QPoweredCompound2DeNovoDrugPropMax –a novel programmatic tool incorporating deep learning and *in silico* methods for automated in silico bio-activity discovery for any compound of interest

Ben Geoffrey A S*[a], Rafal Madaj[b], Pavan Preetham Valluri[d] , Akhil Sanker[c]

[a] *Independent Researcher*
[b] *Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Poland*
[c] *SRM University, Tamil Nadu 603203, India*
[d] *PSG College of Technology, Coimbatore, India*

*Corresponding author email : bengeof@gmail.com

Abstract

Network data is composed of nodes and edges. Successful application of machine learning/deep learning algorithms on network data to make node classification and link prediction have been shown in the area of social networks through which highly customized suggestions are offered to social network users. Similarly one can attempt the use of machine learning/deep learning algorithms on biological network data to generate predictions of scientific usefulness. In the presented work, compound-drug target interaction network data set from bindingDB has been used to train deep learning neural network and a multi class classification has been implemented to classify PubChem compound queried by the user into class labels of PBD IDs. This way target interaction prediction for PubChem compounds is carried out using deep learning. The user is required to input the PubChem Compound ID (CID) of the compound the user wishes to gain information about its predicted biological activity and the tool outputs the RCSB PDB IDs of the predicted drug target interaction for the input CID. Further the tool also optimizes the compound of interest of the user toward drug likeness properties through a deep learning based structure optimization with a deep learning based drug likeness optimization protocol. The tool also incorporates a feature to perform automated In Silico modelling for the compounds and the predicted drug targets to uncover their protein-ligand interaction profiles. The program is hosted, supported and maintained at the following GitHub repository

https://github.com/bengeof/Compound2DeNovoDrugPropMax

Anticipating the use of quantum computing and quantum machine learning in drug discovery we use the Penny-lane interface to quantum hardware to turn classical Keras layers used in our machine/deep learning models into a quantum layer and introduce quantum layers into classical models to produce a quantum-classical machine/deep learning hybrid model of our tool and the code corresponding to the same is provided below

https://github.com/bengeof/QPoweredCompound2DeNovoDrugPropMax

1 Introduction

A network data is composed of nodes and edges[1]. An example of such network data would be social network data where nodes are people and their interests and edges are inter- connections between them[2-5]. Many useful applications such as customized suggestions for social media users have been developed through the use of Machine/Deep learning algorithms which accomplish this through node classification and link prediction protocols[5-10]. Similar techniques are transferable to gain insights and predictions from biological network data. Biological network data include protein-protein interaction networks, differential gene expression and regulatory networks, metabolic pathways and cell signaling networks, etc [11,12]. Using these techniques Vazquez, Alexei, et al have developed a tool for protein function prediction from protein-protein interaction networks [13]. Similarly, Hashemifar, Somaye, et al, and other groups have developed a tool for predicting protein-protein interaction using deep learning algorithms [14,15]. From gene expression network data different groups have developed tools that use deep learning algorithms to classify cancer types [16-18]. Similarly advances in understanding differential gene expression from gene expression networks have also been carried out using Deep Learning techniques by different groups [19,20]. The previous works of our research group have involved incorporating machine/deep learning techniques for automation in screening PubChem compound library and identifying the best small drug molecules for a particular drug target [21-23]. In keeping with our research focus, the present work presents a complementary approach to drug screening, wherein, given a particular PubChem compound ID for a particular compound, the developed tool predicts the most likely pharmaceutical activity of the compound and followingly performs an automated In Silico modelling to uncover the molecular details of its pharmaceutical activity. To accomplish the task mentioned above we have implemented a multi-class Deep Learning neural network to predict the target labels(PDB IDs} for a given PubChem Compound ID(CiD). To accomplish this the deep neural multi-class classifier was trained on a training dataset on protein-compound interaction network data downloaded from BindingDB with PDB IDs as class labels into which PubChem Compound IDs {CIDs) are classified [24,25]. Further, the tool also optimizes the compound of interest of the user toward drug-likeness properties through a deep reinforcement learning based structure optimization protocol. The tool also performs an automated In Silico modeling and profiling of the interaction of the compound and the predicted targets [26].

2 Material and methods

The bindingDB database [27] was downloaded and a network was constructed using NetworkX[28] wherein the nodes were compounds and proteins and edges were the interactions between them. The lower the IC50 value for a compound to inhibit a particular protein, the shorter the edges were that link them together. Each compound is identified using the PubChem Compound ID (CID) and proteins are identified with the Protein Data Bank ID (PDB ID). The dataset visualized using NetworkX and select visualization is shown in Fig.1. The Dataset consists of 536435 unique CIDs and 2707 unique PDB IDs. To generate embeddings of the network, the node2vec [29] python package was used. The model the embeddings of 65 graphs and they were used to perform a deep-learning based multi-class classification. The model classifies the CIDs into class labels which are PDB IDs [30-35]. The deep neural network architecture involves an input layer, three hidden layers that were activated by a RELU activation function, and an output layer that uses a sigmoid activation function to perform the multi-label classification. The categorical labels were vectorized using the OneHotEncoder method. Given an input node which is a PubChem compound ID (CID), the program generates a sub-network of structurally related CIDs to the input CID and performs a multi-class classification using the Deep Neural Network to classify the input CID into the PDB ID class it belongs to or to say it otherwise, predict the PDB ID of the protein the compound with a given input CID is likely to interact with. Dropouts were used as a regularization technique to overcome over-fitting and the neural network performed prediction with a validated accuracy of 83%. This protocol we call compound2drug and it is block-diagrammatically represented in Fig.2. This is followed by the Compound2DeNovoDrugPropMax protocol wherein the actor-critic based reinforcement model of DeepFMPO [36] is used to optimize the compound of interest toward possessing drug likenesses properties. The DeepFMPO [36] reinforcement learning model takes as input SMILES of compounds that are active against predicted targets and is used to generate compounds de novo which are drug-like optimized versions of the input molecules. PAINS, BRENK, NIH and ZINC filters are also added to the optimization process[37-40] The tool also performs an automated In Silico modeling and profiling of the interaction of the optimized compounds and the predicted targets and stores the results in the working folder of the user. To assess the interaction of the target and the lead compounds, the tool performs a fast and computationally cheap and efficient *In Silico* modeling and analysis of the protein-ligand interaction using AutoDock-Vina and stores the results in the working folder. Following this, a computationally more expensive In Silico modeling method of molecular dynamics using GROMACS is used to study the protein-ligand interaction and complex formation for the protein-ligand complex associated with the lowest binding energy score from the AutoDock-Vina based virtual screening protocol. Our main program through the ‘runGromacs.sh’ bash script initiates an automated MD protocol which first identifies the complex with best protein-ligand interaction from the autodock-vina virtual screening using the ‘find_best_affinity.py’ and generates

RMSD(Root Mean Square Deviation), RoG(Radius of Gyration), RMSF(Root Mean Square Fluctuation) plots which reveal the stability of the protein-ligand system based on the Molecular Dynamics simulation carried out for the protein-ligand system using the bash script ‘runGromacs.sh’. The ligand parameterization used in the automated MD protocol follows Bernardi et. al. [41]

The program is required to be run in python3 environment in the Linux OS with the following dependencies, code files, and models kept in the working folder of the user which are downloadable from the links given below.

