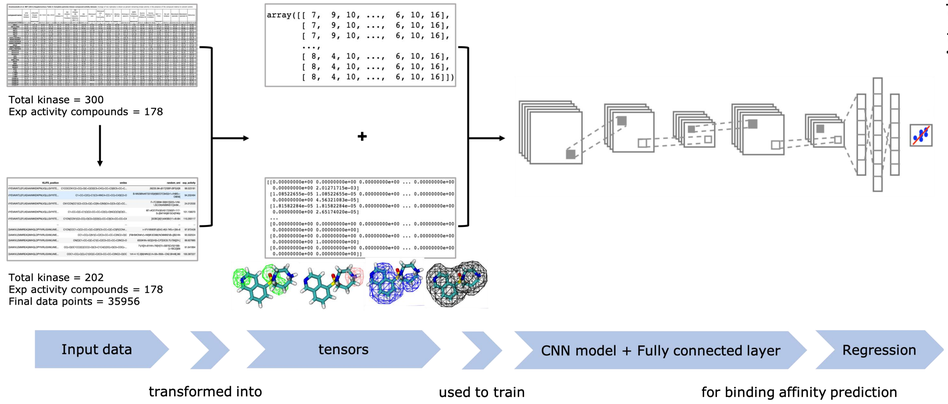
KinBADL: Developing a 3D CNN to Predict Kinase-Inhibitor Binding Affinity

Wern Juin Gabriel Ong[1 2], Palani Kirubakaran[1], Grigorii V. Andrianov[1], and John Karanicolas[1] | Karanicolas Lab -- Fox Chase Cancer Center
Fox Chase Cancer Center[1] and Bowdoin College[2]

Motivations

How do we develop a generalizable deep learning model for predicting kinase-inhibitor binding affinities?

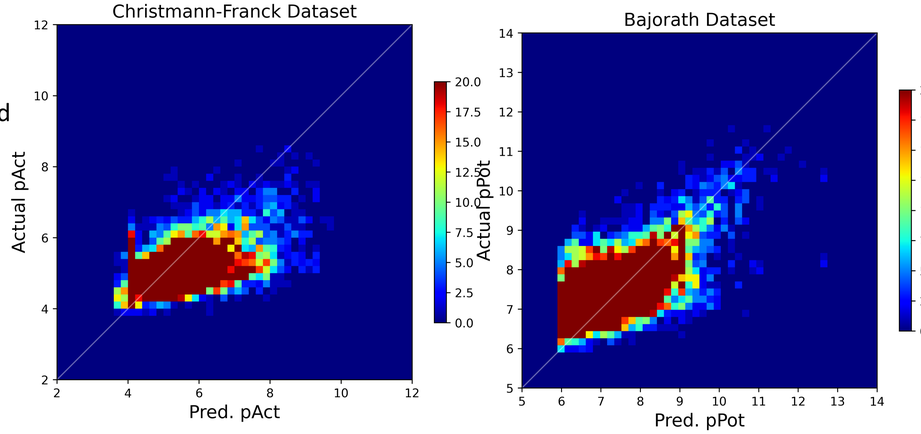
- Kinases are highly druggable biological targets broadly implicated in biological disorders.
- Developing a screening pipeline that can accurately identify kinase inhibitors will significantly streamline the screening of compound libraries.
- Previous work in the Karanicolas Lab has demonstrated the need for structure-based ligand representations.



Methods

- Ligand Representation: 3D Pharmacophores
  - Conformers for compound are generated, aligned with closest ligand in PDB.
  - The best aligned conformer is used to generate the 3D pharmacophore for model training
- Protein Representation: SeqVec and Onehot Encoded Representations from KLIFS sequences (ATP binding site).
  - SeqVec is a novel protein representation inspired by NLP model ELMo.
  - No difference between SeqVec and Onehot Enoded Representations
- Model Training
  - Training was conducted on both Argonne National Lab's GPU cluster and Google Colab.

Results



Dataset	Pearson R	MSE	RMSE	MAE	CI
Bajorath	0.672	0.579	0.760	0.580	0.724
Christmann-Franck	0.531	0.918	0.958	0.746	0.668

Discussion

- Our model performs comparably to others in the literature, exhibiting a lower MSE (RosENet: 1.24; Kdeep: 1.27).
- Our model also displays a surprisingly high concordance index – even if our model is not predicting actual binding affinities, it is able to preserve the rank of inhibitors in its predictions.
- Performance varies among kinases: e.g. PIM3 has Pearson Correlation of 0.79 and CDK9 has Person Correlation of 0.51
- The Bajorath dataset contains more ligand diversity than the Christmann-Franck dataset and is the likely cause of the performance differential.

Future Directions

- Rebuilding the Christmann-Franck dataset, filtering for compounds that have close analogues recorded in PDB.
- Using the top few conformers instead of solely the top 1 conformer.
- A transfer learning regime – first train our model to predict Rosetta Energies and use these weights as the starting point for binding affinity training.
- Adding a structural kinase representation.

Acknowledgements

Many thanks to Kiruba, Grigorii, John, and Austin for their suggestions and assistance throughout this project.

References

Christmann-Franck, Serge, et al. "Unprecedently Large-Scale Kinase Inhibitor Set Enabling the Accurate Prediction of Compound–Kinase Activities: A Way toward Selective Promiscuity by Design?" *Journal of Chemical Information and Modeling*, vol. 204, 2 Aug. 2016, pp. 1654-1675, doi:.org/10.1021/acs.jcim.6b00122.
Francoeur, Paul G., et al. "Three-Dimensional Convolutional Neural Networks and a Cross-Docked Data Set for Structure-Based Drug Design." *Journal of Chemical Information and Modeling*, vol. 60, no. 9, 31 Aug. 2020, doi:.org/10.1021/acs.jcim.0c00411.

Hassan-Harrirou, Hussein, et al. "RosENet: Improving Binding Affinity Prediction by Leveraging Molecular Mechanics Energies with an Ensemble of 3D Convolutional Neural Networks." *Journal of Chemical Information and Modeling*, vol. 60, no. 6, 11 May 2020, pp. 2791-2802, doi:.org/10.1021/acs.jcim.0c00075.

Heinzinger, Michael, et al. "Modeling aspects of the language of life through transfer-learning protein sequences." *BMC Bioinformatics*, vol. 723, 17 Dec. 2019, doi:.org/10.1186/s12859-019-3220-8.

Jiménez, Jose, et al. "KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks." *Journal of Chemical Information and Modeling*, vol. 58, no. 2, 8 Jan. 2018, pp. 287-296, doi:.org/10.1021/acs.jcim.7b00650.
Laufkötter, Oliver, et al. "Identifying representative kinases for inhibitor evaluation via systematic analysis of compound-based target relationships." *European Journal of Medicinal Chemistry*, vol. 204, 15 Oct. 2020, doi:.org/10.1016/j.ejmech.2020.112641.

Rifaioglu, Ahmet S., et al. "MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery." *Bioinformatics*, 7 Dec. 2020, doi:.org/10.1093/bioinformatics/btaa858.