# Deep generative model for anticancer drug design: Application for development of novel drug candidates against chronic myeloid leukemia

Anna D. Karpenko
United Institute of Informatics Problems,
National Academy of Sciences of Belarus
Minsk, Republic of Belarus
rfe.karpenko@gmail.com

Alexander V. Tuzikov
United Institute of Informatics Problems,
National Academy of Sciences of Belarus
Minsk, Republic of Belarus
tuzikov@newman.bas-net.by

Thimothy D. Vaitko
Belarusian State University
Minsk, Republic of Belarus
timvaitko@gmail.com

Alexander M. Andrianov
Institute of Bioorganic Chemistry
National Academy of Sciences of Belarus
Minsk, Republic of Belarus
alexande.andriano@yandex.ru

Keda Yang
Shulan International Medical College,
Zhejiang Shuren University,
Hangzhou, China
kdyang@zjsru.edu.cn

***Abstract*** **— A generative hetero-encoder model for computer-aided design of potential inhibitors of Bcr-Abl tyrosine kinase, the enzyme playing a key role in the pathogenesis of chronic myeloid leukemia, was developed. Training and testing of this model were carried out on a set of chemical compounds containing 2-arylaminopyrimidine, the major pharmacophore present in the structures of many small-molecule inhibitors of protein kinases. The neural network was then used for generating a wide range of new molecules and subsequent analysis of their binding affinity to the target protein using molecular docking tools. As a result, the developed neural network was shown to be a promising mathematical model for de novo design of small-molecule compounds potentially active against Abl kinase, which can be used to develop potent broad-spectrum anticancer drugs.**

***Keywords*** **—** ***machine learning methods, deep learning, generative neural networks, hetero-encoders, Bcr-Abl tyrosine kinase, molecular docking, anticancer drugs, chronic myeloid leukemia***

## I. INTRODUCTION

Currently, machine learning methods have been significantly developed and are used to solve many problems related to various fields of science and technology. The use of these methods in bio- and cheminformatics has made it possible to accelerate the process of designing new drugs and increase the efficiency of pharmaceutical research programs [1, 2]. Modern machine learning algorithms are applied to predict the pharmacological properties of small molecules, obtain information on the molecular mechanisms of protein-protein and protein-ligand interactions, study quantitative structure-activity and structure-property relationships, predict protein structures and protein–ligand binding affinity, as well as for virtual screening of potential drugs [1, 2]. Among the most striking achievements of artificial intelligence technologies, it should be first noted the AlphaFold 2 deep neural network [3, 4], which is based on a new approach to machine learning and uses physical and biological data on the 3D structures of proteins and their amino acid sequences. Using this program, it turned out to predict at the atomic level the spatial structures of some proteins from their primary structures and these structural data are deposited in the AlphaFold protein database, which currently includes more than 2 million protein structures (https://alphafold.ebi.ac.uk) [5]. The use of predictive neural network models for screening of chemical databases allowed one to identify a number of antibacterial and antiviral agents, including HIV-1 and SARS-CoV-2 inhibitors [6-8]. These models have also been successfully applied for screening of the FDA-approved drugs for their repurposing to treat COVID-19 [8] and drug-resistant tuberculosis [9]. In particular, a galicin molecule that is structurally different from conventional antibiotics and exhibits bactericidal activity against a wide phylogenetic spectrum of pathogens, including Mycobacterium tuberculosis and carbapenem-resistant enterobacteria, was recently identified using one of such deep learning neural networks [9]. The results of this study clearly demonstrated the effectiveness of using deep learning methods to predict potential drugs and, in particular, to expand the range of structurally different antibacterial agents [9]. The development of efficient deep learning algorithms has given impetus to the creation of a new line of research focused on the de novo design of molecules with desired pharmacological properties and synthetic availability [10-15]. To date, a large number of generative deep learning models have been proposed, which have demonstrated the promise of their use for generating new drug candidates [10–15]. As successful applications of generative neural networks, the development of a Janus kinase 3 inhibitor and active in vivo inhibitors of discoidin 1 and 2 domain receptors should be noted [15]. However, despite significant progress in the development of deep learning algorithms, their potential in the field of pharmaceutical research has not yet been fully exploited. Development of generative deep learning models with different types of architectures and types of input data for de novo design of promising drug candidates is therefore of great relevance.