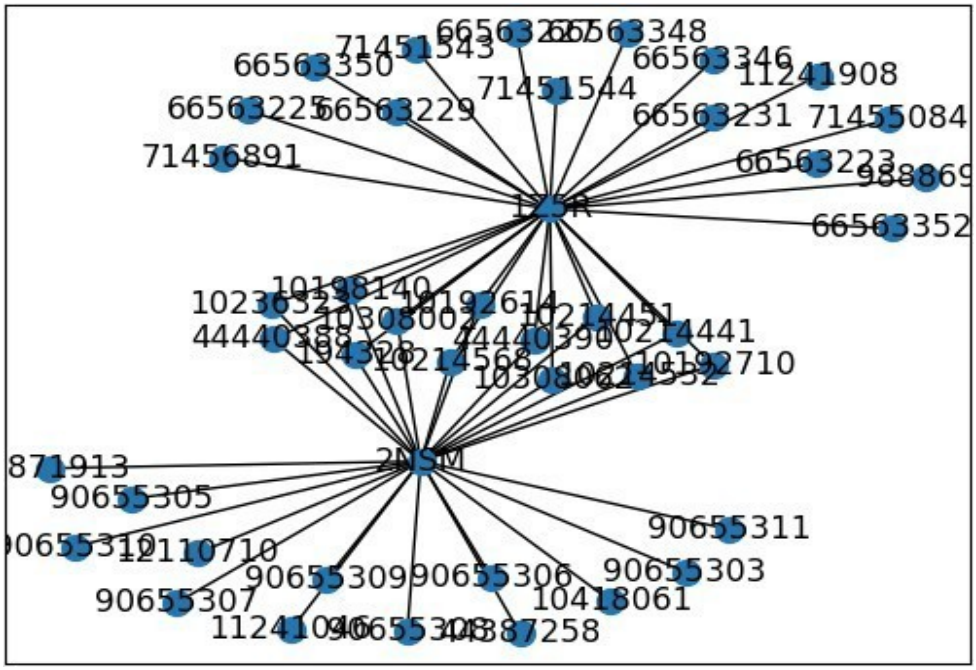

Fig.1 NetworkX visualization of compound-drug target interaction network

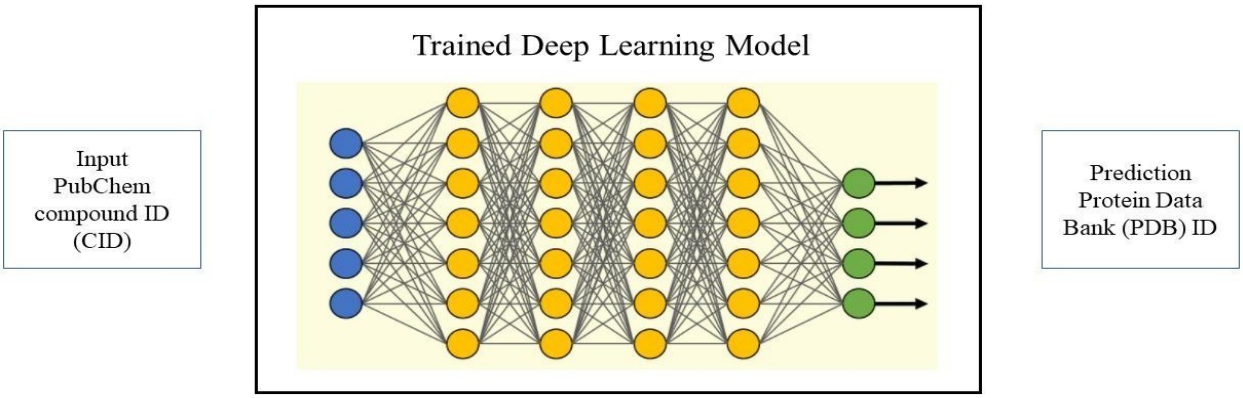

Fig. 2 Overall algorithmic workflow

Dependencies

gensim==3.8.3	tensor2tensor==1.15.7
gunicorn==20.0.4	tensorboard==2.3.0
Keras-Preprocessing==1.1.2	tensorboard-plugin-wit==1.7.0
kfac==0.2.0	tensorflow==2.3.0
matplotlib==3.3.0	tensorflow-addons==0.10.0
networkx==2.4	tensorflow-datasets==3.2.1
node2vec==0.3.2	tensorflow-estimator==2.3.0
nodevectors==0.1.22	tensorflow-gan==2.0.0
numpy==1.19.1	tensorflow-hub==0.8.0
pandas==1.1.1	tensorflow-metadata==0.22.2
scikit-learn==0.23.2	tensorflow-probability==0.7.0
scipy==1.5.2	tensorflow-text==2.3.0
seaborn==0.10.1	xgboost==1.1.1
mgltools==1.5.6	autoDock vina==4.2.6

The command line user interface of the tool is shown below (Fig.3a & b) and the usefulness of the tool is demonstrated by performing a few select examples using a randomly selected CID input. When the user runs the main program he is prompted to enter the CID of the compound for which he requires prediction of drug targets.

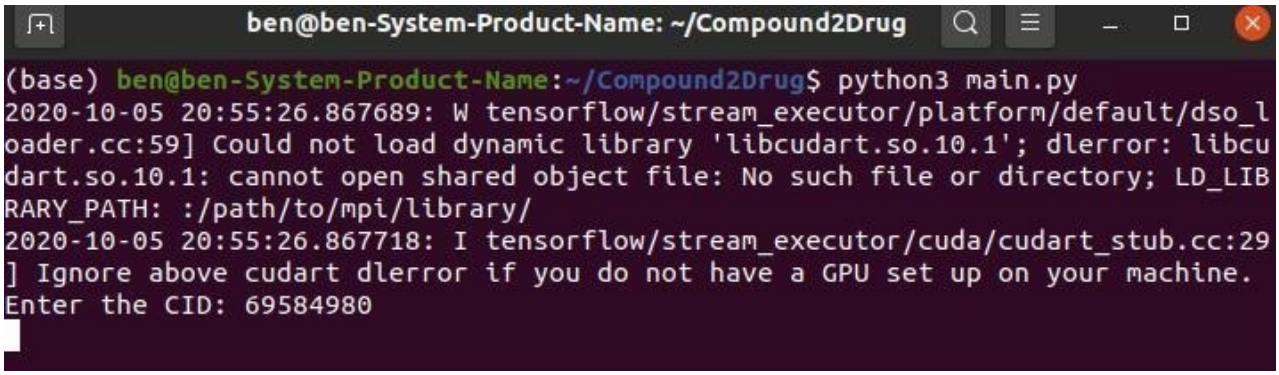

Fig.3a Tool Interface

Following this, the tool carries out the prediction task and prints out the predicted target PDB IDs as follows

