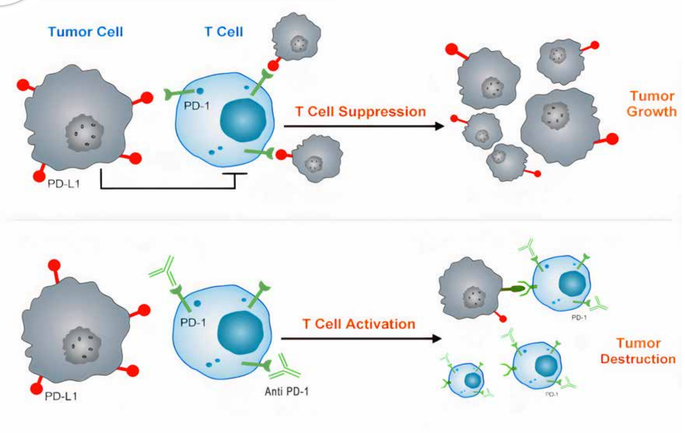
Application of Molecular Similarity and Artificial Neural Networks for PD-L1 inhibitors Virtual Screening

Van-Thinh To, Tieu-Long Phan, Bao-Vy Doan Ngoc, Phuoc-Chung Nguyen Van, Thanh-An Pham, Hoang-Son Lai Le, The-Chuong Trinh, and Tuyen Ngoc Truong

Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam

01 Introduction



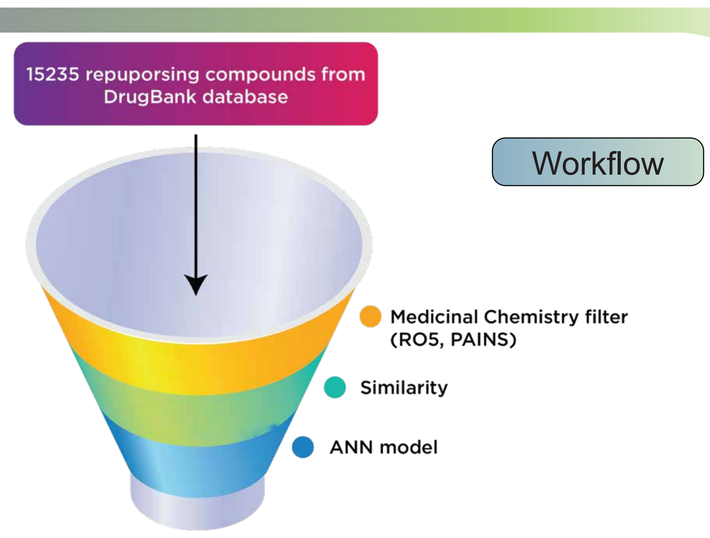
Background

Cancer cells can protect themselves from immune cells by producing PD-L1, which binds to the transmembrane PD-1 protein on T cells and inhibits their activation. Thus, PD-1 and PD-L1 inhibitors can lead to T cell activation, that results in tumor destruction

Research goals

- Building molecular similarity model
- Building ANN model
- Identifying potential drug candidates

Workflow

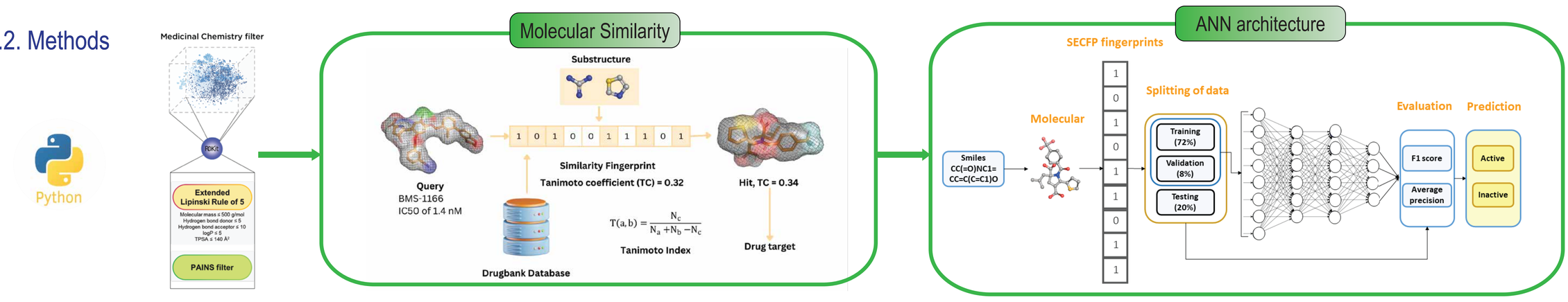


Datasets and methods 02

2.1. Datasets

- Dataset for building the molecular similarity and ANN models: 2,044 substances from Google Patents, splitting them into training, validation, and test sets
- Screening dataset: a repurposing data that contains 15235 compounds from the Drugbank database

