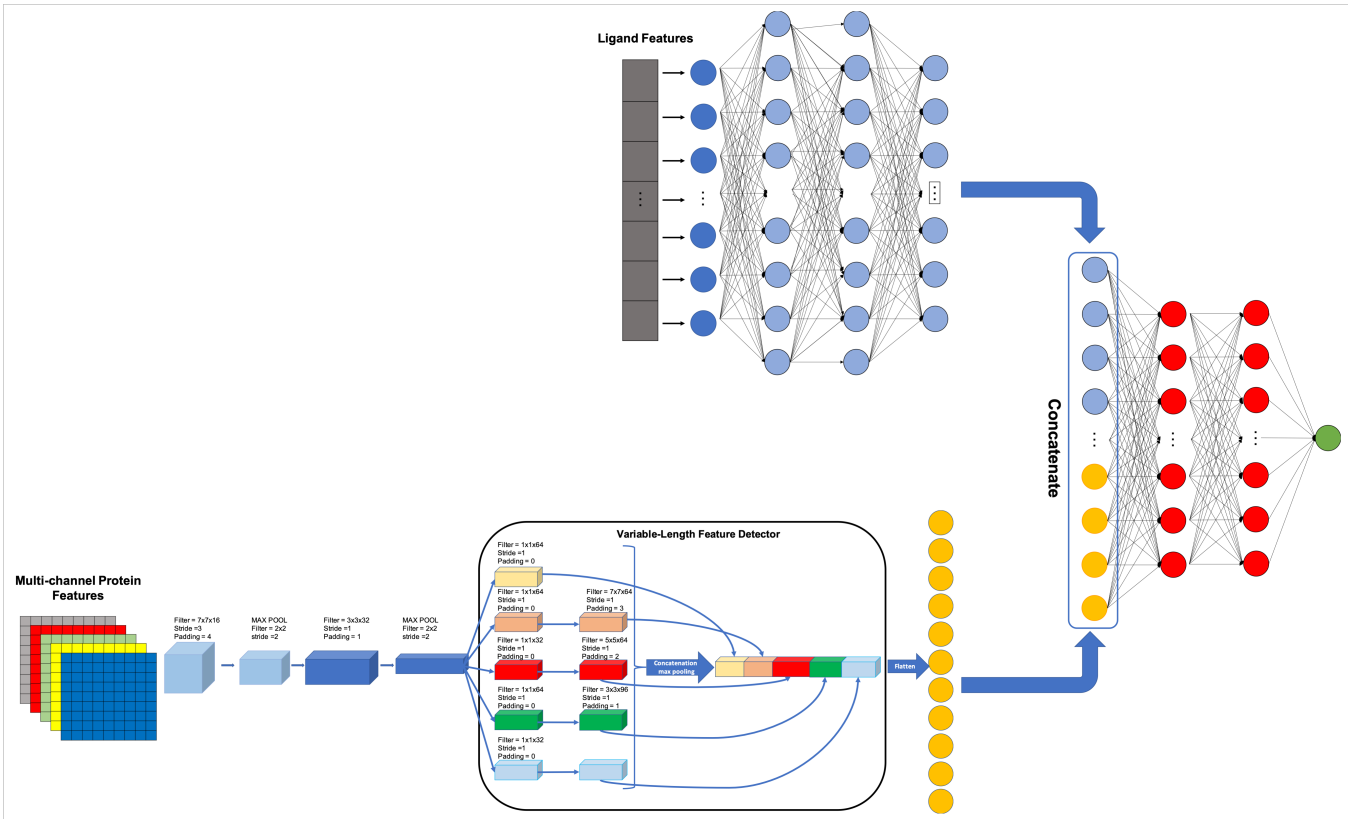

Figure 2: MDeePred Hybrid Deep Neural Network Architecture [13]

statistical potentials, and residue contact energies. We can explain the mathematical description of the given method as this:
Let P be the sequence of the given protein in FASTA format, which is a format for representing sequences where amino acids are shown as single-letter codes. $S_1, S_2 ... S_N$ are the amino acids of the given protein sequence in order.