In this study, a deep generative neural network based on a hetero-encoder model was developed and used in combination with molecular modeling tools for de novo design of small- molecule compounds that can inhibit the ATP-binding site of the native and mutant (T315I) Bcr-Abl tyrosine kinase, the enzyme playing a key role in the

pathogenesis of chronic myeloid leukemia (CML). The first-, second- and third-generation drugs, such as imatinib, nilotinib, ponatinib, and dasatinib, directly interacting with the ATP-binding pocket of the enzyme are currently used in clinical practice to fight against CML [16-20]. However, all these drugs exhibit high toxicity, causing a number of hematological and non-hematological side effects [21]. Additionally, most patients develop resistance to the drugs used, acquired after long-term chemotherapy [21]. In this regard, it is important to search for new inhibitors of Abl kinase, which have less toxicity and reduce the risk of possible resistance to the drugs used.

To reach the object of view, the following studies were carried out: (i) development and implementation of the hetero-encoder architecture, an improved version of the autoencoder capable of simultaneously processing input data on a molecule in several different formats, allowing one to get more stable and cost-effective generative models with improved results compared to autoencoders, (ii) assembly of a training library of small-molecule compounds potentially active towards the native and mutant Bcr-Abl tyrosine kinase which is resistant to a number of anticancer drugs used to treat patients with CML [21], (iii) training of the neural network on a set of drug-like compounds from the assembled molecular library followed by validation of the learning outcomes, (iv) generation of a wide range of potential Abl kinase ligands with a given threshold value of binding free energy (ΔG) using the developed neural network, (v) molecular docking of the generated compounds with the ATP-binding site of the enzyme, (vi) analysis of the data from molecular docking and selection of lead compounds promising for the development of novel inhibitors that can block both Bcr-Abl and Bcr-Abl$^{T315I}$ tyrosine kinase.

## II. Materials and Methods

### A. Hetero-Encoder Architecture

The developed neural network is based on the architecture of a hetero-encoder model, an autoencoder designed to solve the problems in which input data are presented in several different formats [22-24]. Such architecture makes it possible to obtain a more informative latent space due to a larger number of initial features, which expands the possibilities of finding dependencies between them in the process of training a hetero-encoder [22]. In the present study, a heteroencoder model with three encoders and two decoders which uses the Keras open library (https://keras.io) providing operation with artificial neural networks was implemented (Fig. 1) [25]. In this model, the input data are specified in the SMILES (Simplified Molecular Input Line Entry System) and canonical SMILES string formats [26-28], as well as a molecular characteristic vector (https://www.rdkit.org/docs/source/rdkit.Chem.Descriptors.htm) (Figure 1).

Given the specifics of the input data, two sub-models were developed. The architecture with two layers of LSTM (Long Short-Term Memory) was chosen as encoders for the string formats SMILES and canonical SMILES. The input data are processed by two LSTM layers consisting of 128 cells each, and the resulting embeddings for the string format are transferred to the fully connected layer (dense encoder) of the neural network (Fig. 1).

The numerical characteristics of molecules are processed by a fully connected feed-forward neural network, which is represented by an encoder consisting of two fully connected layers with dimensions of 64 and 32, a batch normalization layer and an additional fully connected layer of 16 neurons, which creates embeddings for numerical features. These embeddings go to the concatenating layer, where they form one vector which is normalized on the batch normalization layer and transferred to a fully connected layer of 128 neurons, and the desired value of the binding energy of a molecule to a therapeutic target is then set. The results of this layer operation, that is the processed embeddings and the value of ΔG, form a latent space with a dimension of 129 (Fig. 1).

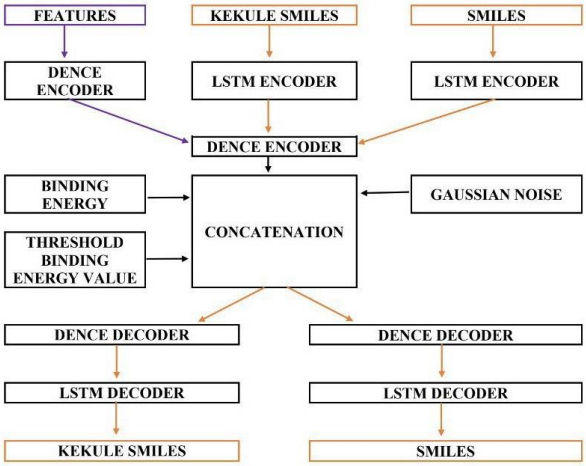


Fig. 1. Architecture of the developed hetero-encoder model.