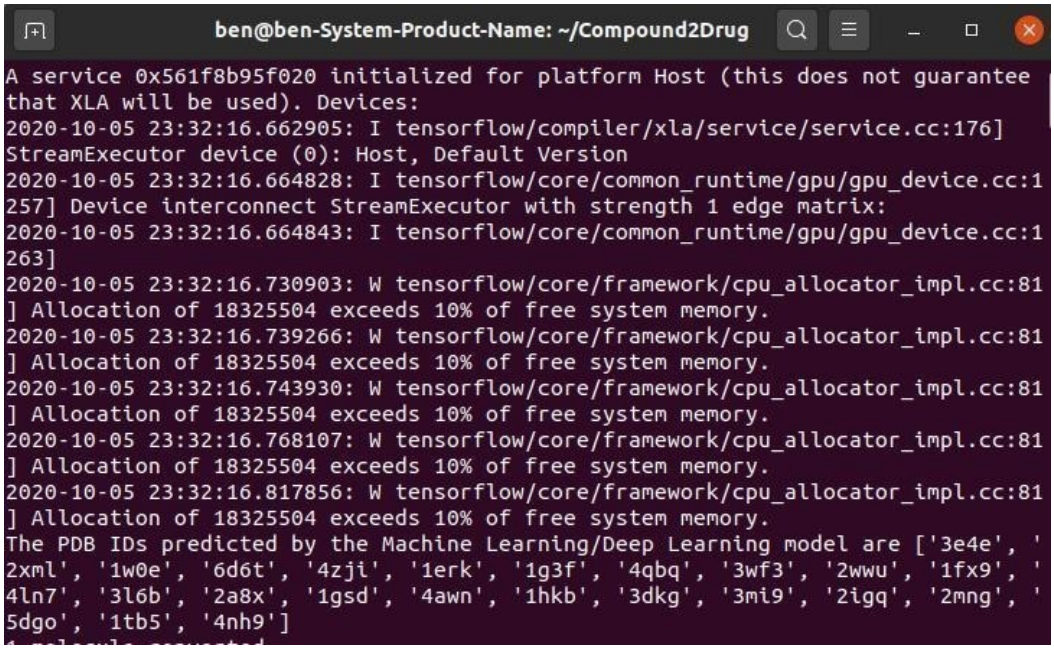

Fig.3b – Drug target prediction by the tool

For each given input CID, the program also performs automated *In Silico* modeling (as shown in Fig.4) and stores the visualized results of protein-ligand interaction in the working folder of the user. The structures of the ligand(compound) and the protein are automatically downloaded from PubChem and RCSB Protein Data Bank and they are prepared for molecular docking using the standard AutoDock scripts available through MGLTools. The program uses Web API to perform PLIP protein-ligand interaction profile and stores the results of the protein-ligand interaction profile in the working folder of the user. The 'runGromacs.sh' bash script which is run through the main program initiates an Automated molecular dynamics protocol in GROMACS for the complex with the lowest binding energy from the molecular docking protocol and generates results such as RMSD, RMSF, and RoG which reveal the stability of complex formation

```
Downloading PDB structure '1gsd'...
/home/ben/anaconda3/lib/python3.7/site-packages/Bio/PDB/StructureBuilder.py:92: PDBConstructionWarning: WARNING: Chain A is disconti
nuous at line 7286.
  PDBConstructionWarning,
/home/ben/anaconda3/lib/python3.7/site-packages/Bio/PDB/StructureBuilder.py:92: PDBConstructionWarning: WARNING: Chain B is disconti
nuous at line 7390.
  PDBConstructionWarning,
setting PYTHONHOME environment
adding gasteiger charges to peptide
Center point of docking grid for /home/ben/Compound2Drug/tmpmynkc64r/1gsd.pdbqt is as follows: x: 67.36, y: 58.9, z: 108.16
Sizes of docking grid are as follows:x: 70.59, y: 20.49, z: 38.35
###############################################################
# If you used AutoDock Vina in your work, please cite:          #
#                                                               #
# O. Trott, A. J. Olson,                                        #
# AutoDock Vina: improving the speed and accuracy of docking    #
# with a new scoring function, efficient optimization and       #
# multithreading, Journal of Computational Chemistry 31 (2010)  #
# 455-461                                                       #
#                                                               #
# DOI 10.1002/jcc.21334                                         #
#                                                               #
# Please see http://vina.scripps.edu for more information.      #
###############################################################

WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Detected 16 CPUs
WARNING: at low exhaustiveness, it may be impossible to utilize all CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: -337853075
Performing search ...
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************
done.
Refining results ... done.
```

Fig.4 – Automated *In Silico* modelling and protein-ligand interaction profiling

The tool is required to be run with the following files as shown in the working folder (Fig.5). They are downloadable from the links given below.

