

## **Innovative Virtual Screening of PD-L1 Inhibitors: The Synergy of Molecular Similarity, Neural Networks, and GNINA Docking**

Van-Thinh To<sup>a</sup>, Tieu-Long Phan<sup>b, c</sup>, Bao-Vy Ngoc Doan<sup>a</sup>, Phuoc-Chung Van Nguyen<sup>a</sup>, Dong-Nghi Hoang Nguyen<sup>a</sup>, Quang-Huy Nguyen Le<sup>a</sup>, Hoang-Huy Nguyen<sup>a</sup>, The-Chuong Trinh<sup>d</sup>, Tuyen Ngoc Truong<sup>\*a</sup>

*<sup>a</sup>University of Medicine and Pharmacy, Ho Chi Minh city, Faculty of Pharmacy, Ho Chi Minh city, 700000, Vietnam*

*<sup>b</sup>Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16-18, 04107, Leipzig, Germany.*

*<sup>c</sup>Department of Mathematics and Computer Science, University of Southern Denmark, Odense M DK-5230, Denmark*

*<sup>d</sup>Faculty of Pharmacy, Grenoble Alpes University, La Tronche, 38700, France*

### **AUTHOR INFORMATION**

\*a Corresponding Author

E-mail address: [truongtuyen@ump.edu.vn](mailto:truongtuyen@ump.edu.vn)

## Abstract

Immune checkpoint inhibitors have garnered significant attention in oncological research over recent years. A plethora of studies have elucidated that inhibitors targeting the Programmed Death-Ligand 1 (PD-L1) play a pivotal role in circumventing the evasion mechanisms of cancer cells against the immune system. This study aimed to develop an integrated screening model combining an Artificial Neural Network (ANN), Molecular Similarity (MS) assessments, and GNINA 1.0 molecular docking, targeting PD-L1 inhibitors. A database of 2044 substances with known PD-L1 inhibitory activity was compiled from Google Patents and used to enhance molecular similarity evaluations and train the machine learning model. For retrospective validation of the docking procedure, the human PD-L1 protein, with the Protein Data Bank (PDB) ID: 5N2F, was employed as a control. In this phase of the study, 15,235 compounds from the DrugBank database were subjected to a series of screening processes: initially through medicinal chemistry filters, followed by MS assessments, the ANN model, and culminating with molecular docking using GNINA 1.0. The decoy generation yielded promising outcomes, evidenced by an AUC-ROC INN value of 0.52 and Doppelganger scores with a mean of 0.24 and a maximum of 0.346, indicating a high resemblance of the decoys to the active set. For MS, the AVALON emerged as the most effective fingerprint for similarity searching, demonstrating an Enrichment Factor (EF) of 1% at 10.96%, an AUC-ROC of 0.963, and an optimal similarity threshold of 0.32. The ANN model demonstrated superior performance in cross-validation, achieving an average precision of  $0.863 \pm 0.032$  and an F1 score of  $0.745 \pm 0.039$ , outperforming both the Support Vector Classifier (SVC) and Random Forest (RF) models, albeit not significantly. In external validation, the ANN model maintained its superiority with an average precision of 0.851 and an F1 score of 0.790. GNINA 1.0, employed for molecular docking, was validated through redocking and retrospective control, achieving an AUC of 0.975, with a critical `cnn_pose_score` threshold of 0.73. From the initial 15,235 compounds, 128 were shortlisted using the MS and ANN models. Further screening through GNINA 1.0

identified 22 potential candidates, among which (3*S*)-1-(4-acetylphenyl)-5-oxopyrrolidine-3-carboxylic acid emerged as the most promising, with a `cnn_pose_score` of 0.79, a PD-L1 inhibitory probability of 70.5%, and a Tanimoto coefficient of 0.35.

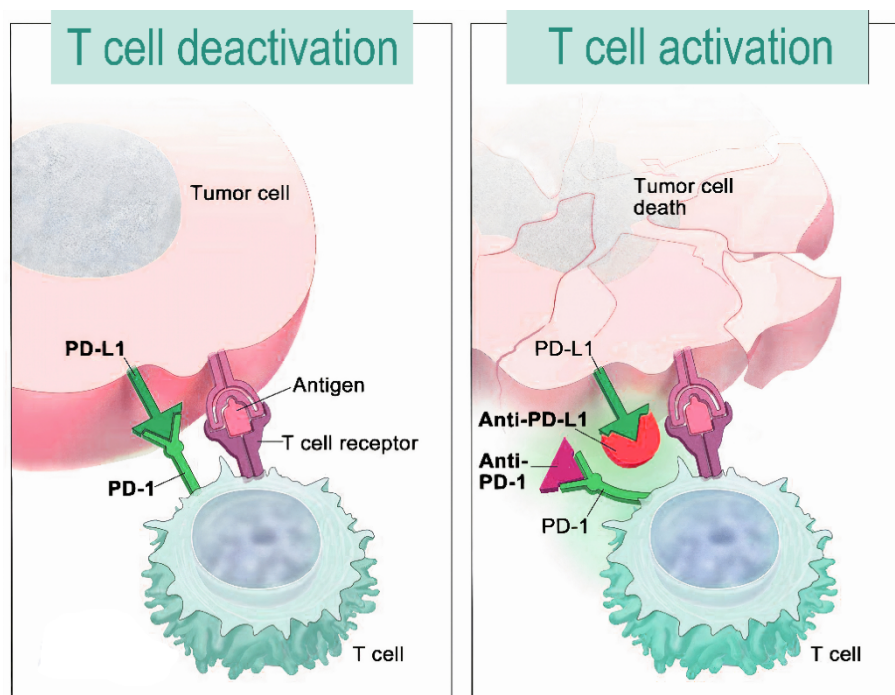
**KEYWORDS:** PD-L1 inhibitors, molecular similarity, GNINA, artificial neural network, virtual screening.

## I. INTRODUCTION

In 2020, the global health landscape was profoundly impacted by cancer, with the International Agency for Research on Cancer (IARC) reporting an staggering 19.3 million new cases and nearly 10 million fatalities<sup>1</sup>. Cancer, fundamentally a conglomerate of diseases manifesting in various organs or tissues, arises from abnormal cell behavior leading to uncontrolled proliferation. These rogue cells not only transgress their original boundaries but also invade adjacent tissues and can metastasize to distant organs<sup>2</sup>. Categorically, cancer is classified into two main types: hematologic cancers, including leukemia, lymphoma, and multiple myeloma, and solid tumor cancers, which impact other body organs or tissues, with breast, prostate, lung, and colorectal cancers being the most prevalent<sup>3</sup>. The genesis of all cancers lies in cellular changes, either in a single cell or a group. Under normal conditions, cellular proliferation is regulated by specific signals that maintain organ-specific cell counts. However, disruptions in these regulatory signals can lead to excessive cell growth, culminating in tumor formation<sup>4</sup>.