(1) $P = \{S_1, S_2, ... S_N\}$
(2) $PxP = \{(S1, S1), (S1, S2) \dots (S1, SN) \dots (SN, SN-1), (SN, SN)\}$

PxP is the matrix whose cells represent all possible amino acids' matchings in the sequence. The diagonals of the matrix are the sequence of the protein itself, while the remaining parts are the matchings between amino acids in sequence.

Then, a new matrix is filled with the information taken from corresponding base feature matrices for each protein property. These base feature matrices taken and prepared from AAindex database [6],

BLOSUM62 scoring matrix – for the evolutionary information over substitution scores
GRAR740104 Grantham matrix – for aminoacid pair's physicochemical differences over composition, polarity, and molecular volumes
SIMK990101 distance-dependent statistical potential matrix – representing 3-D structure features
ZHAC000103 residue contact energy matrix – representing another way of 3-D structural feature (it's used for enriching the information by giving two perspectives to system) .

Each protein feature is represented as four different matrices, as shown in Figure 1. Since the protein sequence length varies, matrices are either 500x500 or 1000x1000 depending on the maximum length of the sequence. For convenience, the rest of the protein sequences are truncated to achieve fixed-sized matrices, and if the sequences are shorter than the size, the remaining parts are filled with zero.
For compound features, ECFP4 fingerprints of compounds as 1024 dimensional feature vectors are used. These fingerprints are generated by RDKit using their simplified molecular-input line-entry system (SMILES) strings, whether provided by users or gathered from the CROssBAR database. [4]

Using 2-D matrices instead of conventional 1-D vectors gives the model additional insight into properties that cannot be shown in 1-D vectors, such as repeats in protein sequences. Using this approach aims to introduce a widened perspective of protein features to the neural network. [13]

3.2 Hybrid Deep Neural Networks

In the MDeePred study, a pairwise-input hybrid neural network architecture is used for modeling compound-protein interactions,

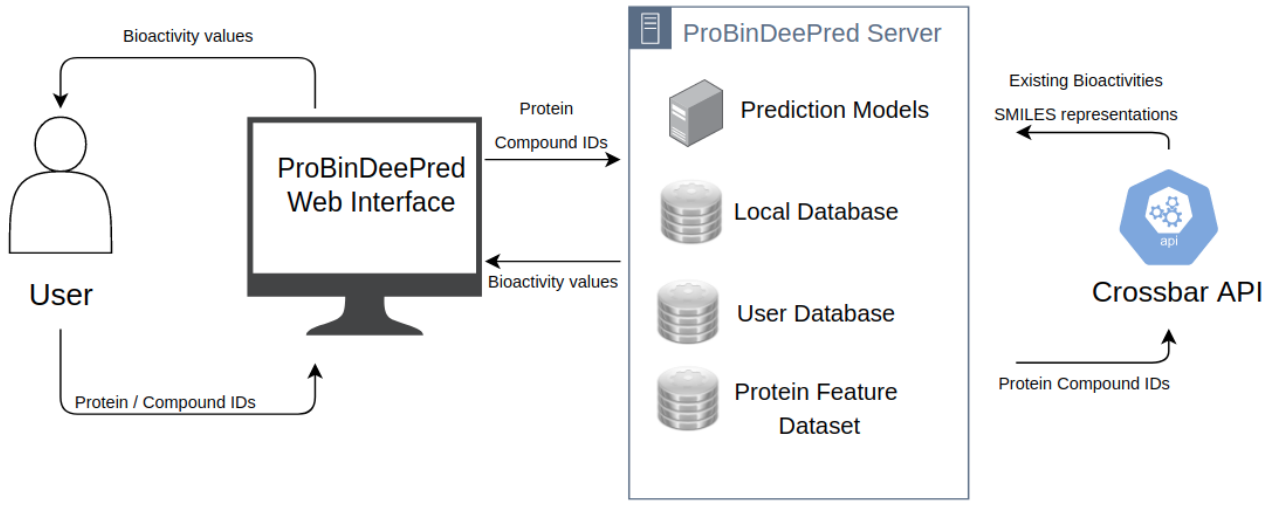

Figure 3: Overwiew of the complete system

as shown in Figure 2. For the proteins, a convolutional neural network is used with the input of protein feature matrices and the sequence matrices of proteins produced as described in Section 2.1. The convolutional part consists of two convolutional + pooling layers followed by an inception module. For the compounds, the generated ECFP4 fingerprint feature vectors of compounds as described in Section 3.1 are used for the feed-forward neural network. The output of the inception module of the protein part is flattened and concatenated with the compound part's last layer and passed in two fully connected layers. Output is a single neuron that gives the prediction of the binding affinity value for given drug-protein pair. The model aims to minimize the error (mean-squared error) during the training, which shows how much the prediction differs from the ground-truth values. [13]