The developed hetero-encoder model includes two identical decoders designed to obtain a description of a molecule in two string formats from the latent space vectors (Fig. 1). The decoders operate as follows: the latent space vector is fed to two independent fully connected layers of dimension 128 each and, after passing them, is normalized on the batch normalization layers. The output generates two numeric vectors that are passed as initialization vectors to the LSTM layer, and the input of this layer additionally receives a string format (for each layer, it is own). The dimension of the LSTM layer in the decoders is also 128. After passing through the LSTM layer, the data are transmitted to a fully connected layer with a softmax activation function, which processes it in such a way as to obtain the probabilities of the next symbols as an output. For all other fully connected layers, the ReLu activation function is used, and for LSTM layers, the Tahn function is applied.

The developed hetero-encoder model has the following specific features (Fig. 1):

- During the preparation of the input data, characters are added to the beginning and end of the string for training the LSTM layers to “remember dependencies in strings”; therefore, the input of the hetero-encoder is a string without the last character, and the output is expected to be a string without the first character.

- A neuron has been added to the latent layer, which allows one to use the value of ΔG as an additional parameter; this neuron is not associated with encoders and is used only in decoders to generate molecules with the desired binding affinity to a molecular target.

- Batch normalization layers are used for more efficient and stable training of the neural network in its coding and decoding parts.

- At the stages of encoding and decoding, data formats are not related to each other, making it possible to expand the network architecture if it is necessary to repurpose it for other therapeutic targets.

- All encoders and decoders are trained together and simultaneously in the general structure of the hetero-encoder.

### *B. Input Data Preparation*

To form a training molecular library, 120,000 compounds containing 2-arylaminopyrimidine were selected from the chemical database PubChem (https://pubchem.ncbi.nlm.nih.gov/) [29]. Chemical structures of these compounds were then converted to the SMILES and canonical SMILES formats. SMILES gives information on the composition and chemical structure of a molecule using an ASCII character string, while canonical SMILES is a version of the SMILES specification including canonization rules that allow the molecular formula of any substance to be written in an unambiguous way. These rules concern the choice of the first atom in the record, the direction of bypassing molecular cycles, and the choice of the direction of the molecule main chain at branching.

The resulting molecular descriptors were integrated into the training set and then transformed and filtered using the procedure described below. For each molecule, the lengths of string formats were checked, and, in cases where they were outside the range of 35–75 characters, the molecule was removed from the data set. Further, all atoms in the string record were changed to their single-character equivalents to prevent additional difficulties in the operation of the neural network. The first characters of all strings were then replaced with a new string opening character, which had not previously been found in the sample, and termination characters were added to all strings, moreover, in such a way that all strings after conversion had the same lengths. After that, the strings were converted to vector format. First, for each string format, unique characters were extracted and each of them was assigned a unique index within the data format. After that, each character of the string was replaced by a numeric vector with a dimension equal to the number of unique characters in the format, and consisting of zeros and a single one in place of the character index, i.e. each row was represented as a matrix of zeros and ones (this method is also known as One-Hot-Encoding; https://machinelearningmastery.com/why-one-hot-encode-data-in-machine-learning/). In the case of numerical embeddings, a standardization procedure was used to balance their impact on the learning process.

After filtering, a sample of 108,410 molecules was obtained in the formats chosen for training the neural network. Molecular docking program AutoDock Vina (https://vina.scripps.edu) [30] was then used to generate complexes of these molecules with the structure of Bcr-Abl tyrosine kinase in the crystal (https://www.rcsb.org; PDB ID: 3OXZ), as well as to calculate the values of ΔG. Molecular docking was carried out in the approximation of rigid receptor and flexible ligands. The grid box for docking included the ATP-binding site of the enzyme and had the following parameters: ΔX = 31 Å, ΔY = 23 Å, ΔZ = 23 Å centered at X = 18 Å, Y = 8 Å, Z = 6 Å. The value of the exhaustiveness parameter setting the number of individual sample runs was equal to 100 [30]. The prepared training library including 108,410 compounds and the corresponding values of ΔG formed the dataset for training and testing the neural network, which was divided into training and testing subsets in the proportion of 80% and 20%, respectively, of the total number of compounds.

