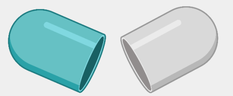
DTBAPRED: IMPROVED PREDICTION OF DRUG-TARGET BINDING AFFINITY USING MACHINE LEARNING APPROACH

Mohamed M. Hussein, Saleh Musleh, Tanvir Alam
College of Science and Engineering, Hamad Bin Khalifa University (HBKU), Doha, Qatar.

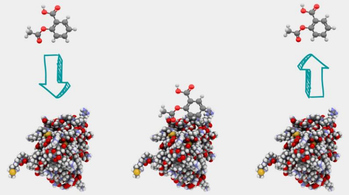
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY
عضو في مؤسسة قطر
Member of Qatar Foundation

SCAN ME



Introduction

Drug-Target Interaction



The drug interacts with the target protein at the binding site, forming temporary bonds.

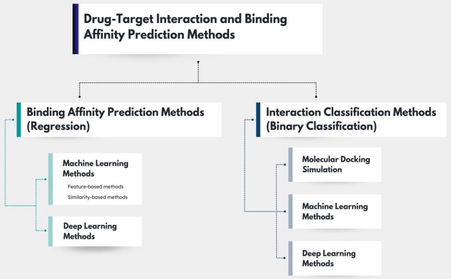
The drug induces a positive or negative change in the biological target.

The drug consequently leaves the target

Drugs function by targeting specific proteins, enzymes, or other biological molecules that play a role in the disease process. By interacting with these targets, drugs can help treat or cure the illness.

Drugs function by targeting specific proteins, enzymes, or other biological molecules that play a role in the disease process. By interacting with these targets, drugs can help treat or cure the illness.

Literature Review



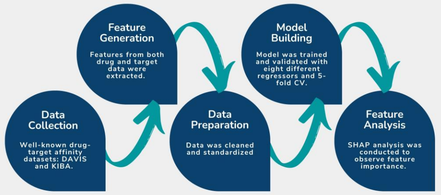
Molecular Docking vs Machine Learning

Molecular Docking	Machine Learning
Require 3D structure availability.	Can work with 2D datasets
Require large computational resources	Can work with accessible computational resources
Low efficiency for large-scale datasets	High efficiency for large-scale datasets

Model	Publication Year	Drug Representation	Protein Representation	Proposed Model
KronRLS	2014	PubChem Sim Chemical kernels	Smith-Waterman similarity score	KronRLS model (Similarity-based)
SimBoost	2017	PubChem Sim and network features	Smith-Waterman similarity score	Gradient boosting
GANsDTA	2020	SMILES + GAN	Amino-acid sequences + GAN	CNN with ReLU
DeepCDA	2020	SMILES (LSTM+CNN)	Amino-acid sequences (LSTM+CNN)	Two-sided Attention Mechanism
Affinity2Vec	2022	SMILES + Seq2seq embedding fingerprints	Amino-acid sequences	XGBoost

Methods

DTBAPred Workflow

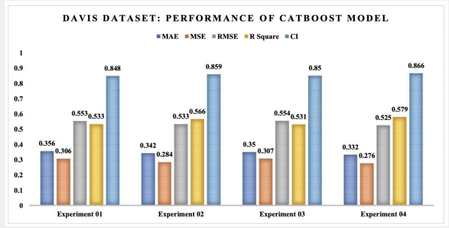


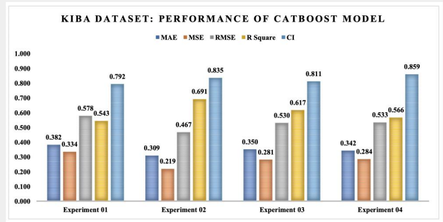
The model was evaluated on DAVIS and KIBA datasets to determine which has the best overall performance. Four experimental settings were tested.

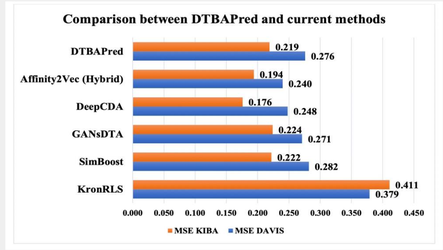
Experiment 01	- ***Drug Features***: *Morgan fingerprints* - ***Protein Features***: *AAC*
Experiment 02	- ***Drug Features***: *Morgan fingerprints* - ***Protein Features***: *AAC, CTDC, CTDD, CTDT, P-AAC*
Experiment 03	- ***Drug Features***: *RDKit 2D descriptors* - ***Protein Features***: *AAC*
Experiment 04	- ***Drug Features***: *RDKit 2D descriptors* - ***Protein Features***: *AAC, CTDC, CTDD, CTDT, P-AAC*

Results

CatBoost regressor had the best overall performance with the lowest MAE, MSE, and RMSE, and high R-Square and CI values.







Feature	Type		Group	Explanation
	Drug	Target		
Xc1.Y		✅	P-AAC	Frequency of Tyrosine (Y) in a protein sequence.
SlogP_VSAB	✅		RDKit-2D	The partition coefficient of a drug.
fr_aniline	✅		RDKit-2D	Number of Aniline substructures in a drug.
NumHDonors	✅		RDKit-2D	Number of hydrogen atoms that can act as hydrogen bond donors.
NHOHCount	✅		RDKit-2D	Number of nitrogen-oxygen bonds in a drug.

Conclusion

DTBAPred was developed with CatBoost to predict DTBA. Two drug feature groups and five protein feature groups were used. The combination of RDKit-2D descriptors and five protein descriptors showed promising results. SHAP analysis explained important features.

Acknowledgements

This work was supported by the College of Science and Engineering, Hamad Bin Khalifa University (HBKU), Doha, Qatar.

سدرة للطب
Sidra Medicine

Welcome to our tech-focused conference! Our poster presentation features an innovative AR experience that you can access by scanning the image below. See the concepts come to life and get a deeper understanding of the topic at hand.

Don't forget to turn up the volume on your device for an explanation. During the networking breaks, feel free to connect with the presenters to discuss the topic further.

Thanks for joining us!