3.3 Training Dataset

In this study, to train and test the MDeePred, gold standard target protein families [17], named Kinase, G Protein-coupled Receptors (GPCR), and Nuclear Receptors are used. For each protein, their FASTA formatted sequences are taken from UniProt [3] database and their registered bioactivities from ChEMBL [10].
For the Kinase family, 410 proteins and 158353 bioactivities,
For the GPCR family, 219 proteins with 163107 bioactivities and for the Nuclear Receptors, 75 proteins with 24584 bioactivities are used, as shown in Table 1.
For training of the preliminary models, 10000 ± 250 bioactivities are randomly sampled from each dataset. Later, for feature vectors, each compound's ECFP4 fingerprints are generated using their SMILES strings and stored in a separate file. Finally, each protein's feature matrices are generated according to their sequences using AAindex [6] scoring index matrices described in section 3.1.

Table 1: Datasets used in model trainings : 10000 ± 250 **bioactivities are randomly sampled from each dataset for training**

Protein Family Family	Number of Proteins	Number of Ligands	Number of Bioactivities	Database
Kinase	410	100228	158353	ChEMBL
GPCR	219	106211	163107	ChEMBL
Nuclear Receptor	75	16785	24584	ChEMBL

3.4 Web Service

In order to serve bioactivities to users, a web-service is constructed. It's architecture is shown in Figure 3. The web service consist of three main parts:
-Pre-trained models for target-protein families,
-Protein feature datasets,
-Local databases to store predictions and results.
Providing users reliable information with minimal latency is a priority. Since prediction operation can take time, users are expected to provide an email to send the results. In order to achieve this goal, the following method of preparing the end product is used:
As shown in Figure 4, users provide their email, proteins ids from a protein family selected by them, and the chemical compound ids or SMILES representations. The proteins' and chemicals' validity is checked from the CROssBAR [4] database and inform the user if any erroneous input is given. All possible pairs are calculated and checked from the CROssBAR database and local database for the given proteins and chemical compounds if there already exists a bioassay or prediction stating their binding affinity and saved in a separate CSV file. The remaining pairs' feature matrices are then