Cancer treatment methodologies encompass a variety of modalities, including surgery, radiotherapy, hormone therapy, chemotherapy, targeted therapies, immunotherapy<sup>5</sup>, and gene therapy<sup>6</sup>. These techniques range from physically removing tumors, utilizing ionizing radiation, manipulating hormone production, to correcting genetic defects. Immunotherapy, particularly notable among these, activates the body's immune system, comprising granulocytes, macrophages, natural killer cells, T and B cells, and the lymphatic system, to combat cancer. Immunotherapy, a subset of biological therapy, utilizes bio-derived substances to tackle cancer<sup>7</sup>. Immunotherapy includes various treatments such as Immune Checkpoint Inhibitors (ICIs)<sup>8</sup>, T-cell Transfer Therapy<sup>9</sup>, Monoclonal Antibodies<sup>10</sup>, Immune System Modulators<sup>11</sup>, and Cancer Treatment Vaccines<sup>12</sup>. ICIs function by interfering with the interaction between T cell proteins and their partner proteins on other cells, including cancer cells, which normally send an inhibitory signal to T cells.

By blocking this interaction, ICIs enable T cells to attack cancer cells effectively. PD-L1 and PD-1 inhibitors, a subset of ICIs, specifically target the checkpoint proteins PD-L1 and PD-1 (Figure 1). Tumors often evade immune response by overexpressing PD-L1, which binds to PD-1 on T cells, leading to their deactivation<sup>13</sup>. As of 2023, the FDA has approved six PD-1/PD-L1 inhibitors: Pembrolizumab, Nivolumab, Cemiplimab, Atezolizumab, Avelumab, and Durvalumab<sup>14</sup>.



**Figure 1.** Cancer cells generate a protein known as PD-L1. This protein interacts with the PD-1 protein, which is a transmembrane protein found in T cells. The result of this interaction is the deactivation of T cells. However, the introduction of anti PD-L1 and anti PD-1 can prevent this complex from forming, thereby allowing the T cells to activate. This activation leads to the destruction of cancer cells.

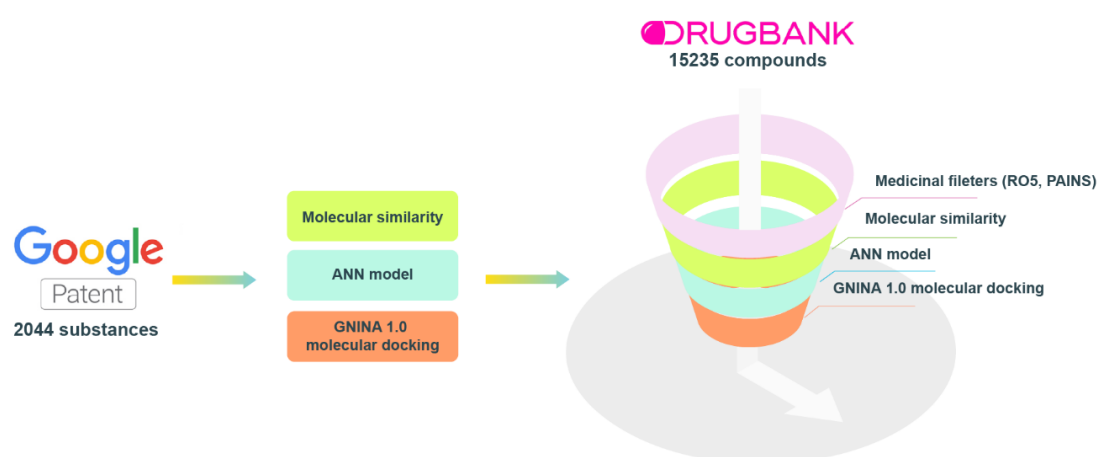
Numerous investigations have been carried out to establish machine learning models for predicting the bioactivity of compounds targeting PD-1 and PD-L1. A study by Geethu et al (2022)<sup>15</sup> screened 32,552 organic compounds from the Atlas database, successfully identifying the top five compounds with the highest potential. Concurrently, Sachin et al (2022)<sup>16</sup> utilized a Random Forest model to analyze 16 million biological molecules, identifying two compounds with inhibitory

concentration (IC<sub>50</sub>) values of 22.35 and 33.65  $\mu$ M, respectively.

The objective of this research was to identify PD-L1 inhibitors using a three-step screening process involving molecular similarity, an ANN model, and GNINA 1.0 molecular docking. The principle of the molecular similarity model lies in evaluating how closely a compound resembles a reference compound in the chemical space<sup>17</sup>. This similarity is measured using techniques such as molecular descriptors, molecular fingerprints, and similarity coefficients. Descriptors provide numerical representations of molecular characteristics, while fingerprints are binary codes representing molecular substructures. Similarity coefficients, such as the Tanimoto and Dice coefficients, quantify similarity based on shared attributes. In this study, molecular fingerprints were utilized to create a two-dimensional similarity model.

## II. METHODS AND MATERIALS

The study was conducted on a system with an AMD Ryzen 9 3900X CPU (12 processors, 3.79 GHz), equipped with 500 GB storage and 96 GB RAM, running Linux 22.04. Molecular fingerprints were generated using RDKit 2020.3.1<sup>18</sup>, the ANN model was developed via PyTorch library<sup>19</sup>, and GNINA 1.0<sup>20</sup> was used for molecular docking. Consistency and reproducibility were ensured by maintaining a constant random state (value = 42) throughout all phases of the study, as illustrated in Figure 2.



**Figure 2.** The workflow of this study.

## Dataset

The models in this study were developed using a dataset of 2,044 two-dimensional (2D) structures, derived from patents US10710986, US10899735, US20200017471A1, and WO2018195321A1. These structures were delineated using ChemDraw 18.1 and saved in both \*.sdf and \*.csv file formats. Energy minimization was subsequently performed using RDKit. Additionally, a repurposing dataset from the DrugBank database, consisting of 15,235 structures either in phase II/III clinical trials or approved as drugs<sup>21</sup>, was utilized for virtual screening.

