Poster: Machine learning applications for drug discovery

Author

Mariya Ivanova

New Vistas

Volume 12, Issue 1, 2026, UWL Annual Doctoral Students' Conference 2025

Abstract

To aid drug discovery in neurodegeneration, we created four unique computational methodologies, leveraging the capabilities of scikit-learn and PyTorch. Despite being developed for neurodegeneration, these methodologies hold potential for application in other medical fields.

The ML predictions of the first variant were based on carbon-13 isotope and proton nuclear magnetic resonance (13CNMR ,1HNMR) spectroscopic data originating from the Simplified Molecular Input Line Entry System (SMILES) notations of small biomolecules. The conversion into spectroscopic data was carried on by the NMRDB software. We utilized case studies to illustrate the predictive modelling of the DNA Damage-Inducible Transcript 3 (CHOP); Transthyretin transcription activators; human dopamine D1 receptor antagonists.

The second approach was based on atomic features of small biomolecules provided by PubChem, the world's largest collection of freely accessible chemical information, or calculated additionally by us. The case studies were on predicting the active G9a inhibitors and their efficacy magnitude.

Despite appearing to contradict established machine learning principles, the third variant predicted small biomolecule functionalities using only PubChem identifiers. We explored this approach because PubChem identifiers encapsulate structural and similarity information. The methodology was demonstrated through the prediction of D3 and D1 dopamine receptors` antagonists; activators of
the Rab9 promoter, inhibitors of the DNA damage?inducible transcript 3 (CHOP), antagonists of the M1 muscarinic receptor and G9a inhibitors, Tyrosyl?DNA phosphodiesterase 1(TDP1) inhibitors, and the orphan G-protein coupled receptor 151 (GPR 151).

In the fourth methodology, we extracted information from chemical names generated according to the International Union of Pure and Applied Chemistry (IUPAC) nomenclature. The aim was to order the small biomolecule's functional groups by decreasing expected functionality.

The case study was focused on the Tyrosyl-DNA phosphodiesterase 1(TDP1) inhibitors.These applications would reduce drug discovery
costs and time beyond the studied cases.

Keywords

Machine learning, Drug discovery