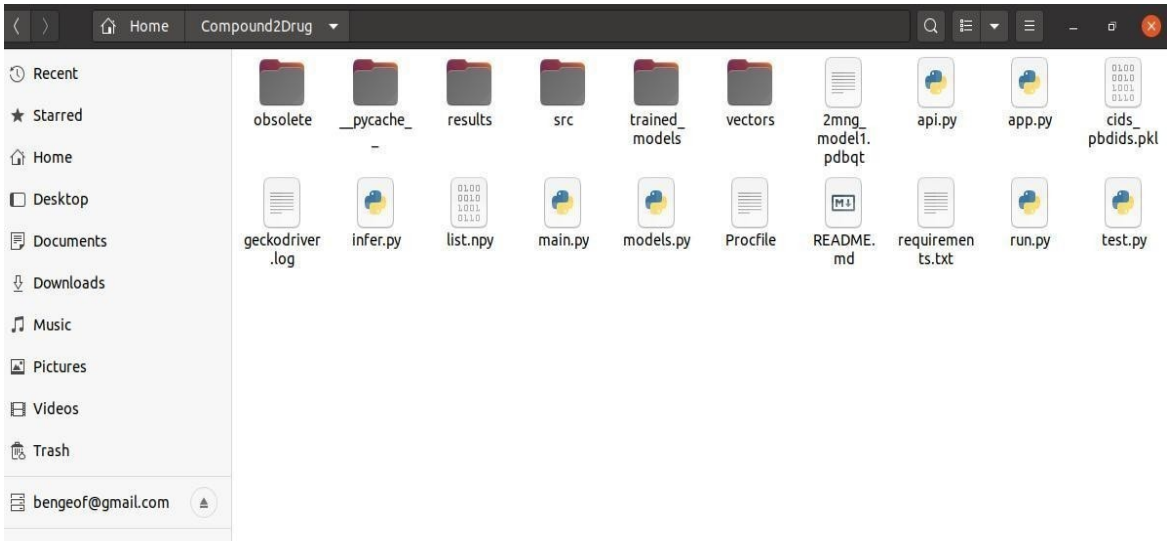

Fig.5 – Working folder

The trained models, vectors, pickle file can be downloaded from the drive link given below https://drive.google.com/drive/folders/1wwgrS6EWCnUFnPRohDFmzzShjZDb0GFe?usp=sharing https://drive.google.com/drive/folders/1JOpIdckxhCVz1A5R67YzXPxBWOlkFLJs?usp=sha ring https://drive.google.com/file/d/1ENt5pb7liNctR_8CE54g35hBU1WQ1TPx/view?usp=sharin g

The bindingDB database used can be obtained from the following link

https://drive.google.com/file/d/1s6c4k7RgBS4reF6b_K9l7-4bqK397JAS/view?usp=sharing

The code is downloadable from the GitHub repository link given below

https://github.com/bengeof/Compound2DeNovoDrugPropMax

Anticipating the exciting prospect of quantum computing and with the developers of quantum hardware such as IBM, Google, Microsoft provide programming interface to their hardware which go by names Qiskit, Cirq and Q# respectively for the development of real-world applications that leverage the of quantum hardware using quantum algorithms [42-46] we provide a version of our tool that utilizes quantum machine learning algorithms. Although fault-tolerant quantum computers may be far off, solving real world quantum machine learning problems using near-term quantum devices is possible through Pennylane which programming interface to use any of the quantum hardware providers mentioned above [47]

We use the pennylane interface to quantum hardware to turn classical Keras layers used in our machine/deep learning models into a quantum layer and introduce quantum layers into classical models to produce a quantum-classical machine/deep learning hybrid model of our tool and the code corresponding to the same is provided below

https://github.com/bengeof/QPoweredCompound2DeNovoDrugPropMax

3 Results and Discussion

To demonstrate the use of the tool with a randomly selected user input, the tool was run as described in the methodology section with a randomly chosen PubChem CID : 69584980. The tool generated a list of predicted targets and automatically estimated the strength of interaction of the compound with the predicted targets and the results are given below in Table 1. The strongest interaction was found to be with the target identified with PDB ID : 1gsd which is identified to be the enzyme Glutathione Transferase. Glutathione Transferase inhibitors increase the sensitivity of cancer cells to anti-cancer drugs and also possess several other therapeutic applications [48].

Table 1 – Results of protein-ligand interaction prediction and modeling by the tool

Compound Information	Target Information	Interaction Strength
PubChem CID	RCSB PDB ID	Binding Affinity (Kcal/mol)
69584980	3e4e	-7.6
	2xml	-9.2
	1w0e	-9.1
	6d6t	-7
	4zji	-7.1
	1erk	-7
	1g3f	-6.4
	4qbq	-7
	3wf3	-8.6
	2wwu	-7.5
	1fx9	-8.4
	4ln7	-7.4
	3l6b	-7
	2a8x	-9.1
	1gsd	**-9.9**
	4awn	-6.7
	1hkb	-7.3
	3dkg	-7.1
	3mi9	-7.9
	2igq	-7.9
	5dgo	-7.3
	1tb5	-9.3
	4nh9	-8.6

The results of AutoDock-Vina based In Silico modelling of the interaction of the drug like optimized compounds and the target can be accessed using the link and a selection of the results are tabulated in Table 2 along with their protein-ligand interactions shown in Fig.6
https://drive.google.com/drive/folders/1M6wOXG2Z1g9EDJUxCG7xB_ZDPx4S8mam?usp=sharing

Table 2 – *In Silico* Modelling Compoundt2DeNovoDrugPropMax

S.No.	Ligand (*de novo* generated SMILES)	Target	Binding Energy (Kcal/mol)
1-SMI OPT	Cc1cc2cn(C3CCC4NNC(NC(=O)C5CC(Cc6nc(I)ncc6F)CN(C)C5)C4C3)nc2cc1	1gsd	-10.9
2-SMI OPT	Cc1cc2ccccc2nc1OC(=O)c1ccc[n+](Br)c1		-10.4
3-SMI OPT	Cc1cc(C2NNC3CCC(C(=O)NN4CCC(Br)(Cn5cc6ccccc6n5)C4)CC32)ccn1		-10.8
4-SMI OPT	Cc1ccc(CNc2ccc(SCc3cnc4[nH]cc(C)c4c3)c(I)n2)nc1		-10.2
5-SMI OPT	Cc1ccc(C(=O)C2COCCN2Cc2cc(-c3nocn3)cnc2)cc1-n1nccn1		-10.7