## Molecular similarity model

BMS-1166 was selected as the query substance for its potent PD-L1 inhibitory activity, evidenced by an IC<sub>50</sub> value of 1.4 nM in a homogeneous time-resolved fluorescence (HTRF) binding assay<sup>22</sup>. The study's active set comprised 2,044 substances, each with a PD-L1 inhibitory IC<sub>50</sub> value under 100 nM. Decoys were generated at a 1:10 ratio (one active substance to ten decoys) using the DeepCoy library<sup>23</sup>. Both active substances and decoys were employed to create five types of molecular fingerprints: AVALON, MACCS, ECFP4, RDK5, and MAP4. A similarity matrix was computed for BMS-1166 in comparison with each active and decoy compound for all fingerprint types, utilizing Tanimoto and Dice coefficients. The Tanimoto coefficient is mathematically defined as:  $T_c(A, B) = \frac{c}{a+b-c}$ ; where 'a' and 'b' represent the number of features in compounds A and B, respectively, and 'c' is the number of features that A and B have in common<sup>24</sup>.

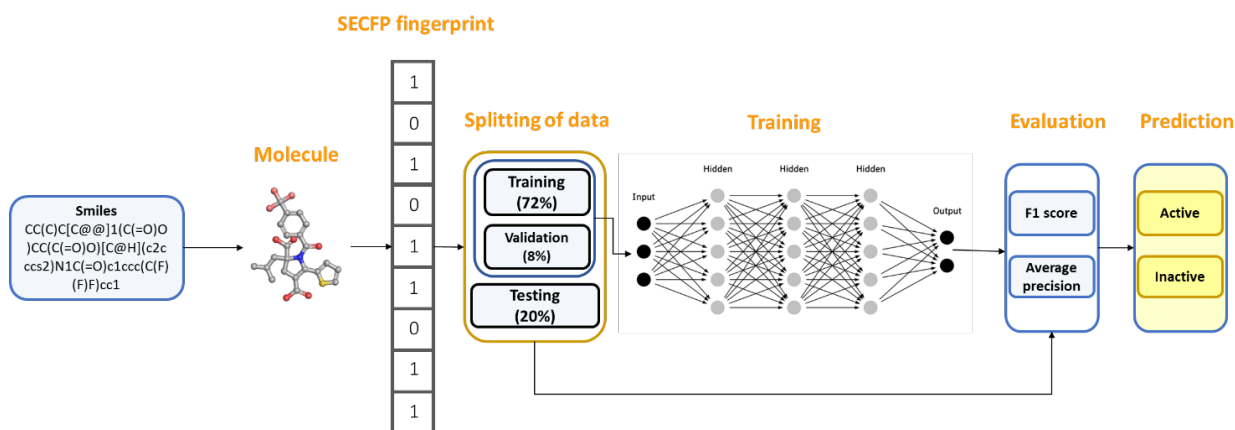
The calculation yielded a data table comprising ten columns of coefficients. To ascertain the most efficient fingerprint and its ideal similarity threshold for virtual screening, the study employed a validation process. This involved analyzing the receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC), enrichment factor (EF), and average precision metrics. The EF is a metric that quantifies the active compounds found within a specified "early recognition" portion of an ordered list, compared to what would be expected from a random distribution.

$$EF_{\chi} \% = \frac{n_s}{N_s} \times \frac{N}{n}$$

where 'n<sub>s</sub>' is the number of active substances in the χ% dataset, 'N<sub>s</sub>' is the number of active substances in the entire dataset, 'n' is the number of substances in the χ% dataset and 'N' is the number of substances in the entire dataset<sup>25</sup>.

### Artificial Neural Networks model

The ANN model, utilizing the SECFP fingerprint as input data, was developed using the PyTorch library and executed on a GPU for accelerated computational efficiency. The dataset, comprising 2,044 substances, was partitioned into training, external validation, and internal validation sets, with allocation ratios of 0.72, 0.2, and 0.08, respectively. For compatibility with GPU-based computations and deep learning operations within PyTorch, the data was transformed into a Tensor network. The training set was further segmented into batches, optimizing computational resource usage, enhancing randomness, and ensuring representative data sampling.



**Figure 3.** Summary of the process of building the ANN model

In the development of the ANN model (illustrated in Figure 3), several critical steps were undertaken. Initially, the number of hidden layers was determined, informed by the problem's complexity and data volume, as the correct number significantly enhances the model's learning capability. Subsequently, the node count in each hidden layer was established, balancing the complexity of the problem against



the model's learning capacity; too few nodes risk underfitting due to limited learning of complex patterns, whereas excess nodes may cause overfitting. Finally, the dropout rate was set. Dropout<sup>26</sup> is a technique used in neural network training to randomly deactivate certain nodes, which helps in mitigating the risk of overfitting. The detailed configuration of the ANN model is outlined in Table 1.

**Table 1.** The specific architecture of the ANN model

<b>Layer</b>	<b>Parameter</b>
<i>Input layer</i>	Linear (in_features = 2048, out_features = 2048)
<i>Activation function</i>	ReLU
<i>Dropout</i>	Dropout (p = 0.5)
<i>First hidden layer</i>	Linear (in_features = 2048, out_features = 1024)
<i>Activation function</i>	ReLU
<i>Dropout</i>	Dropout (p = 0.5)
<i>Second hidden layer</i>	Linear (in_features = 1024, out_features = 256)
<i>Activation function</i>	ReLU
<i>Dropout</i>	Dropout (p = 0.5)
<i>Output layer</i>	Linear (in_features = 256, out_features = 1)
<i>Activation function</i>	Sigmoid

The performance of a model can be evaluated by its learnability from data and generalizability on unseen datasets, performed through internal and external validation, respectively. Internal validation (IV) involves cross-validation techniques for training models and hyperparameter tuning. External validation (EV), on the other hand, utilizes a validation dataset from an independent source to assess the model's performance unbiasedly. As such, the results of EV provide crucial evidence for the generalizability of a QSAR model<sup>27</sup>.