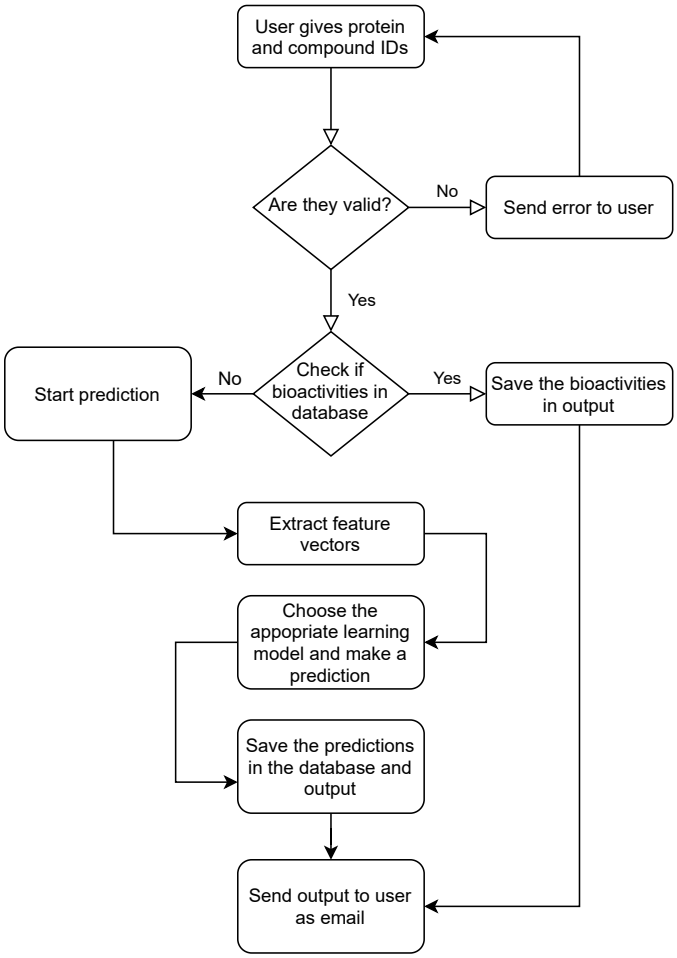

Figure 4: Flowchart of web services operation routine

calculated, and the family-specific trained model makes a prediction. The predictions are saved in a local database and the output. The finalized form of output is then sent to the user's email as an attachment.

3.4.1 *Front-end and Back-end Resources.*

To implement the back-end services of the web service, Python-Flask 2.0 [1], a lightweight web framework, is used. There is only the main page, which consists of forms to get protein ID and compound IDs and the user's email. User inputs are stored as variables during run-time, and the system also has an SQLite-based database to store previously predicted pairs. Finally, the system is connected to MDeePred implementation to make predictions for given protein–compound pairs. The front end is designed with HTML5 and CSS, using the Bootstrap framework [8] to provide a lightweight user interface.

3.4.2 *Web Service Inputs.*

Proteins and compounds are represented in various encodings, such as UniProt accessions or ChEMBL IDs, bounding users with a specific type of encoding is not favorable. Also, since models are trained for specific protein families, it is needed to know the family of proteins. Therefore, users are expected to select the protein family of given inputs. Protein IDs can be given in the form of either ChEMBL ID or UniProt accession ID, but not both. On the compound side, users can write existing chemical compounds by giving ChEMBL IDs of them or just writing SMILES representations of their chemicals. Finally, users also leave their email so the output can be sent to them.

3.4.3 *Protein and Compound Datasets .*

Even with the vast data space of ChEMBL, information on proteins and related bioassays are dispersed on the internet. Since each database, such as UniProt [3] or ChEMBL[10], has its web service, checking each service is both time and resource-consuming. It is needed to use a comprehensive service consisting of those important databases. The CROssBAR Database [4] is used for this purpose. It is a database that integrates a vast amount of information, such as bioassays, protein, and compound features, from UniProt, ChEMBL, PubChem. In Section 3.4.2, it's stated that experiments can vary in terms of encodings of proteins and compounds, such as UniProt accession and ChEMBL ID for proteins, but models are trained in a strict type of encoding. Hence, to make inputs more flexible, UniProtKB is used to translate the protein input in the desired form for trained models.

3.4.4 *Output .*

Since there is a vast amount of bioassay already done in vitro, users should also need to refer to which bioactivities are taken from bioassays and which ones are predicted. The results of the requested service is listed in a CSV, as shown in Table 2. This output is then sent to the user with an email as an attachment.

Table 2: Example output of service in table format

Bioassay ID	Protein ID	Compound ID	Binding Affinity Value
CHEMBL1062787	Q9H2G2	CHEMBL388978	10.62
CHEMBL1908783	Q9H2G2	CHEMBL384304	7.02
PREDICTION	Q9H2G2	CHEMBL3577124	7.09
PREDICTION	Q9H2G2	CHEMBL2216826	6.70

4 RESULTS

The evaluation of MDeePred has been done in the original study under multiple settings. MDeePred has been tested with various state-of-the-art solutions at the time of research, such as CGKronRLS, DeePDTA, and SimBoost. Some of these performance tests can be seen in Table 3 and Figure 5. As seen in the tables and figure, MDeePred yields better performance than the state-of-the-art methods at the time.