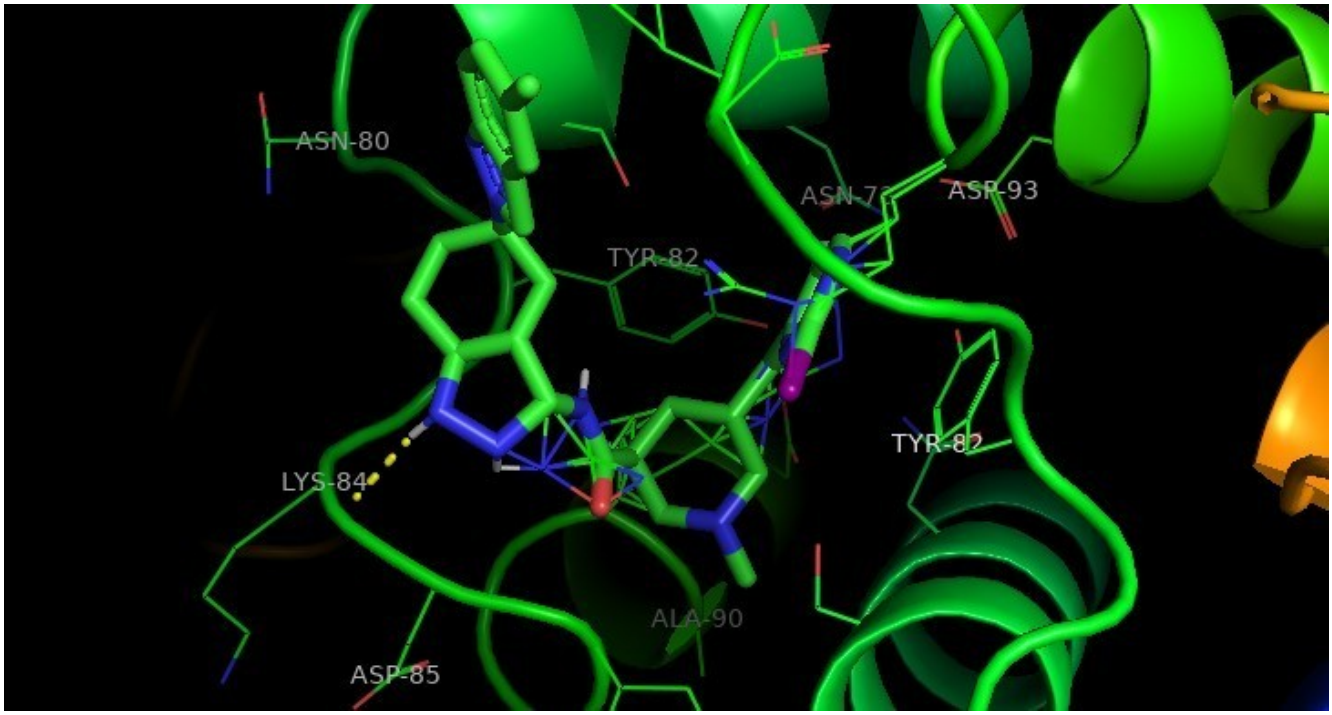

Fig.6a - Target-compound interaction of the *de novo* generated SMILES 1-OPT-SMI

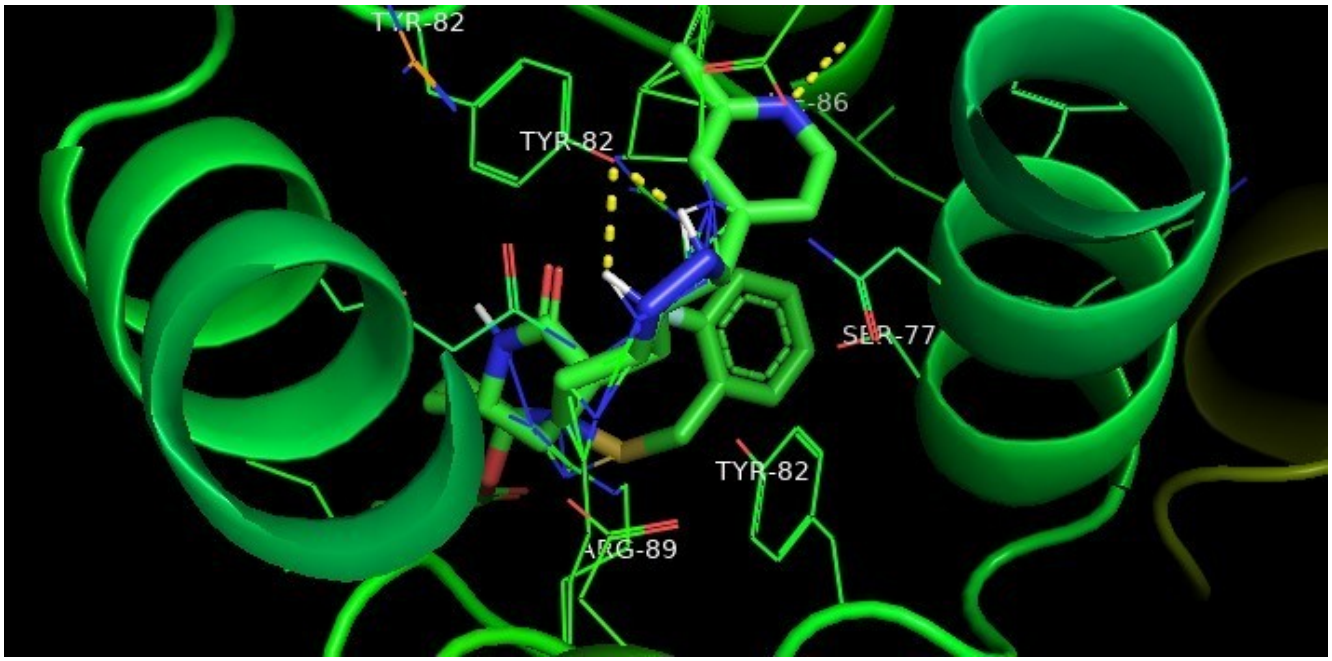

Fig.6b - Target-compound interaction of the *de novo* generated SMILES 2-OPT-SMI

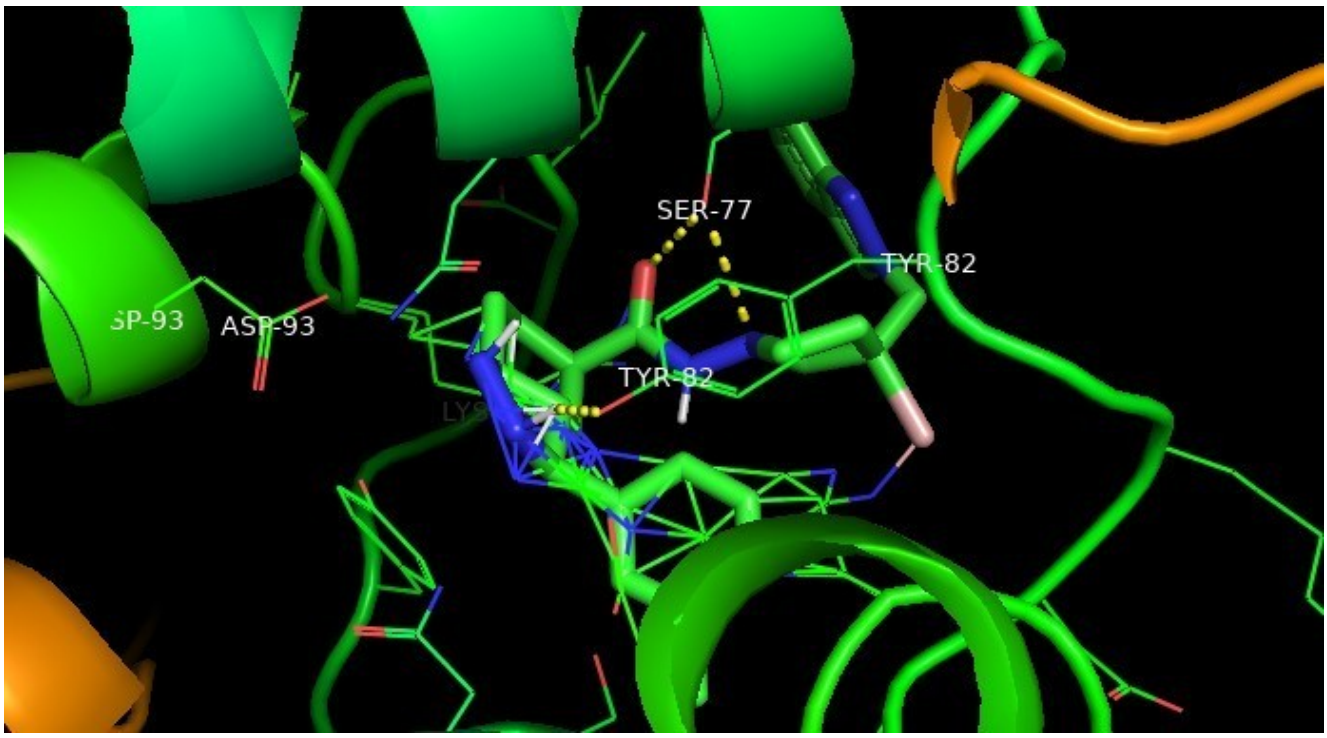

Fig.6c - Target-compound interaction of the *de novo* generated SMILES 3-OPT-SMI

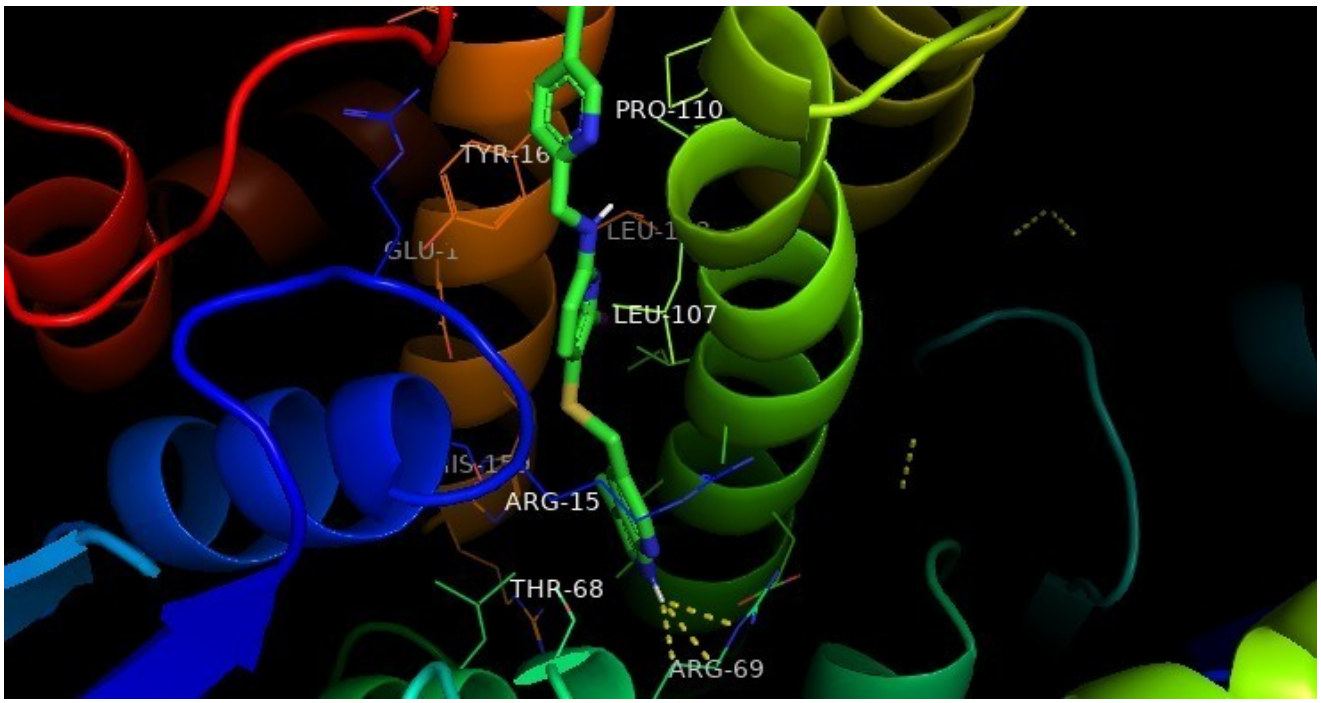

Fig.6d - Target-compound interaction of the *de novo* generated SMILES 4-OPT-SMI