The models' performance in this study were assessed using statistical parameters such as F1-score, and average precision (AP). AP is computed as the weighted mean of precision at each threshold, with the weight being the increase in recall from the previous threshold:<sup>28</sup>  $AP = \sum_n (\text{Recall}_n - \text{Recall}_{n-1}) \times \text{Precision}_n$

The F1-score is calculated as the harmonic mean of precision and recall, providing a measure of the trade-off between them<sup>29</sup>:  $F1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$

### **GNINA 1.0 – molecular docking**

In this research, the PD-L1 protein structure was sourced from the Protein Data Bank (PDB ID: 5N2F)<sup>30</sup>, and underwent standardization, including restoration of missing residues and local energy minimization. MGLTools was used to add both polar and non-polar hydrogen atoms and assign Gasteiger charges to the structure. The docking gridbox was set as a  $22.5 \times 22.5 \times 22.5$  Å cube, centered at coordinates  $x = 32.391$  Å,  $y = 12.721$  Å, and  $z = 133.816$  Å. Butina algorithm<sup>31</sup> was employed to categorize the 2044 substances into 17 clusters, from which 17 centroids were selected alongside a co-crystal ligand, yielding 18 active substances. Utilizing the DeepCoy library, 1800 decoys were generated from these actives at a 1:100 ratio. The resultant dataset of 1818 compounds (actives and decoys) was then used for docking. Initial docking yielded `cnn_pose_score` and `cnn_affinity` scores; GNINA's CNN algorithm then prioritized conformations based on `cnn_pose_score` (ranging from 0 to 1, indicating binding potential), while `cnn_affinity` provided additional binding potential assessment, with higher values indicating stronger binding.

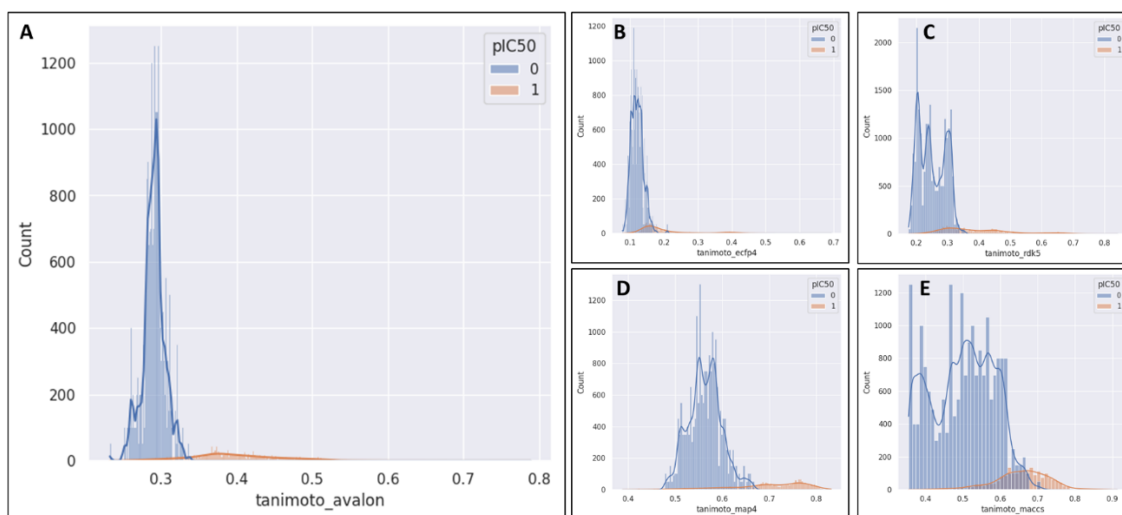
The assessment of the docking model's performance was rigorously conducted through two principal methodologies. Redocking validation was employed to ascertain docking power, quantified by a Root Mean Square Deviation (RMSD) criterion of  $\leq 2$  Å. For evaluating the model's screening efficacy, retrospective control was executed, with the ROC-AUC serving as the key metric. To further refine the analysis, the Geometric Mean (G-Mean) was applied to determine the optimal cut-off for the ROC curve<sup>32</sup>. The G-Mean is an analytical metric designed to evaluate the

equilibrium in classification accuracy between the majority and minority classes, ensuring a comprehensive assessment of the model's performance.

### III. RESULTS AND DISCUSSION

#### Molecular similarity

A total of 22,484 decoys were generated, exhibiting close resemblance to the active compound set, as indicated by the monitoring metrics: AUC-ROC for 1-Nearest Neighbor (1NN) at 0.52, AUC-ROC for Random Forest (RF) at 0.65, and Doppelganger scores with a mean of 0.24 and a maximum of 0.346. The study's initial phase included the computation of a similarity matrix, accessible on the accompanying GitHub repository. As illustrated in Figure 4, the distribution of Tanimoto coefficients for five distinct fingerprints was analyzed. The AVALON fingerprint demonstrated a bifurcation in distribution between decoys and actives. In contrast, the distributions for ECFP4, RDK5, MAP4, and MACCS fingerprints presented considerable overlap, complicating the distinction between decoys and actives. Therefore, the Tanimoto coefficients derived from the AVALON fingerprint are indicative of its superior discriminative capability between decoys and actives.



**Figure 4.** The distribution of Tanimoto coefficients for (A) AVALON fingerprint, (B) ECFP4 fingerprint, (C) RDK5 fingerprint, (D) MAP4 fingerprint, (E) MACCS fingerprint, where '0' indicates a decoy and '1' signifies an active.

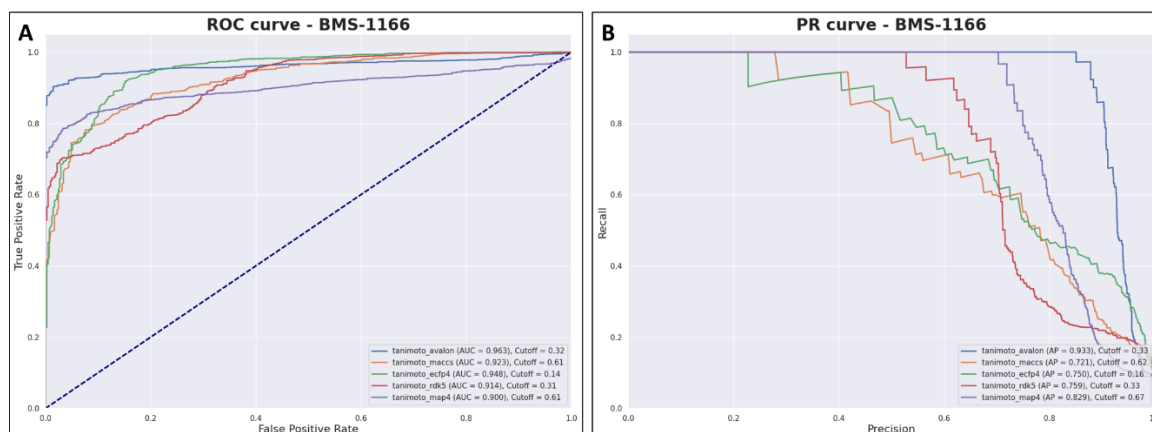
This study utilized EF metrics at the 1%, 5%, and 10% levels of the dataset to

determine the optimal molecular fingerprint for similarity-based virtual screening. Concurrently, two critical metrics in the classification model, namely AUCROC and AP, were utilized to assess the performance of the model (Table 2).

