COMPUTATIONALLY ACCELERATING PROTEIN-LIGAND MATCHING FOR NEGLECTED TROPICAL DISEASE: A CASE IN LEISHMANIASIS

Kane Mohamed Hassan;Nkwate Ebenezer;Loic Kwate Dass;

;

IDRC Grant/ Subvention du CRDI: 109187-002-Laying the foundations for artificial intelligence for development (AI4D) in Africa

Computationally Accelerating Protein-Ligand Matching For Neglected Tropical Disease : a case study on Leishmaniasis

Outline

- Main Goal.
- Data Source and Preprocessing.
- First Target Selection.
- Ligand-Target Affinity Score Prediction with Deep Learning (DeepPurpose).
- Molecular Docking with AutoDock Vina.
- Statistical Analysis.

Main Goal

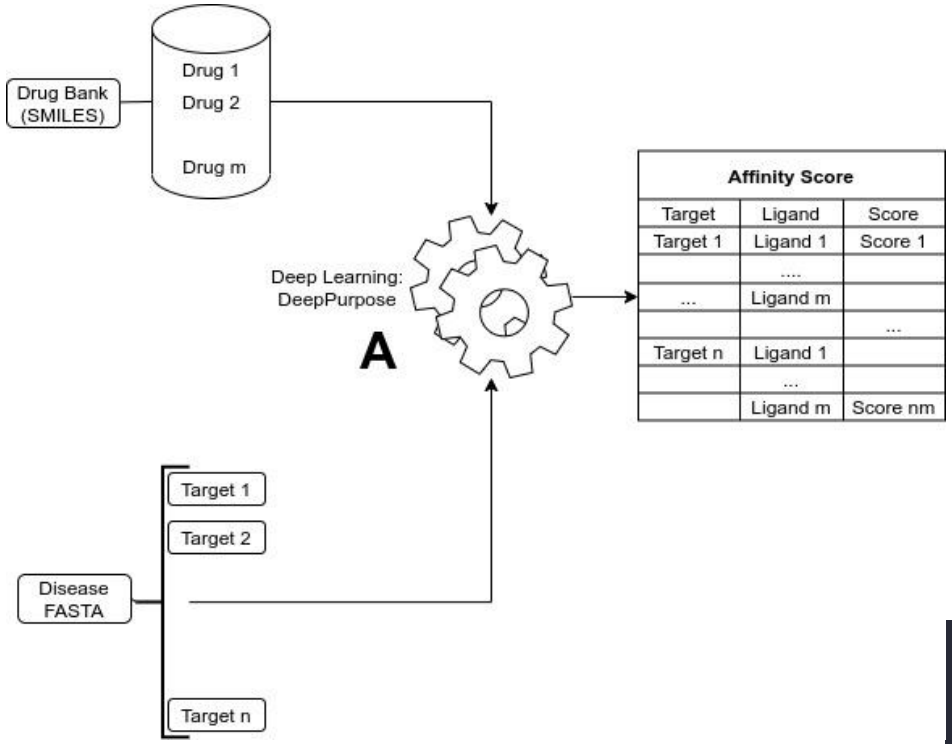
- Applying Drug Repurposing with computational methods to find out medications to cure Leishmaniasis:
  - Deep Learning (DeepPurpose)
  - Molecular Docking (Autodock Vina).
- Analyzing the correlation between Deep Learning based methods and molecular docking tools

Data Source and Preprocessing

- Data source : Indaba Grand Challenge
  - 4020 Ligands in PDB (Protein Data Bank), SDF (Structure Data File) format
  - 507 Targets in PDB format
- Data Processing:
  - Protein conversion (PDB → Fasta Amino Acid Sequence)
    - Tool : OpenBabel
    - Timeout : 10 second/conversion
    - Initial protein list reduced by 32%
  - Ligand conversion (PDB → SMILES)
    - SMILES : Simple Molecular Input Line Entry System

