Cleveland Clinic Innovations

ImageMol/VideoMol: Framework for Predicting Drug Target and Molecular Properties

Feixiong Cheng, PhD

PRODUCT
Deep learning model

INDICATION
Drug discovery and development

VALUE PROPOSITION
- Unsupervised pretraining deep learning system
- Molecular encoder is designed to extract latent features from ~10 million molecular images

DEVELOPMENT STAGE
Computational model validated for drug discovery for COVID-19

INTELLECTUAL PROPERTY
Patent Pending

RELATED PUBLICATIONS
Zeng, X., Xiang, H., Yu, L. et al. Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. Nat Mach Intell 4, 1004–1016 (2022).

CONTACT INFORMATION
Giedre Ruzgaite, PhD, CLP
Associate Director of Business Development & Licensing
ruzgaig@ccf.org
216.704.0352
CCF ref: IDF 2022-119

UNMET NEED

Drug discovery and development is currently a long and expensive endeavor, with a high failure rate (an average of 12 years from preclinical testing to final approval, at an estimated cost of $3 billion, with an 80-90% failure rate). Computational approaches leveraging biological and chemical datasets to aid in new target identification, candidate molecule design, evaluation of efficacy and compound properties offer promising solutions. A fundamental challenge for computational approaches is accurate molecular representation when considering large (millions) existing and novel compound libraries. Traditional approaches use hand-crafted (e.g. physicochemical) and pharmacophore-based fingerprints which rely on a large amount of domain knowledge, lending to subjective bias. Additional accuracy challenges exist based on limitations of the types and ways in which descriptive molecular and biological characteristics are extracted and represented, often not capturing the dynamic nature of molecule-target interactions in 2D and 3D geometric space.

SOLUTION

The Cheng Lab at the Cleveland Clinic has addressed a number of the above limitations by generating a 2D and 3D based self-supervised image or video representation learning framework for analyzing molecular compounds. The unsupervised pretraining deep learning system receives unlabeled molecular images or videos as input and a molecular encoder that extracts latent features of the molecular inputs. The molecular encoder can be pre-trained according to a variety and plurality of tasks based on structural molecular characteristics. The ImageMol and VideoMol offer a powerful toolbox for computational drug discovery and development by demonstrating high performance in evaluation of molecular properties (e.g. drug's metabolism, brain penetration, toxicity), molecular target profiles (e.g. beta-secretase enzyme, kinases) and identification of clinical drug candidates (e.g. 3C-like protease inhibitors for potential treatment of COVID-19).