### *C. Hetero-Encoder Training*

The hetero-encoder model included 784,537 parameters (weights), of which 781,369 parameters were used to train the neural network. In the learning process, the loss function (LF) of the following form was used:

$$LF(s) = CCE(s) + 0.1 \cdot CCL(s),$$

where CCE(s) is the categorical cross entropy [31], s is a molecule in the SMILES format, and CCL(s) (CustomChemLoss) is the function that imposes penalties for violations of a molecule stereochemistry and the absence of 2-arylaminopyrimidine in its chemical structure. The value of the weight factor for the penalty function was chosen by sorting through a discrete number of coefficients aimed at the determination of the value of this parameter providing stability of the neural network training.

The categorical cross entropy CCE(s) was calculated using the formula

$$CCE(s) = -\sum_{s_i \in s} p(s_i) \log q(s_i),$$

where $p(s_i)$ and $q(s_i)$ are the true and predicted probabilities of generating the character $s_i$ of the string s, respectively. The CCL(s) penalty function was calculated using the following criteria:

$$CCL(s) = \begin{cases} 0, \text{ if } s \text{ is correct and contains 2-arylaminopyrimidine} \\ 1, \text{ if } s \text{ is correct and does not contain 2-arylaminopyrimidine} \\ 5, \text{ if } s \text{ is not correct} \end{cases}$$

During the learning process, the loss function for the training set varied from 1.867 to 1.0375, and, for the testing set, it changed from 1.943 to 1.0445. A method for stochastic optimization Adam [32] was used as an optimizer.

The following parameters were used to train the hetero-encoder:

– Factor of conservation of the first moment $\beta_1$ was equal to 0.9;

– Factor of conservation of the second moment $\beta_2$ was set to 0.999;

– Smoothing parameter $\zeta$ was equal to $10^{-7}$;

– Object containing information about the computing node $\eta$ was equal to 0.005;

– Initial value of learning rate was set to 0.005;

– Number of complete iterations of the network training was equal to 25;

– Sub-sample size at one training step was equal to 256.

Graphs of the loss function for the training and testing datasets indicate their similarity and final convergence, which allows one to conclude that the neural network was successfully trained and there was no retraining (Fig. 2).

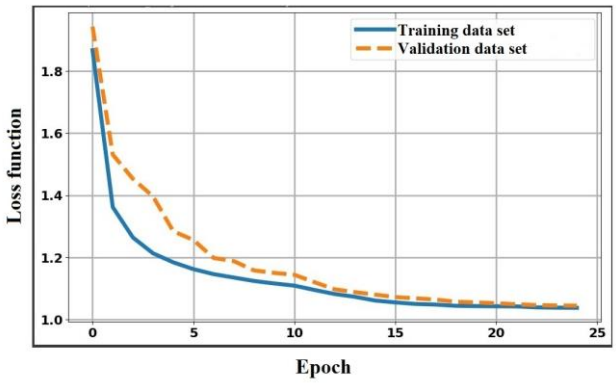


Fig. 2. Training and validation losses for the developed hetero-encoder model.

### *D. Compound Generation*

The developed hetero-encoder was used to generate a wide range of high-affinity ligands of Bcr-Abl tyrosine kinase for subsequent identification of potential inhibitors of this enzyme by molecular docking methods. To implement the generation process, a representation of the latent space was obtained using the coding part of the model from the molecules of the training library with the values of ΔG lower than –9 kcal/mol. Normally distributed noise was then introduced into the resulting vectors to generate new latent vectors, which, together with a given threshold energy, were fed to the decoding part of the model as initializing vectors. At the same time, the start symbol for symbol-by-symbol generation was the line start symbol added earlier each time. Characters were generated sequentially until a line ending character was obtained. As a result of the hetero-encoder operation, linear SMILES representations for 1,117 molecules were obtained, which were cleared of duplicates, checked for correctness, interpretability, and presence of 2-arylaminopyrimidine using the RDKit module (http://www.rdkit.org/) [33] and transformed from the SMILES format to chemical structures. After filtration procedure, 1,083 compounds were selected and their potential inhibitory activity against Bcr-Abl and Bcr-Abl$^{T315I}$ tyrosine kinase was evaluated by molecular docking tools.