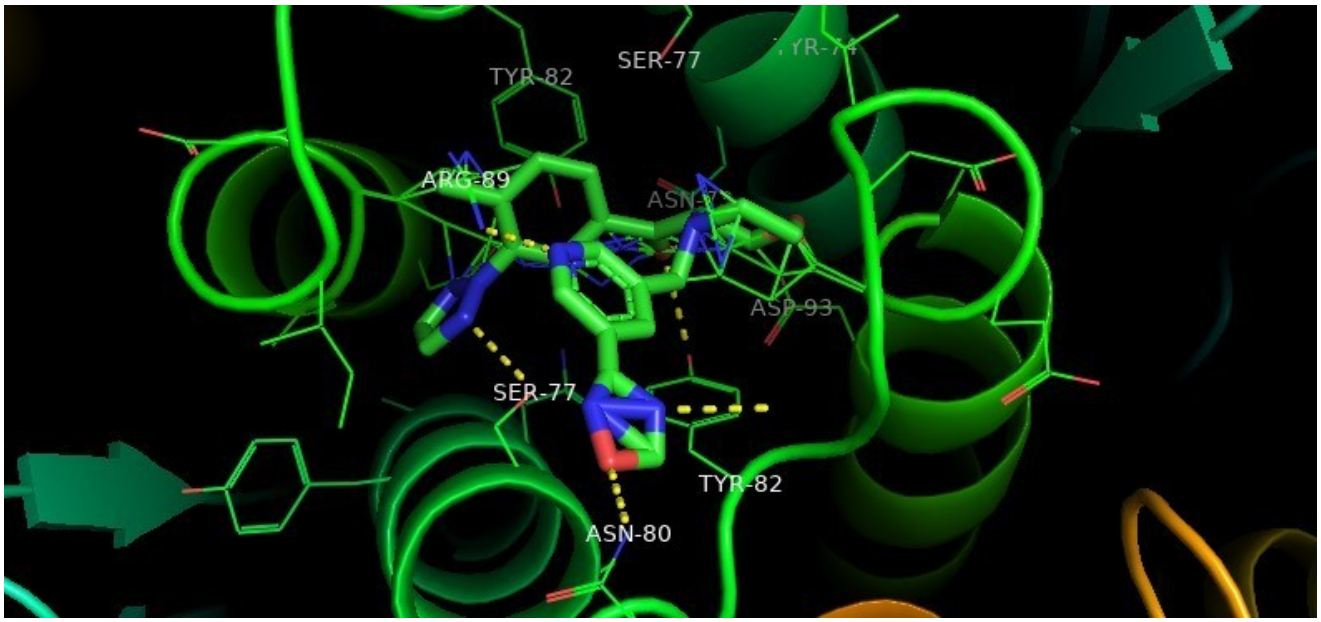

Fig.6e - Target-compound interaction of the *de novo* generated SMILES 5-OPT-SMI

The RMSD plot revealing the stability of complex formation for the complex associated with the lowest binding energy from autodock-vina virtual screen is shown below in Fig 7. The plots are automatically generated in the workflow using the 'runGromacs.sh' bash script which identifies the complex associated with the lowest binding energy and performs molecular dynamics simulation using the GROMACS package.