2.2. Methods

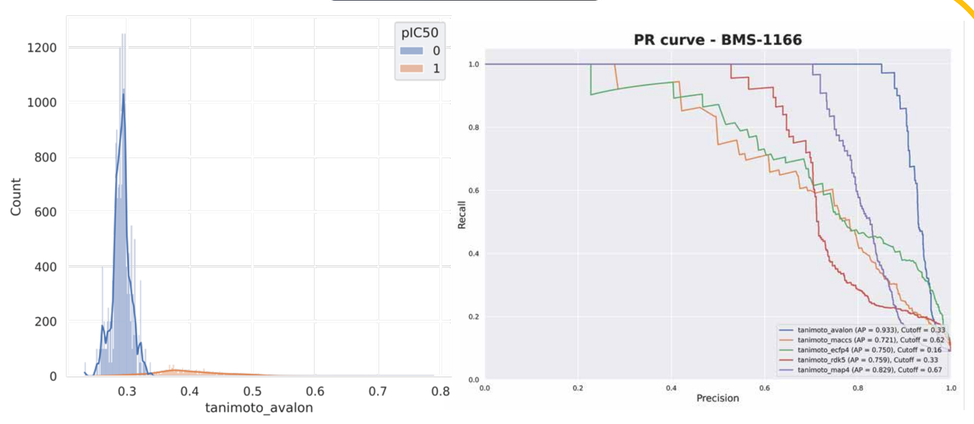


03 Results

Decoy generation results

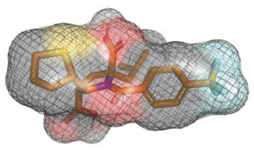
The decoy generation achieved promising results, with AUC-ROC 1NN of 0.52, AUC-ROC RF of 0.65, Doppelganger scores mean of 0.24, and Doppelganger scores max of 0.346, indicating that the decoys closely resemble the active set.

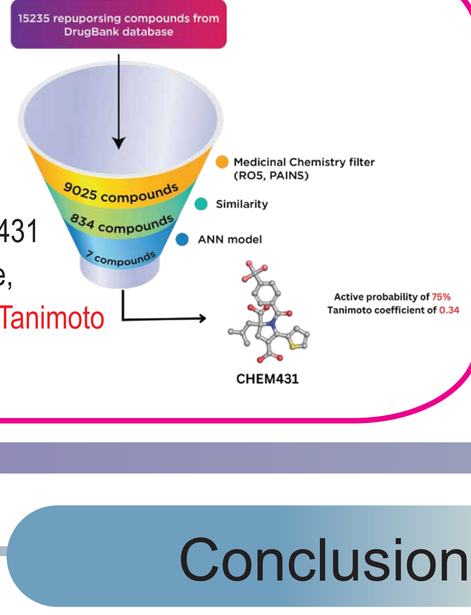
Similarity results



The Avalon fingerprint was the best nominee for similarity searching, with EF1% of 10.99%. The best ROC-AUC of 0.963 was achieved by using BMS-1166 and AVALON fingerprint to measure molecular similarity with a cutoff of 0.32.

Virtual screening results





Only 7 molecules from the Drugbank database fulfilled three filters, with CHEM431 emerging as the most optimistic candidate, possessing active probability of 75% and Tanimoto coefficient of 0.34.

ANN model results

Internal validation

- F1 score 0.745±0.039
- Average precision 0.863±0.032

External validation

- F1 score 0.799
- Average precision 0.747

These metrics were higher than those of the SVC and RF models

The performance of the ANN model showed a significant difference compared to that of the RF model.

Conclusions 04

This study's virtual screening resulted in the 7 most potential substances for PD-L1 inhibitory activity in vitro assay. We recommend to conduct synthesis and test the activity of the four most potential substances. Design more molecular frameworks and optimize in silico processes.

Reference

1. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: An overview. International journal of cancer. 2021;149(4):778-789.
2. Ohaegbulam KC, Assal A, Lazar-Molnar E, Yao Y, Zang X. Human cancer immunotherapy with antibodies to the PD-1 and PD-L1 pathway. Trends in molecular medicine. 2015;21(1):24-33.

Author: Van Thinh To (0972.051.347)
Email: tvthinh.d19@ump.edu.vn