**Table 2.** Results of five molecular similarity models, encompassing AP, AUCROC, LogAUCROC, EF<sub>1%</sub>, EF<sub>5%</sub>, and EF<sub>10%</sub>

Model	AP	AUCROC	Log AUCROC	EF <sub>1%</sub>	EF <sub>5%</sub>	EF <sub>10%</sub>
<b>tanimoto_avalon</b>	0.933	0.963	0.913	10.959	10.998	9.041
<b>tanimoto_maccs</b>	0.721	0.923	0.64	10.959	9.53	6.756
<b>tanimoto_ecfp4</b>	0.75	0.948	0.658	10.959	9.53	7.04
<b>tanimoto_rdk5</b>	0.759	0.914	0.712	10.959	10.558	7.035
<b>tanimoto_map4</b>	0.829	0.9	0.794	10.959	10.998	7.818

Table 2 reveals that the AVALON fingerprint outperformed in similarity searching, achieving an EF<sub>1%</sub> of 10.959%, EF<sub>5%</sub> of 10.998%, and EF<sub>10%</sub> of 9.041%, alongside an AUC-ROC of 0.963 and an AP of 0.933. To define a molecular similarity threshold, AUC-ROC and AP plots were depicted in Figure 5, from which a threshold of 0.32 was derived for optimal virtual screening efficacy.



**Figure 5.** (A) The Area Under the Receiver Operating Characteristic plots and (B) The Average Precision plots for five fingerprints, accompanied by their cutoffs.

## ANN model

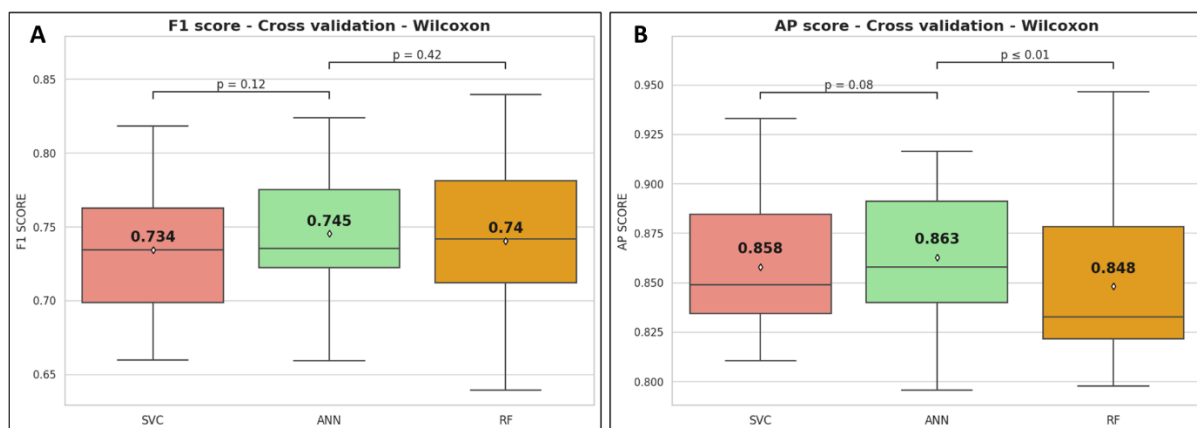
Figure 6 demonstrates a sustained decline in training and validation errors across 170 epochs, indicating the ANN model's convergence, as further training showed no improvement. Simultaneously, a progressive increase in the F1 score and Average Precision (AP) was observed, suggesting efficient learning without overfitting. The training epoch experienced some fluctuations in validation loss and metrics, possibly due to the increased *weight\_decay* hyperparameter that enhanced the rigor of the learning process. Nevertheless, the stable trend in validation loss corroborates the absence of overfitting.



**Figure 6.** The diagram illustrates the training process of ANN model

During the internal cross-validation, the model attained AP of  $0.863 \pm 0.032$  and F1 score of  $0.745 \pm 0.03$ . In the external validation, the model yielded an F1 score of 0.799 and an AP of 0.854.

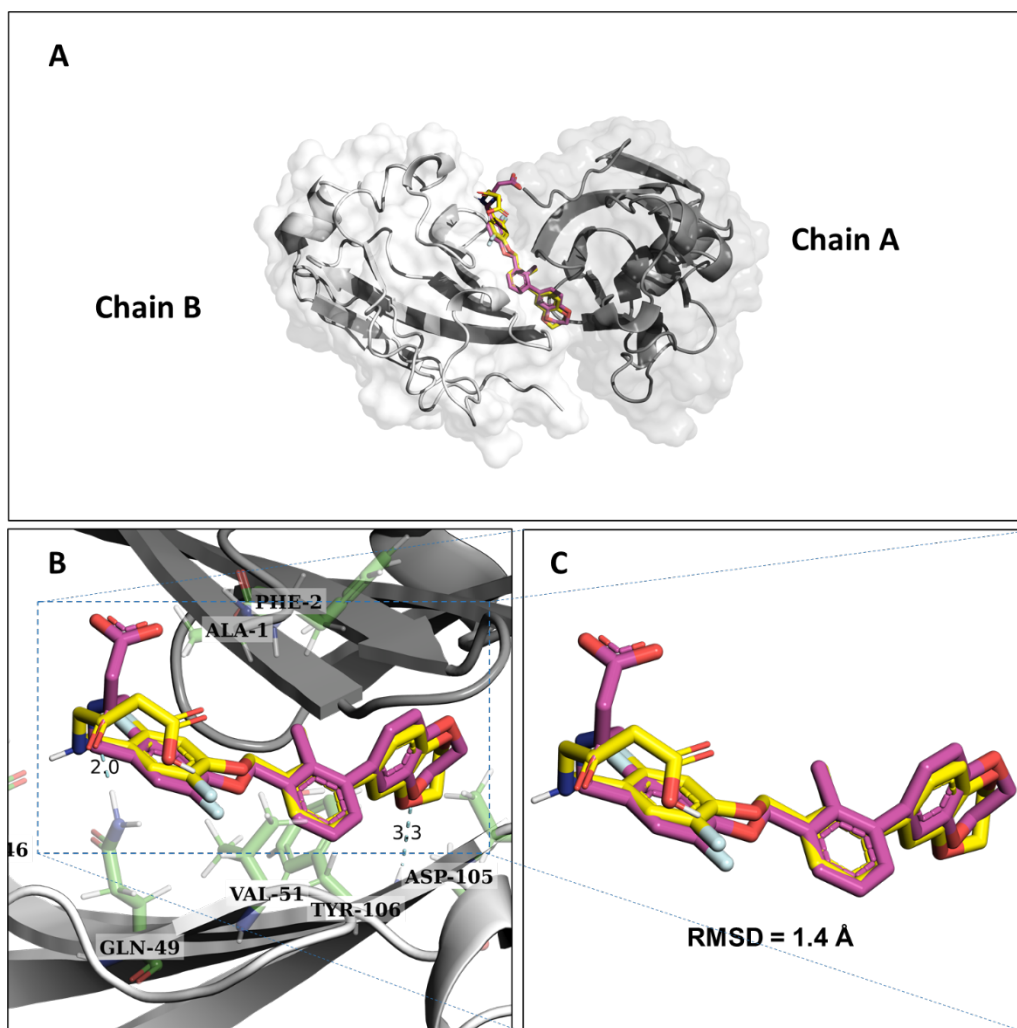
Our study also included a comparative analysis between the ANN model and two other machine learning models based on RF and SVC algorithms. Figure 7 presents the comparative results. The cross-validated F1 score of the ANN model was not statistically significant compared to RF and SVC models, with p-values of 0.42 and 0.12, respectively. However, the ANN model's Average Precision (AP) from cross-validation was statistically significant when compared with the RF model (p-value  $\leq 0.01$ ) but not significant compared to the SVC model (p-value = 0.08).