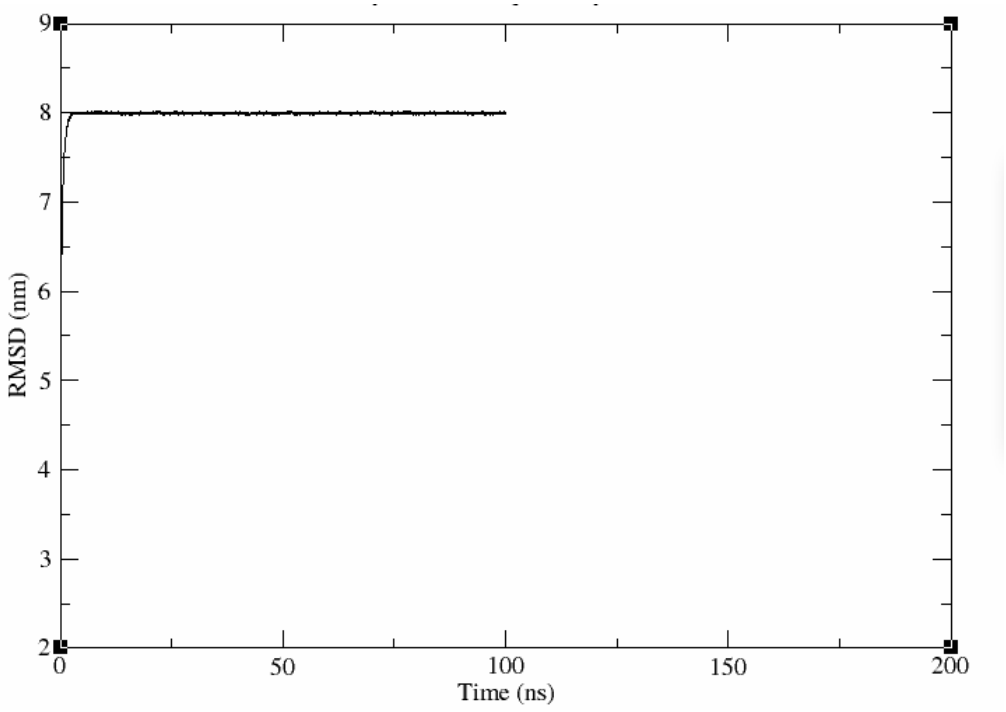

Fig. 7 – RMSD plot generated by the 'runGromcas.sh' script for complex with lowest binding energy

4 Conclusion

In the present work, the compound-drug target interaction data set from bindingDB has been used to train a deep learning multi-class classifier which was used to predict the drug targets for any PubChem compound. The user is required to input the PubChem Compound ID (CID) of the compound the user wishes to gain information about its predicted biological activity and the tool outputs the RCSB PDB IDs of the predicted drug targets for the compound. The tool also generates de novo drug-like optimized versions of the compound of interest for the user. The tool also incorporates a feature to perform automated *In Silico* modeling and profiling of the protein-ligand interaction of compounds and the predicted drug targets. To demonstrate the use of the tool a randomly selected PubChem Compound ID (CID) was given as input to the program and the use of the tool in identifying the in silico bio- activity of the compound was demonstrated.

Data and Software availability

https://github.com/bengeof/Compound2DeNovoDrugPropMax

https://github.com/bengeof/QPoweredCompound2DeNovoDrugPropMax

Acknowledgments

Acknowledge the support of the computational resources of the PLGrid Infrastructure

References

1 Trinajstic, Nenad. *Chemical graph theory*. Routledge, 2018.

2 Lee, Yofay Kari, Peter Xiu Deng, and Luke Andrew DeLorme. "Automatically generating nodes and edges in an integrated social graph." U.S. Patent No. 8,572,129. 29 Oct.2013.

3 Al Hasan, Mohammad, and Mohammed J. Zaki. "A survey of link prediction in social networks." *Social network data analytics*. Springer, Boston, MA, 2011. 243-275.

4 da Silva Soares, Paulo Ricardo, and Ricardo Bastos Cavalcante Prudêncio. "Time series based link prediction." *The 2012 international joint conference on neural networks (IJCNN)*. IEEE, 2012.

5 Zheleva, Elena, et al. "Using friendship ties and family circles for link prediction." *International Workshop on Social Network Mining and Analysis*. Springer, Berlin, Heidelberg, 2008.

6 Aggarwal, Charu C. "An introduction to social network data analytics." *Social network data analytics*. Springer, Boston, MA, 2011. 1-15.

7 Murata, Tsuyoshi, and Sakiko Moriyasu. "Link prediction based on structural properties of online social networks." *New Generation Computing* 26.3 (2008): 245- 257.

8 Aggarwal, Charu C. "An introduction to social network data analytics." *Social network data analytics*. Springer, Boston, MA, 2011. 1-15.

9 Jalili, Mahdi, et al. "Link prediction in multiplex online social networks." *Royal Society open science* 4.2 (2017): 160863.

10 Tang, Jiliang, Charu Aggarwal, and Huan Liu. "Node classification in signed social networks." *Proceedings of the 2016 SIAM international conference on data mining*. Society for Industrial and Applied Mathematics, 2016.

11 Almansoori, Wadhah, et al. "Link prediction and classification in social networks and its application in healthcare and systems biology." *Network Modeling Analysis in Health Informatics and Bioinformatics* 1.1-2 (2012): 27-36.

12 Pavlopoulos, Georgios A., et al. "Using graph theory to analyze biological networks." *BioData mining* 4.1 (2011): 10.

13 Vazquez, Alexei, et al. "Global protein function prediction from protein-protein interaction networks." *Nature biotechnology* 21.6 (2003): 697-700.

14 Hashemifar, Somaye, et al. "Predicting protein–protein interactions through sequence- based deep learning." *Bioinformatics* 34.17 (2018): i802-i810.

15 Zhang, Buzhong, et al. "Sequence-based prediction of protein-protein interaction sites by simplified long short-term memory network." *Neurocomputing* 357 (2019): 86-100.

16 Lyu, Boyu, and Anamul Haque. "Deep learning based tumor type classification using gene expression data." *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*. 2018.