Various performance evaluation metrics are used for this purpose. In this study, several protein family datasets mentioned in section 3.3 are also tested. In addition to these, extensive tests with various

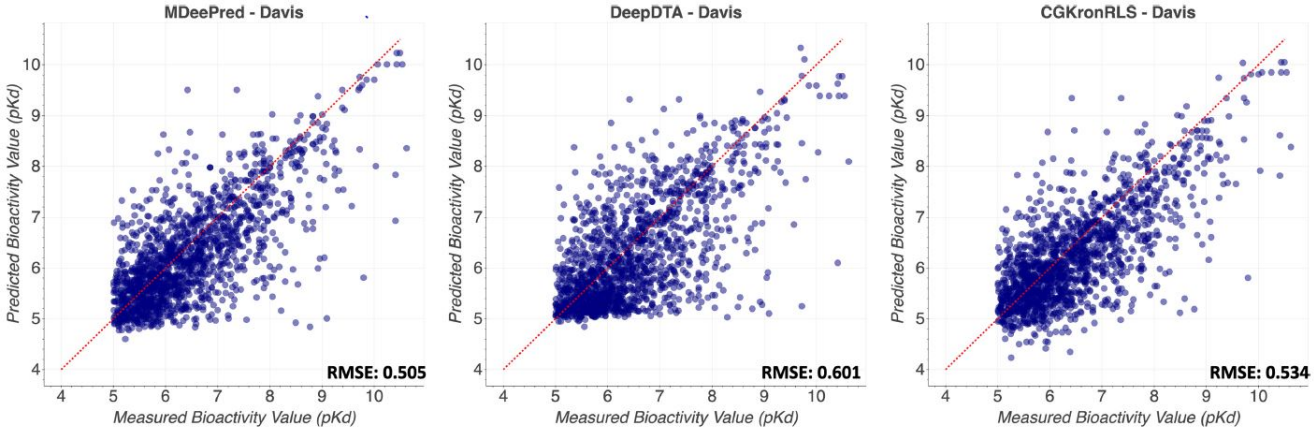

Figure 5: Comparison of MDeePred with state-of-the-art methods on Davis Dataset [13]

Table 3: Evaluation score of GPCR, Kinase and Nuclear Receptor datasets

Method	CI	RMSE	Spearman	Average AUPRC	MCC(1 uM)	MCC (100 nM)	MCC (30 nM)
MDeePred	0.886	0.505	0.69	0.744	0.691	0.644	0.626
CGKronRLS	0.873	0.534	0.671	0.724	0.677	0.632	0.604
DeepDTA	0.867	0.601	0.665	0.679	0.644	0.596	0.576

comparisons to state-of-the-art solutions and *in-vitro* tests are mentioned in the original MDeePred study. [13]

4.1 Performance evaluation metrics

Concordance index (CI) and Spearman rank correlation are used to evaluate the performance in binding affinity prediction as a continuous value. The CI measures correction of the probability of randomly selected pairs' binding affinity values order:

$$CI = \frac{1}{Z} \sum_{y_i > y_j} h(f_i - f_j)$$

Where f_i, y_i are the predicted and measured binding affinity values for i^{th} pair. Z is the total number of the pairs and $h(x)$ is the step function as shown:

$$h(x) = \begin{cases} 1.0 & x > 0 \\ 0.5 & x = 0 \\ 0.0 & x < 0 \end{cases}$$

In addition, four different bioactivity cut-off values (10 uM, 1 uM, 100 nM and 30 nM) are also used to measure performance of MDeePred as if it is a binary classification problem with Matthews Correlation Coefficent (MCC) and Area Under Precision-Recall Curves (AUPRC) evaluation metrics. Precision, Recall and MCC metrics are : [13]

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}$$

TP = True positive, TN = True negative,
FN = False negative, FP = False positive

4.2 Training, Validation and Test Settings

For training, 10000 ± 250 bioactivities are randomly sampled from each dataset split into five parts and another 2500 bioactivities are seperated for testing. Then for training, a fivefold cross-validation method is used with estimated hyper-parameters based on previous MDeePred study. In table 4 and figure 6, the performance of study's predictions on these randomly sampled datasets are shown; it shows resemblance with the original study of MDeePred and a promising performance is demonstrated on both in tables and in the figure.