## III. RESULTS AND DISCUSSION

To evaluate the efficiency of the hetero-encoder operation, complexes of the generated compounds with the X-ray structures of Bcr-Abl tyrosine kinase (PDB ID: 3OXZ; https://www.rcsb.org) and its mutant form Bcr-Abl$^{T315I}$ (PDB ID: 3OY3; https://www.rcsb.org) were built using the AutoDock Vina program (https://vina.scripps.edu). Molecular docking was performed via the computational protocol identical to that used for the formation of the training dataset. Under the calculation data, the generated compounds showed the values of binding energies to the native and mutant Bcr-Abl tyrosine kinase ranging from –6.5 to –13.8 kcal/mol. At the same time, 569 molecules with the values of energy from –9.0 to –13.8 kcal/mol were selected for the further analysis. For these molecules, a more accurate assessment of the protein-ligand binding affinity was performed using three scoring functions, namely AutoDock Vina [30], NNScore 2.0 [34], and RF-Score-4 [35]. For this purpose, the ranks of all compounds were determined according to each scoring function and the value of the exponential consensus rank (ECR) was calculated for each compound by the formula [36]

$$ECR = \sum_{sf} \frac{1}{\sigma_{sf}} \cdot \exp(-\frac{rank_{sf}}{\sigma_{sf}}),$$

where $rank_{sf}$ is the rank of the compound according to the scoring function *sf*, $\sigma_{sf}$ is the parameter that controls the influence of the scoring function *sf* on the results of consensus selection (ECR was calculated using $\sigma_{sf} = 10$ for all considered *sf*, since the contributions of the individual scoring functions were taken equal).

To identify compounds potentially active against both therapeutic targets, cross exponential consensus rank (crossECR) was calculated for all selected molecules using the formula

$$crossECR(i) = \frac{ECR_1(i)}{max_i\{ECR_1(i)\}} + \frac{ECR_2(i)}{max_i\{ECR_2(i)\}},$$

where $ECR_1(i)$ is the ECR value of ligand i for the first target (native Abl kinase), and $ECR_2(i)$ is the ECR value of ligand i for the second target (mutant Abl kinase).

Molecules with the low crossECR values were assigned to the group of promising drug candidates, dual-targeted anticancer compounds able to inhibit the catalytic activity of the native and mutant Bcr-Abl tyrosine kinase. Analysis of the data from molecular docking revealed four lead compounds that showed a high-affinity binding to the both considered therapeutic targets. These compounds are characterized by the low values of ΔG predicted for the ligand/Bcr-Abl complexes using classical and machine-learning scoring functions, which are comparable with those calculated by the same computational protocol for the potent FDA-approved anticancer drug Ponatinib (Table I). These findings testify to that the developed neural network is a promising computational model for de novo design of small-molecule compounds potentially active against Bcr-Abl and Bcr-Abl$^{T315I}$ tyrosine kinase, which can be used to develop novel, potent and broad-spectrum anticancer agents.

## IV. CONCLUSION

The hetero-encoder model was developed to generate novel potential inhibitors of Bcr-Abl tyrosine kinase, the enzyme playing a key role in the pathogenesis of CML. The neural network was trained and tested and the results of its operation were analyzed. In the process of validation of the neural network, 1,083 molecules were generated and their binding affinity to the catalytic site of the native and mutant Bcr-Abl tyrosine kinase was evaluated using molecular docking tools.

As a result, four lead compounds were identified presenting considerable interest for further theoretical and

TABLE I. CrossECR VALUES AND BINDING ENERGIES FOR FOUR NEURAL NETWORK-GENERATED COMPOUNDS I–IV AND PONATINIB (V) IN THE COMPLEXES WITH THE NATIVE AND MUTANT BCR-ABL TYROSINE KINASE [I]

| Ligand | CrossECR value | Binding energy (kcal/mol) Native Bcr-Abl tyrosine kinase | | | Binding energy (kcal/mol) Mutant Bcr-Abl tyrosine kinase | | |
|---|---|---|---|---|---|---|---|
| | | AutoDock Vina | RF-Score-4 | NNScore 2.0 | AutoDock Vina | RF-Score-4 | NNScore 2.0 |
| I | 0.0674 | -13.8 | -11.5 | -9.8 | -11.3 | -11.3 | -8.9 |
| II | 0.0674 | -13.0 | -11.6 | -10.1 | -11.0 | -11.5 | -9.0 |
| III | 0.0835 | -10.4 | -11.6 | -11.7 | -10.0 | -11.3 | -11.3 |
| IV | 0.0931 | -13.4 | -11.3 | -9.3 | -11.4 | -11.4 | -5.8 |
| V | 0.0399 | -12.0 | -11.4 | -12.2 | -12.2 | -11.2 | -12.4 |

experimental studies, including computer-based generation of their modified forms with improved pharmacological properties, synthesis and detailed biomedical assays.