**Figure 7. (A)** The F1-score comparison of the ANN model with RF and SVC based on the Wilcoxon signed-rank test, **(B)** The AP comparison of the ANN model with RF and SVC based on the Wilcoxon signed-rank test.

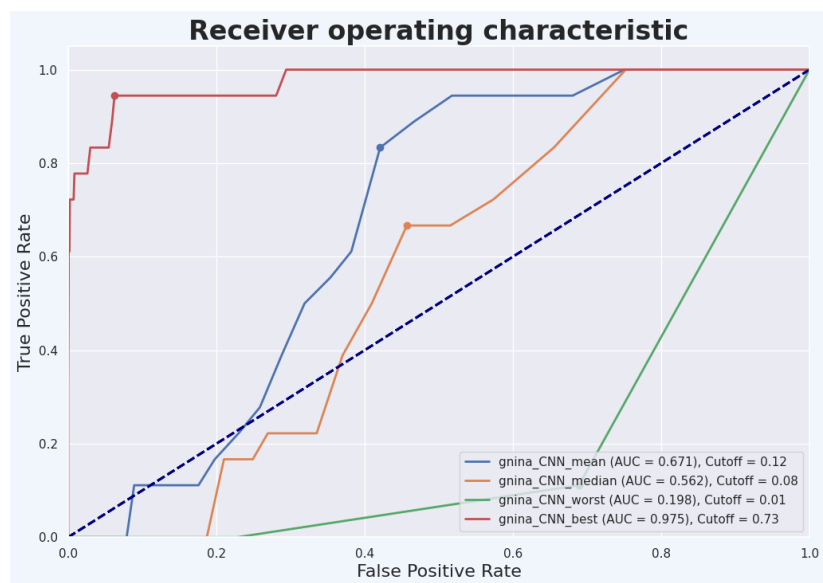
### GNINA 1.0

Utilizing GNINA for molecular docking, the study processed protein and ligand structures in the \*.pdb format. Following the redocking phase, the ten highest-ranking conformations by `cnn_pose_score` were juxtaposed with the co-crystal ligand reference. The most accurately docked conformation exhibited an RMSD of 1.40 Å, well within the acceptable threshold of 2 Å, demonstrating precise docking fidelity as depicted in Figure 8.



**Figure 8.** (A) The best docked conformation is shown in yellow, while the co-crystal ligand is displayed in pink. (B) The binding pocket highlights the hydrogen bonds formed with the sidechain of Gln-49 a characteristic seen in the co-crystal-protein complex. © The RMSD value was found to be 1.4 Å.

According to the AUC-ROC curve in Figure 9, the AUC value was 0.975 for `cnn_pose-score (gnina_CNN_best)`. From this curve, a docking threshold of 0.73 was extrapolated, as per the ROC-AUC analysis.

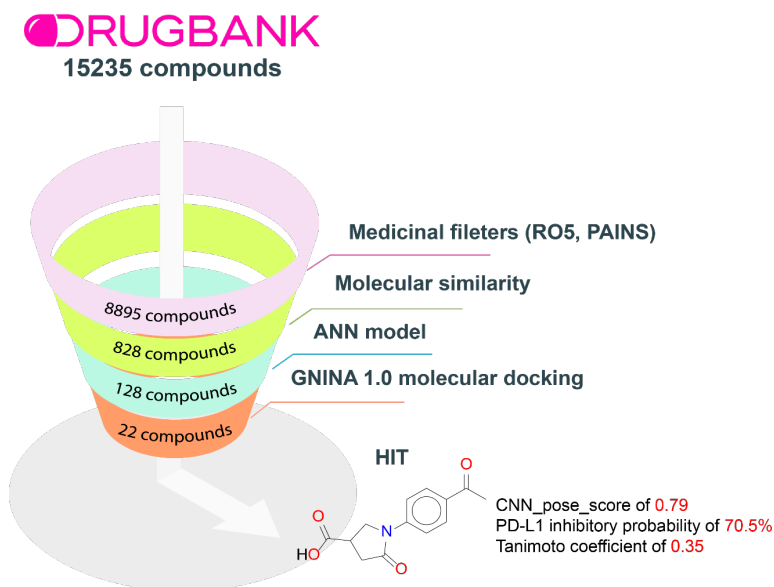


**Figure 9.** The retrospective control was evaluated utilizing four types of conformations, including the smallest positive (gnina\_CNN\_min), the most positive (gnina\_CNN\_max), the median (gnina\_CNN\_median), and mean (gnina\_CNN\_mean) of docked conformation distribution.