4.3 Web Service

For the performance of the web service, execution times of variant-sized queries have been measured from request of query to delivery of results. Being run in local-host, execution times of web-service are shown in the table 5 . As seen in the table, there is a logarithmic relation between input size and execution time. Since establishing connections to remote databases such as ChEMBL [10] and CROssBAR [4] takes a considerable amount of time, it results in a

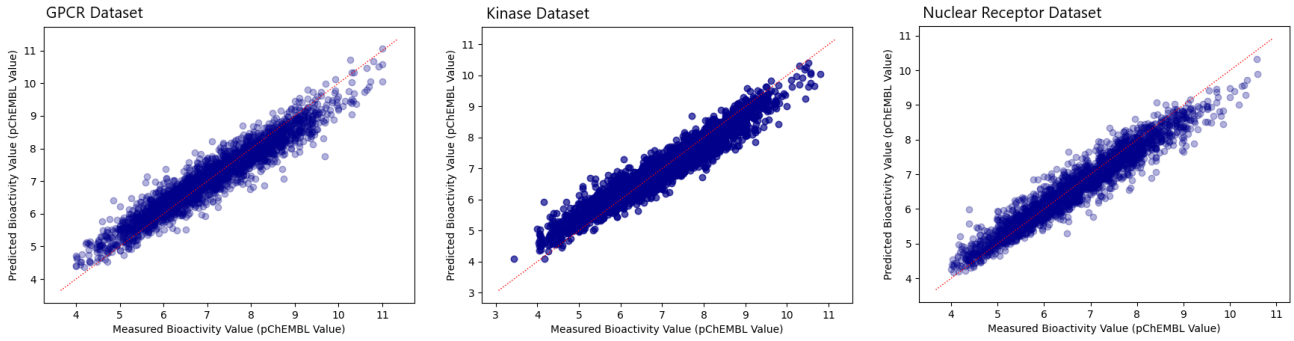

Figure 6: Scatter plots of dataset's accuracy trained with MDeePred

Table 4: Evaluation score of GPCR, Kinase and Nuclear Receptor datasets [13]

Family	CI	RMSE	Spearman	Average AUPRC	MCC(1 uM)	MCC (100 nM)	MCC (30 nM)
Nuclear Receptor	0.855	0.581	0.883	0.893	0.71	0.712	0.692
GPCR	0.834	0.701	0.839	0.912	0.65	0.672	0.675
Kinase	0.842	0.655	0.860	0.922	0.668	0.702	0.695

higher minimum time for smaller size input sizes. However, since execution time tends to speed up as input size gets larger, execution time is about reasonable limits.
any of the existing web services for DTI purposes are servicing either predicted targets for a given compound or predicted compounds for a given protein, such as Swiss Target Prediction [12] or SuperPred [11], they are kept out of comparison. Therefore, ProBinDeePred is compared with The iGPCR- Drug web service [16], only online web-service of same purpose at the time. In comparison, the iGPCR -Drug Service yields a consistent performance of 50 ± 2 seconds for each prediction, while ProBinDeePred yields better performance as the size of input increases.

Table 5: Execution times of Web-Service with variant input sizes

Input Size (pair count)	Execution Time (in seconds)
1	43
10	52
100	90
1000	516
10000	4605

5 CONCLUSION

In this study, a web service to serve users drug-target interactions is presented. Using the proposed service aims to eliminate advanced-level computer proficiency for applying deep-learning methods for DTI purposes. In detail, a deep-learning method named MDeePred is used to produce prediction models for individual target-protein families so that a prediction can be made if an unknown drug-target pair is given. For already measured interactions, a connection is established with CROssBAR, a comprehensive database consisting of many protein and drug databases. Protein and drug information are gathered from CROssBAR along with bioactivity values of existing drug-target pairs and serviced to users with predicted values of non-existing pairs. The reason for choosing the MDeePred is that it's improved performance compared to state-of-the-art methods at the time by using both protein and drug features in a multi-channel approach. Evaluations in addition to the original MDeePred study are made with gold standard protein families known as Kinase, Nuclear Receptor, and GPCR.
With this study, it's hoped to improve drug-discovery research by eliminating the complexity of developing deep-learning models for DTI purposes. Implementing individual models instead of a big-scale one aims to increase the modularity of the system and accuracy by guiding models to focus on specific features of target protein families. As improvements, models with more data points for each target protein family can be prepared to get more comprehensive and stable results and new models for other target protein families can be added.

6 ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Dr. Mehmet Volkan Atalay Dr. Ahmet Rifaioglu and Assoc. Prof. Dr. Tunca Doğan for their guidance in this project, helping me to understand the MDeePred and developing the system, and their valuable critiques about the study.