## ACKNOWLEDGMENTS

This study was supported by the State Program of Scientific Research "Convergence 2025" (subprogram "Interdisciplinary research and emerging technologies", project 3.4.1).

## REFERENCES

[1] J. Vamathevan et al., "Applications of machine learning in drug discovery and development," Nature Reviewers. Drug Discovery, vol. 18, no. 6, pp. 463–477, 2019.

[2] C.F. Lipinski, V.G. Maltarollo, R.P. Oliveira, A.B.F. da Silva, and K.M. Honorio, "Advances and perspectives in applying deep learning for drug design and discovery," Frontiers in Robotics and AI, vol. 6: 108, 2019.

[3] P. Cramer, "AlphaFold2 and the future of structural biology," Nature Structural & Molecular Biology, vol. 28, no. 9. pp. 704-705, 2021.

[4] P. Bryant, G. Pozzati, and A. Elofsson, "Improved prediction of protein-protein interactions using AlphaFold2, "*Nature Communications*," vol. 13, no. 1: 1265, 2022.

[5] A. David, S. Islam, E. Tankhilevich, and M.J. Sternberg, "The AlphaFold database of protein structures: a biologist's guide," Journal of Molecular Biology, vol. 434. no. 2: 167336, 2022.

[6] P.B. Timmons and C.M. Hewage, "ENNAVIA is a novel method which employs neural networks for antiviral and anti-coronavirus activity prediction for therapeutic peptides," Briefings in Bioinformatics, bbab258. 2021. doi: 10.1093/bib/bbab258.

[7] A.M. Andrianov, G.I. Nikolaev, N.A. Shuldov, P.B. Bosko, A.I. Anischenko, and A.V. Tuzikov, "Application of deep learning and molecular modeling to identify small drug-like compounds as potential HIV-1 entry inhibitors," Journal of Biomolecular Structure and Dynamics, vol. 40, no. 16, pp. 7555-7573, 2022.

[8] Y. Zhang, T. Ye, H. Xi, M. Juhas, and J. Li, "Deep learning driven drug discovery: Tackling Severe Acute Respiratory Syndrome Coronavirus 2," Frontiers in Microbiology, 2021. doi: 10.3389/fmicb.2021.739684

[9] J.M. Stokes et al., « A deep learning approach to antibiotic discovery," Cell, vol. 180, no. 4, pp. 688–702, e13, 2020.

[10] R. Mercado et al., "Practical notes on building molecular graph generative models," ChemRxiv.Preprint, 2020. doi: 10.26434/chemrxiv.12888383.

[11] J. Arús-Pous, T. Blaschke, S. Ulander, J.L. Reymond, H. Chen, and O. Engkvist, "Exploring the GDB-13 chemical space using deep generative models," Journal of Cheminformatics, vol. 11: Article 20, 2019. doi: 10.1186/s13321-019-0341-z.

[12] O. Prykhodko, S.V. Johansson, P.C. Kotsias, J. Arús-Pous, E.J. Bjerrum, O. Engkvist, and Chen, "De novo molecular generation method using latent vector based generative adversarial network," Journal of. Cheminformatics, vol. 11, no 1: Article 74, 2019. doi: 10.1186/s13321-019-0397-9.

[13] D. Polykovskiy et al., "Entangled conditional adversarial autoencoder for de novo drug discovery," Molecular Pharmaceutics, vol. 15, no. 10, pp. 4398-4405, 2018. doi:10.1021/acs.molpharmaceut.8b00839s.

[14] J. Zhang, R. Mercado, O. Engkvist, and H. Chen, "Comparative study of deep generative models on chemical space coverage," ChemRxiv. Preprint., 2020. doi: 10.26434/chemrxiv.13234289.v1.

[15] A. Zhavoronkov et al., "Deep learning enables rapid identification of potent DDR1 kinase inhibitors," Nature Biotechnology, vol. 37, no. 9, pp. 1038-1040, 2019. doi: 10.1038/s41587-019-0224-x.