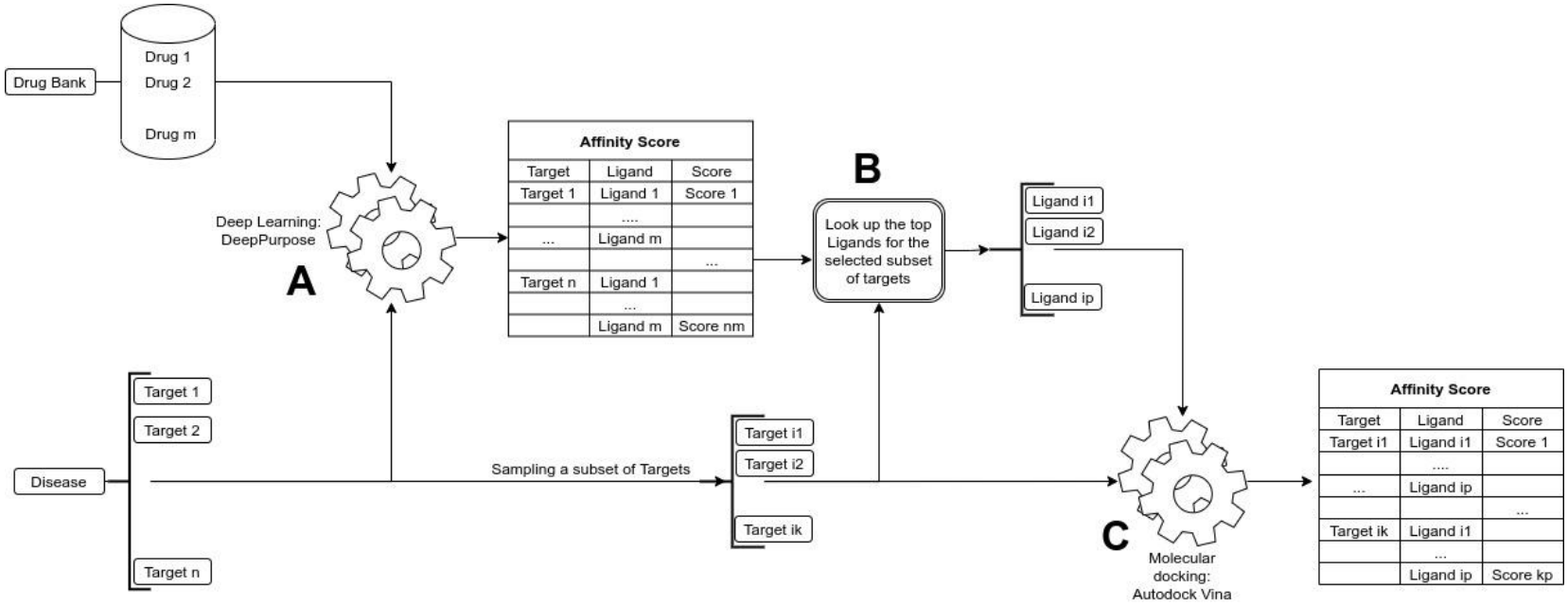
Ligand-Target Affinity Score Prediction with Deep Learning (DeepPurpose)

- Computing Affinity Score energies between all Protein-Ligands Pairs
- Selecting the top 50 drugs for each targets



Molecular Docking (Autodock Vina)

- Sampled 50 targets from the initial list
- Looked up the top 50 drugs for these targets according to Deep Learning outputs
- Computed the best ligand configurations and its corresponding energies using Autodock Vina.



Molecular docking pipeline

Statistical Analysis

- Data form for the analysis:
  - Series of (Ligand, Target, Energy) from DeepPurpose
  - Series of (Ligand, Target, Energy) from Autodock Vina
- Statistical Indicators
  - Kendall rank correlation coefficient
  - Pearson correlation coefficient
  - Spearman's correlation coefficient

Statistical Analysis

	Kendall Coef	Spearman Coef	Pearson Coef
min	-0.33	-0.45	-0.30
max	0.10	0.15	0.07
mean	-0.09	-0.13	-0.12

- Degree of Ligand Overlapping between the two series

Target	% overlapping	Target	% overlapping
A0A5K1USX6.0.apo	60%	P42556.0.apo	50%
E9AQ39.0.apo	50%	Q2PYN0.0.apo	40%
Q4QFE2.0.apo	50%	A1Y2D8.0.apo	40%