REFERENCES

[1] [n. d.]. Welcome to Flask - Flask Documentation (2.0.x). Retrieved July 03 , 2021 from https://flask.palletsprojects.com/

[2] Maryam Bagherian, Elyas Sabeti, Kai Wang, Maureen A. Sartor, Zaneta Nikolovska-Coleska, and Kayvan Najarian. 2020. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. *Briefings in Bioinformatics* 22, 1 (jan 2020), 247-269. https://doi.org/10.1093/bib/bbz157

[3] A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, R. Agivetova, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, B. Bursteinas, H. Bye-A-Jee, R. Coetzee, A. Cukura, A. Da Silva, P. Denny, T. Dogan, T. G. Ebenezer, J. Fan, L. G. Castro, and D. … Teodoro. 2020. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* 49, D1 (nov 2020). https://doi.org/10.1093/nar/gkaa1100

[4] T. Doğan, H. Atas, V. Joshi, A. Atakan, A. S. Rifaioglu, E. Nalbat, A. Nightingale, R. Saidi, V. Volynkin, H. Zellner, R. Cetin-Atalay, M. Martin, and V. Atalay. 2020. CROssBAR: Comprehensive Resource of Biomedical Relations with Deep Learning Applications and Knowledge Graph Representations. *bioRxiv* (sep 2020). https://doi.org/10.1101/2020.09.14.296889

[5] Mingjian Jiang, Zhen Li, Shugang Zhang, Shuang Wang, Xiaofeng Wang, Qing Yuan, and Zhiqiang Wei. 2020. Drug–target affinity prediction using graph neural network and contact maps. *RSC Advances* 10, 35 (2020), 20701-20712. https://doi.org/10.1039/d0ra02297g

[6] S. Kawashima. 2000. AAindex: Amino Acid index database. *Nucleic Acids Research* 28, 1 (jan 2000), 374-374. https://doi.org/10.1093/nar/28.1.374

[7] John M. Last. 2007. *A dictionary of public health.* Oxford University Press, Oxford.

[8] Jacob T. Mark Otto. [n. d.]. Bootstrap · The most popular HTML, CSS, and JS library in the world. Retrieved July 03 , 2021 from https://getbootstrap.com/

[9] UChicago Medicine. [n. d.]. How Aspirin Works. Retrieved July 01 , 2021 from https://www.uchicagomedicine.org/forefront/news/how-aspirin-works

[10] D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, and A. R. … Leach. 2018. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research* 47, D1 (nov 2018). https://doi.org/10.1093/nar/gky1075

[11] Janette Nickel, Bjoern-Oliver Gohlke, Jevgeni Erehman, Priyanka Banerjee, Wen W. Rong, Andrean Goede, Mathias Dunkel, and Robert Preissner. 2014. SuperPred: update on drug classification and target prediction. *Nucleic Acids Research* 42, W1 (may 2014). https://doi.org/10.1093/nar/gku477

[12] Swiss Institute of Bioinformatics. [n. d.]. SwissTargetPrediction. Retrieved July 09 , 2021 from http://www.swisstargetprediction.ch/

[13] A. S. Rifaioglu, R. Cetin Atalay, D. Cansen Kahraman, T. Dogan, M. Martin, and V. Atalay. 2020. MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery. *Bioinformatics* 37, 5 (oct 2020), 693-704. https://doi.org/10.1093/bioinformatics/btaa858

[14] Mohammed S. Salahudeen and Prasad S. Nishtala. 2017. An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice. *Saudi Pharmaceutical Journal* 25, 2 (feb 2017), 165-175. https://doi.org/10.1016/j.jsps.2016.07.002

[15] Richard T. Scheife. 1989. Protein Binding: What Does it Mean? *DICP* 23, 7-8 (jul 1989). https://doi.org/10.1177/106002808902300706

[16] Xuan Xiao, Jian-Liang Min, Pu Wang, and Kuo-Chen Chou. 2013. iGPCR-Drug: A Web Server for Predicting Interaction between GPCRs and Drugs in Cellular Networking. *PLoS ONE* 8, 8 (aug 2013). https://doi.org/10.1371/journal.pone.0072234

[17] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M Kanehisa. 2008. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, 13 (jun 2008), 232-240. https://doi.org/10.1093/bioinformatics/btn162