### Virtual screening

From the initial set of 15,235 substances in the DrugBank database, 8,895 structures were retained after a medicinal chemistry filter screening. These were then subjected to the molecular similarity model, which identified 828 compounds as active. Subsequently, these 828 compounds were processed through the ANN model, narrowing the list down to 128 compounds. Finally, molecular docking of these 128 compounds led to the identification of 22 promising candidates, comprising 10 known medicines and 12 new hits. The outcomes of this virtual screening process are detailed in Figure 10.

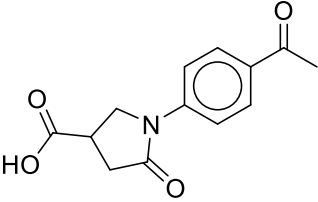


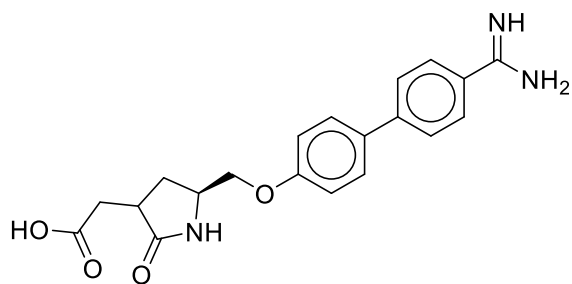


**Figure 10.** The results of virtual screening process

Table 3 highlights the identification of 5 substances characterized by a `cnn_pose_score` exceeding 0.7 and a PD-L1 inhibitory probability surpassing 65%.

**Table 3.** Five substances from the repurposing dataset have a `cnn_pose_score` greater than 0.7 and PD-L1 inhibitory probability greater than 65%

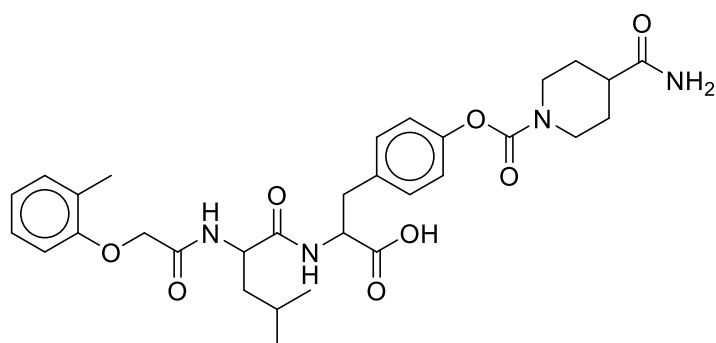
Structure	PD-L1 inhibitory probability (%)	Cnn pose score	Tanimoto coefficient
 <p><b>(3S)-1-(4-acetylphenyl)-5-oxopyrrolidine-3-carboxylic acid</b></p>	70.5	0.79	0.35



69.7

0.89

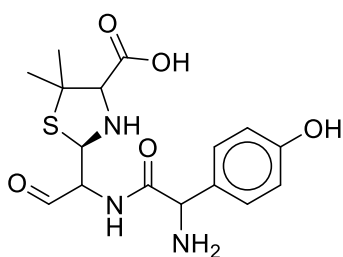
0.34

**Fradafiban**

67.3

0.82

0.38

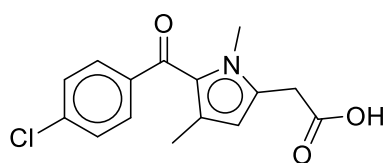
**GW-559090**

66.1

0.76

0.33

**(2R,4S)-2-[(1R)-1-[(2R)-2-Amino-2-(4-hydroxyphenyl) acetyl] amino]-2-oxoethyl]-5,5-dimethyl-1,3-thiazolidine-4-carboxylic acid**



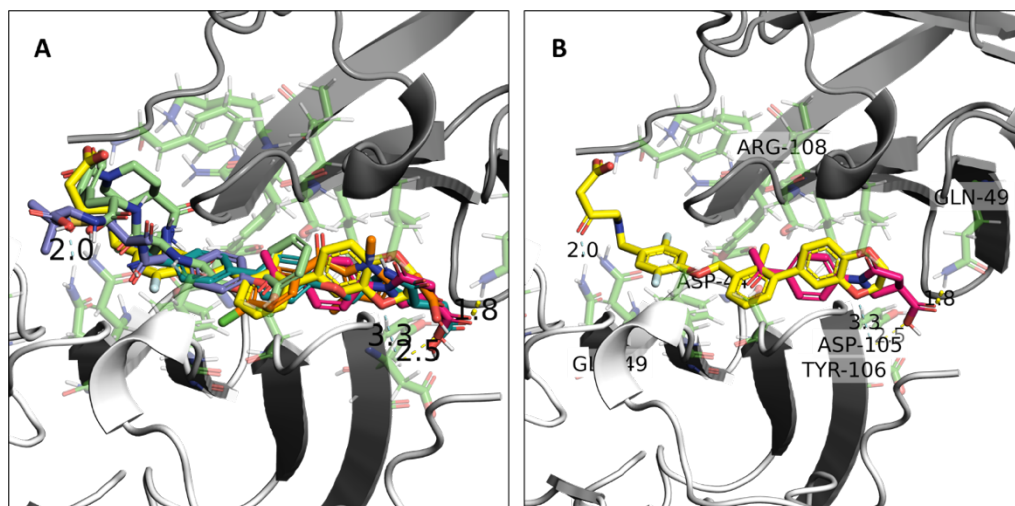
65

0.84

0.34

**Zomepirac**

Out of the five identified compounds, only (3*S*)-1-(4-acetylphenyl)-5-oxopyrrolidine-3-carboxylic acid exhibited a binding pocket analogous to that of the co-crystal ligand. It establishes two hydrogen bonds with Gln49 and Asp105, a crucial interaction anticipated to bolster the conformational stability of (3*S*)-1-(4-acetylphenyl)-5-oxopyrrolidine-3-carboxylic acid within the binding pocket. This stabilization mechanism plays a pivotal role in its PD-L1 inhibitory activity.



**Figure 11.** (A) Binding conformation in the binding pocket of the co-crystal ligand and five candidates (B) Interaction between (3*S*)-1-(4-acetylphenyl)-5-oxopyrrolidine-3-carboxylic acid and the co-crystal ligan with vital amino acid in the active site.