[16] W.J. Köstler and C.C. Zielinski, "Targeting receptor tyrosine kinases in cancer. In receptor tyrosine kinases: Structure, functions and role in human disease. New York: Spring, pp. 78-225, 2015.

[17] H.M. Kantarjian et al., "Nilotinib versus imatinib for the treatment of patients with newly diagnosed chronic phase, Philadelphia chromosome-positive, chronic myeloid leukaemia: 24-month minimum follow-up of the phase 3 randomised ENESTnd trial," The Lancet Oncology, vol.12, no. 9, pp. 841–851, 2011. doi: 10.1016/S1470-2045(11)70201-7.

[18] F.H. Tan, T.L. Putoczki, S.S. Stylli, and R.B. Luwor, **"**Ponatinib: a novel multi-tyrosine kinase inhibitor against human malignancies," OncoTargets and Therapy, vol. 12, pp. 635–645, 2019. doi: 10.2147/OTT.S189391.

[19] T. O'Hare, "A decade of nilotinib and dasatinib: From in vitro studies to first-line tyrosine kinase inhibitors," Cancer Research, vol.76, no. 20, pp. 5911–5913, 2016. doi: 10.1158/0008-5472.CAN-16-2483.

[20] T.H. Brümmendorf et al., "Bosutinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukaemia: Results from the 24-month follow-up of the BELA trial," British Journal of Haematology, vol. 168, no. 1, pp. 69–81, 2015. doi: 10.1111/bjh.13108.

[21] K.S. Bhullar, N.O. Lagarón, E.M. McGowan, J. Parmar, A. Jha, B.P. Hubbard, and H.P.V. Rupasinghe, "Kinase-targeted cancer therapies: progress, challenges and future directions," Molecular Cancer, vol. 17: 48, 2018. https://doi.org/10.1186/s12943-018-0804-2.

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

[23] M. Hwang at al., "A local region proposals approach to instance segmentation for intestinal polyp detection," International Journal of Machine Learning and Cybernetics, vol. 14, no. 5, pp. 1591-1603, 2023.

[24] A. Huang, X. Ju, J. Lyons, D. Murnane, M. Pettee, and L. Reed, "Heterogeneous graph neural network for identifying hadronically decayed Tau Leptons at the high luminosity LHC," arXiv preprint arXiv:2301.00501, 2023.

[25] K. Keras (2019), "The Python Deep Learning Library," 2015. Available in: https://keras. io.

[26] D. Weininger, “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules,” Journal of Chemical Information and Computer Sciences, vol. 28, no. 1, pp. 31–36, 1988. doi.org/10.1021/ci00057a005.

[27] D. Weininger, A. Weininger, and J.L. Weininger, “SMILES. 2. Algorithm for generation of unique SMILES notation,” Journal of Chemical Information and Computer Sciences, vol. 29, no. 2, pp. 97-101, 1989.

[28] N.M. O'Boyle, “Towards a Universal SMILES representation-A standard method to generate canonical SMILES based on the InChI,” Journal of Cheminformatics, vol. 4, pp. 1–14, 2012.

[29] S. Kim et al., “PubChem 2019 update: improved access to chemical data,” Nuclear Acids Research, vol. 47 (D1), pp. D1102-D1109, 2019.

[30] O. Trott and A.J., “AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,” Journal of. Computational Chemistry, vol. 31, no. 2, pp. 455–461, 2010. doi: 10.1002/jcc.21334.

[31] Y. Ho and S. Wookey, “The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling,” IEEE Access, vol. 8, pp. 4806–4813, 2019.

[32] D.P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[33] G. Landrum, “RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling,” Greg Landrum, 8, 2013.

[34] J.D. Durrant and J.A. McCammon, “NNScore 2.0: A neural-network receptor–ligand scoring function,” Journal of Chemical Information and Modeling, vol. 51, no. 11, pp. 2897–2903, 2011. doi: 10.1021/ci2003889.

[35] M. Wójcikowski, P.J. Ballester, and P. Siedlecki, “Performance of machine-learning scoring functions in structure-based virtual screening,” Scientific Reports, vol. 7, no. 1, pp. 1–10, 2017.

[36] K. Palacio-Rodríguez, I. Lans, C.N. Cavasotto, and P. Cossio, “Exponential consensus ranking improves the outcome in docking and receptor ensemble docking,” Scientific Reports, vol. 9, no. 1: 5142, 2019. doi: 10.1038/s41598-019-41594-3.