eMOSAIC: Multi-modal Out-of-distribution Uncertainty Quantification Streamlines Large-scale Polypharmacology

a model agnostic approach

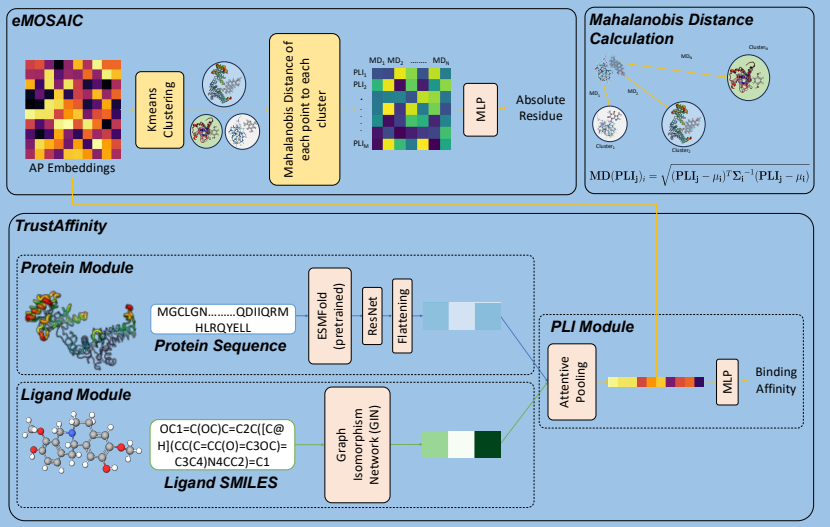
BACKGROUND

Polypharmacology requires screening large libraries of compounds across multiple protein targets, demanding accurate, scalable methods for predicting binding affinity. Existing structure-based and sequence-based methods struggle with accuracy and scalability.

METHODS

1. Trained on the ChEMBL31 dataset, with 350,400 protein-ligand interaction (PLI) pairs. Data is split using OOD scaffold-based split, to evaluate generalization on unseen chemical scaffolds same as real-world screening.
2. eMOSAIC integrates a multi-modal deep learning framework called TrustAffinity, which combines protein sequence and ligand embeddings. Protein sequences are processed with a pre-trained protein language model, while ligands are represented as graphs using a graph neural network. These embeddings are fused using attention pooling to predict binding affinities.
3. eMOSAIC clusters training embeddings and computes Mahalanobis distances for new cases to detect outliers. This distance-based approach enables accurate uncertainty estimation for predictions, enhancing model confidence in out-of-distribution (OOD) cases.
4. We evaluate the performance of eMOSAIC and compare it to baselines including DL-based sequence-based, traditional docking, and DL=based docking.

Can Deep Learning Derisk and Accelerate Drug Discovery?



Yes, it can, with promising results!

Amitesh Badkul

Li Xie

Shuo Zhang

Lei Xie

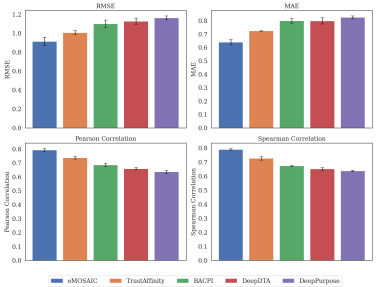
THE GRADUATE CENTER CITY UNIVERSITY OF NEW YORK

Weill Cornell Medicine

HUNTER

RESULTS

Binding Affinity Prediction



Polypharmacology Results

Protein Type	Model	Accuracy ↑	Precision ↑	Recall ↑	F-1 Score ↑	Hamming Loss ↓
GPCR	AutoDock	0.502	0.501	0.484	0.387	0.499
	KarmaDock	0.51	0.51	0.143	0.124	0.49
	DeepPurpose	0.724	0.696	0.678	0.668	0.277
	DeepDTA	0.743	0.716	0.699	0.688	0.256
	BACPI	0.74	0.74	0.722	0.652	0.26
	eMOSAIC	**0.902**	**0.886**	**0.869**	**0.871**	**0.098**
Kinase	AutoDock	0.366	0.367	**0.858**	0.416	0.633
	KarmaDock	0.414	0.414	0.735	0.38	0.586
	DeepPurpose	0.698	0.609	0.637	0.595	0.302
	DeepDTA	0.714	0.635	0.627	0.605	0.286
	BACPI	0.681	**0.681**	0.407	0.312	0.319
	eMOSAIC	**0.758**	0.662	0.648	**0.622**	**0.242**

Uncertainty Quantification Results