17 Fakoor, Rasool, et al. "Using deep learning to enhance cancer diagnosis and classification." *Proceedings of the international conference on machine learning*. Vol.28.New York, USA: ACM, 2013.

18 Xiao, Yawen, et al. "A deep learning-based multi-model ensemble method for cancer prediction." *Computer methods and programs in biomedicine* 153 (2018): 1-9.

19 Tasaki, Shinya, et al. "Deep learning decodes the principles of differential gene expression." *Nature Machine Intelligence* (2020): 1-11.

20 Sekhon, Arshdeep, Ritambhara Singh, and Yanjun Qi. "DeepDiff: DEEP-learning for predicting DIFFerential gene expression from histone modifications." *Bioinformatics* 34.17 (2018): i891-i900.

21 Geoffrey A S, Ben; Madaj, Rafal; Sanker, Akhil; Tresanco, Mario Sergio Valdés; Davidd, Host Antony; Roy, Gitanjali (2020): A programmatic tool for automatic ease in coronavirus drug discovery through programmatically automated data mining, QSAR and In Silico modelling. ChemRxiv. Preprint.

22 AS, B. G., Sanker, A., Madaj, R., Tresanco, M. S., Upadhyay, M., & Gracia, J. (2020). A program to automate the discovery of drugs for West Nile and Dengue virus—programmatic screening of over a billion compounds on PubChem, generation of drug leads and automated In Silico modelling. BioRxiv.

23 Madaj, R., AS, B. G., Sanker, A., David, H. A., Faletif, A. I., Gracia, J., ... & Verma, S. (2020). Automated identification of small drug molecules for Hepatitis C virus through a novel programmatic tool and extensive Molecular Dynamics studies of select drug candidates. *bioRxiv*.

24 Liu, Tiqing, et al. "BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities." *Nucleic acids research* 35.suppl_1 (2007): D198-D201.

25 Gilson, Michael K., et al. "BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology." *Nucleic acids research* 44.D1 (2016): D1045-D1053.

26 Jaghoori, Mohammad Mahdi, Boris Bleijlevens, and Silvia D. Olabarriaga. "1001 ways to run AutoDock Vina for virtual screening." *Journal of computer-aided molecular design* 30.3 (2016): 237-249.

27 Gilson, Michael K., et al. "BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology." *Nucleic acids research* 44.D1 (2016): D1045-D1053.

28 Hagberg, Aric, Pieter Swart, and Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. No. LA-UR-08-05495; LA-UR-08-5495. Los Alamos National Lab. (LANL), Los Alamos, NM (United States), 2008.

29 Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 2016.

30 Farooq, Ammarah, et al. "A deep CNN based multi-class classification of Alzheimer's disease using MRI." *2017 IEEE International Conference on Imaging systems and techniques (IST)*. IEEE, 2017.

31 Maxwell, Andrew, et al. "Deep learning architectures for multi-label classification of intelligent health risk prediction." *BMC bioinformatics* 18.14 (2017): 523.

32 Wang, Sheng, Ashwin Raju, and Junzhou Huang. "Deep learning based multi-label classification for surgical tool presence detection in laparoscopic videos." *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017.

33 Murthy, Venkatesh N., et al. "Deep decision network for multi-class image classification." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

34 Szymański, Piotr, and Tomasz Kajdanowicz. "A scikit-based Python environment for performing multi-label classification." *arXiv preprint arXiv:1702.01460* (2017).

35 Xu, Youjun, Jianfeng Pei, and Luhua Lai. "Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction." *Journal of chemical information and modeling* 57.11 (2017): 2672-2685.

36 Ståhl, Niclas, Goran Falkman, Alexander Karlsson, Gunnar Mathiason, and Jonas Bostrom. "Deep reinforcement learning for multiparameter optimization in de novo drug design." Journal of chemical information and modeling 59, no. 7 (2019): 3166-3176.

37 Baell, Jonathan B., and Georgina A. Holloway. "New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays." *Journal of medicinal chemistry* 53, no. 7 (2010): 2719-2740.

38 Brenk, Ruth, Alessandro Schipani, Daniel James, Agata Krasowski, Ian Hugh Gilbert, Julie Frearson, and Paul Graham Wyatt. "Lessons learnt from assembling screening libraries for drug discovery for neglected diseases." *ChemMedChem* 3, no. 3 (2008): 435.

39 Doveston, Richard G., Paolo Tosatti, Mark Dow, Daniel J. Foley, Ho Yin Li, Amanda J. Campbell, David House, Ian Churcher, Stephen P. Marsden, and Adam Nelson. "A unified lead-oriented synthesis of over fifty molecular scaffolds." *Organic & biomolecular chemistry* 13, no. 3 (2015): 859-865.

40 Zuegg, Johannes, and Matthew A Cooper. "Drug-likeness and increased hydrophobicity of commercially available compound libraries for drug screening." *Current topics in medicinal chemistry* 12, no. 14 (2012): 1500-1513.

41 Bernardi, Austen, Roland Faller, Dirk Reith, and Karl N. Kirschner. "ACPYPE update for nonuniform 1–4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS." SoftwareX 10 (2019): 100241.

42 Aleksandrowicz, G., Alexander, T., Barkoutsos, P., Bello, L., Ben-Haim, Y., Bucher, D., Cabrera-Hernández, F.J., Carballo-Franquis, J., Chen, A., Chen, C.F. and Chow, J.M., 2019. Qiskit: An open-source framework for quantum computing. Accessed on: Mar, 16.

43 Cross, A., 2018. The IBM Q experience and QISKit open-source quantum computing software. In APS March Meeting Abstracts (Vol. 2018, pp. L58-003).

44 MacQuarrie, E.R., Simon, C., Simmons, S. and Maine, E., 2020. The emerging commercial landscape of quantum computing. Nature Reviews Physics, 2(11), pp.596-598.

45 Hassija, V., Chamola, V., Saxena, V., Chanana, V., Parashari, P., Mumtaz, S. and Guizani, M., 2020. Present landscape of quantum computing. IET Quantum Communication, 1(2), pp.42-48.

46 Cao, Y., Romero, J., Olson, J.P., Degroote, M., Johnson, P.D., Kieferová, M., Kivlichan, I.D., Menke, T., Peropadre, B., Sawaya, N.P. and Sim, S., 2019. Quantum chemistry in the age of quantum computing. Chemical reviews, 119(19), pp.10856-10915.

47 Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., Alam, M.S., Ahmed, S., Arrazola, J.M., Blank, C., Delgado, A., Jahangiri, S. and McKiernan, K., 2018. Pennylane: Automatic differentiation of hybrid quantum-classical computations. arXiv preprint arXiv:1811.04968

48 Allocati, Nerino, et al. "Glutathione transferases: substrates, inihibitors and pro-drugs in cancer and neurodegenerative diseases." *Oncogenesis* 7.1 (2018): 1-15.