#### IV. CONCLUSIONS

Our study implemented three computational models—molecular similarity, ANN, and the GNINA 1.0 docking procedure—to screen for PD-1/PD-L1 inhibitors. In molecular similarity assessments, the AVALON fingerprint emerged as the superior method, achieving the EF1% of 10.959%, the AUC-ROC of 0.963, and a similarity threshold of 0.32. The ANN model surpassed the SVC and RF models in cross-validation with an AP of  $0.863 \pm 0.032$  and an F1 score of  $0.745 \pm 0.039$ , albeit not significantly. Further external validation reaffirmed the ANN model's superior performance with an AP of 0.854 and an F1 score of 0.799. Following the screening of the repurposing structure library, 22 promising compounds were identified. It is

suggested that these compounds undergo synthesis and subsequent biological activity evaluation to confirm their therapeutic potential.

## **Acknowledgments**

We would like to express our deepest gratitude to the MedAI lab for their generous support, without which this work would not have been possible. Their commitment to fostering innovation and advancing research in the field of medical AI has been instrumental in the successful completion of this study.

## **Data and Software Availability**

The source code, notebooks, and datasets are hosted on GitHub at the following location: [https://github.com/ThinkUMP/PD1\\_PDL1](https://github.com/ThinkUMP/PD1_PDL1)

## **Author Contributions**

V.T.T developed and organized source code for the experiment, wrote and organized the manuscript. T.L.P and B.V.D.N developed the molecular similarity model. P.C.N.V and D.N.H.N collected data, conducted enrichment analysis and visualization task for molecular docking. T.C.T compiled the package and deployed models on Github. H.H.N and Q.H.N.L conducted enrichment analysis source code for molecular docking. T.N.T suggested some concepts and helped to revise the manuscript and guided the development of the project. All authors read and approved the final manuscript.

## References

- 1 Ferlay, J. *et al.* Cancer statistics for the year 2020: An overview. *International journal of cancer*, doi:10.1002/ijc.33588 (2021).
- 2 Organization, W. H. *Cancer*, <[https://www.who.int/health-topics/cancer#tab=tab\\_1](https://www.who.int/health-topics/cancer#tab=tab_1)> (
- 3 Society, A. C. *What is Cancer?*, <<https://www.cancer.org/cancer/understanding-cancer/what-is-cancer.html>> (2022).
- 4 UK, C. R. *How does cancer start?*, <<https://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancer-starts>> (2023).
- 5 Yildizhan, H. *et al.* in *Drug Targeting and Stimuli Sensitive Drug Delivery Systems* (ed Alexandru Mihai Grumezescu) 1-37 (William Andrew Publishing, 2018).
- 6 Cross, D. & Burmester, J. K. Gene therapy for cancer treatment: past, present and future. *Clinical medicine & research* **4**, 218-227, doi:10.3121/cm.4.3.218 (2006).
- 7 Institute, N. C. *Immunotherapy to Treat Cancer*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy>> (2019).
- 8 Institute, N. C. *Immune Checkpoint Inhibitors*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/checkpoint-inhibitors>> (2022).
- 9 Institute, N. C. *T-cell Transfer Therapy*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/t-cell-transfer-therapy>> (2022).
- 10 Institute, N. C. *Monoclonal Antibodies*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/monoclonal-antibodies>> (2019).
- 11 Institute, N. C. *Immune System Modulators*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/immune-system-modulators>> (2019).
- 12 Institute, N. C. *Cancer Treatment Vaccines*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/cancer-treatment-vaccines>> (2019).
- 13 Institute, N. C. *Immune Checkpoint Inhibitors*, <<https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/checkpoint-inhibitors>> (2022).
- 14 Twomey, J. D. & Zhang, B. Cancer Immunotherapy Update: FDA-Approved Checkpoint Inhibitors and Companion Diagnostics. *The AAPS journal* **23**, 39, doi:10.1208/s12248-021-00574-0 (2021).
- 15 Kumar, G. S., Moustafa, M., Sahoo, A. K., Malý, P. & Bharadwaj, S. J. L. Computational investigations on the natural small molecule as an inhibitor of programmed death ligand 1 for cancer immunotherapy. **12**, 659 (2022).
- 16 Patil, S. P. *et al.* Machine-learning guided discovery of bioactive inhibitors of PD1-PDL1 interaction. **15**, 613 (2022).
- 17 Medina-Franco, J. L. & Maggiora, G. M. J. C. f. d. d. Molecular similarity analysis. 343-399 (2013).
- 18 Landrum, G. J. R. Rdkit documentation. **1**, 4 (2013).
- 19 Manaswi, N. K. in *Deep Learning with Applications Using Python : Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras* (ed Navin Kumar Manaswi) 31-43 (Apress, 2018).
- 20 McNutt, A. T. *et al.* GNINA 1.0: molecular docking with deep learning. *Journal of cheminformatics* **13**, 43, doi:10.1186/s13321-021-00522-2 (2021).
- 21 Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for

2018. **46**, D1074-D1082 (2018).
- 22 Chen, F. F., Li, Z., Ma, D. & Yu, Q. Small-molecule PD-L1 inhibitor BMS1166 abrogates the function of PD-L1 by blocking its ER export. *Oncoimmunology* **9**, 1831153, doi:10.1080/2162402x.2020.1831153 (2020).
- 23 Imrie, F., Bradley, A. R. & Deane, C. M. J. B. Generating property-matched decoy molecules using deep learning. **37**, 2134-2141 (2021).
- 24 Maggiora, G., Vogt, M., Stumpfe, D. & Bajorath, J. Molecular similarity in medicinal chemistry. *Journal of medicinal chemistry* **57**, 3186-3204, doi:10.1021/jm401411z (2014).
- 25 Truchon, J. F. & Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *Journal of chemical information and modeling* **47**, 488-508, doi:10.1021/ci600426e (2007).
- 26 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. J. T. j. o. m. l. r. Dropout: a simple way to prevent neural networks from overfitting. **15**, 1929-1958 (2014).
- 27 Ho, S. Y., Phua, K., Wong, L. & Goh, W. W. B. J. P. Extensions of the external validation for checking learned model interpretability and generalizability. **1** (2020).
- 28 Davis, J. & Goadrich, M. in *Proceedings of the 23rd international conference on Machine learning*. 233-240.
- 29 Fernández, A. *et al.* *Learning from imbalanced data sets*. Vol. 10 (Springer, 2018).
- 30 Guzik, K. *et al.* Small-molecule inhibitors of the programmed cell death-1/programmed death-ligand 1 (PD-1/PD-L1) interaction via transiently induced protein states and dimerization of PD-L1. **60**, 5857-5867 (2017).
- 31 Butina, D. J. J. o. C. I. & Sciences, C. Unsupervised data base clustering based on daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. **39**, 747-750 (1999).
- 32 Akosa, J. in *Proceedings of the SAS global forum*. 1.4.