Rishal Aggarwal

@AggarwalRishal | rishal.aggarwal@gmail.com | Rishal Aggarwal | RishalAggarwal

Education

CMU-Pitt Computaional Biology — Pittsburgh, USA
PhD Student — *Aug. 2022 – Present*

International Institute of Information Technology — Hyderabad, India
MS by Research in Bioinformatics — *Jan. 2021 – July 2022*

Birla Institute of Technology and Sciences, Pilani — Rajasthan, India
Bachelor of Pharmacy — *Aug. 2016 – May 2020*

Publications and Preprints

- **Rishal Aggarwal***, Bhuvanesh Sridharan*, Manan Goel*, Pal, P., and U. Deva Priyakumar, Efficient and Enhanced Sampling of Chemical Space for Targeted Drug Design. *Wiley Interdisciplinary Reviews: Computational Molecular Science,* 2022.
- Kanakala G. Chandan, **Rishal Aggarwal**, Divya Nayar, and U. Deva Priyakumar, Latent Biases in Machine Learning Models for Predicting Binding Affinities using Popular Datasets. *Journal of Chemical Information and Modeling,* (Submitted) 2022.
- **Rishal Aggarwal**, Akash Gupta, Vineeth Chelur, C. V. Jawahar, and U. Deva Priyakumar. DeepPocket: Ligand Binding Site Detection and Segmentation using 3D Convolutional Neural Networks. *Journal of Chemical Information and Modeling,* 2021.
- Viraj Bagal, **Rishal Aggarwal**, P. K. Vinod, and U. Deva Priyakumar. MolGPT: Molecular Generation Using a Transformer-Decoder Model. *Journal of Chemical Information and Modeling,* 2021.
- Andrew T McNutt, Paul Francoeur, **Rishal Aggarwal**, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. GNINA 1.0: molecular docking with deep learning. *Journal of cheminformatics,* 13(1):1–20, 2021.
- **Rishal Aggarwal** and David R Koes. Learning RMSD to Improve Protein-Ligand Scoring and Pose Selection. *ChemArxiv* 2020.

Workshop Papers

- **ICML Computational Biology 2021:** APOBind: A Dataset of Ligand Unbound Protein Conformations for Machine Learning Applications in De Novo Drug Design. **Rishal Aggarwal**, Akash Gupta, U. Deva Priyakumar.
- **AAAI SDA 2021:** LigGPT: Molecular Generation Using a Transformer-Decoder Model. Viraj Bagal, **Rishal Aggarwal**, P. K. Vinod, and U. Deva Priyakumar.

Research Experience

Graduate RA / PhD student - Koes Group — August 2022 – Present
CMU - Pitt Computational Biology — *Pittsburgh, USA*

- **Pharmnn:** Development of neural network models for automated design of pharmacophores.
- **Energy based docking:** Designing deep generative models for molecular conformation generation and small molecule docking

Graduate Research Assistant - Deva Lab — May 2020 – August 2022
Center for Computational Natural Sciences and Bioinformatics — *IIITH, India*

- **DeepPocket:** Developed an AI workflow for the detection and segmentation of ligand binding sites on 3D structures of proteins using Convolutional Neural Networks. **SoTA results published in ACS JCIM. Link**
- **MolGPT:** Showcased the performance of attention models for molecular generation and its utility in the inverse design of molecules conditioned on user specified scaffolds and property values. **Published in ACS JCIM. Link**

- **APOBind:** Generated a dataset of unbound protein structures for the protein-ligand complexes present in PDBbind to improve the performance of binding affinity prediction and binding site detection methods on apo structures. Popular docking algorithms when tested on apo structures of this dataset also performed poorly, highlighting the utility of this dataset for methods development; **Accepted, ICML Computational Biology Workshop 2021. Link**
- **Book Chapter:** Authored a book chapter that reviews recent AI and ML methods for accelerating drug design covering several topics such as Protein Structure Prediction, Binding Site Detection, Virtual Screening, Molecular Generation and Molecular Property Prediction (yet to be published).
- **Binding Affinity Datasets:** Exploring hidden biases in protein-ligand datasets that affect performances of ML models in order to provide a holistic view of deep learning for virtual screening. **Published in ACS Omega. Link**

Research Fellow - Koes Group Aug. 2019 – July. 2020
University of Pittsburgh *Pittsburgh, USA*

- **Protein-Ligand Scoring:** Developed a model to predict the distance between an arbitrary ligand pose and the true binding pose of a protein-ligand complex with high correlation (0.768). Showed improved performance over baseline empirical scoring functions for pose ranking and docking. **Work published on ChemArxiv. Link**
- **Pose Optimization:** Also explored improving deep learning potentials by directly training on a loss applied to the gradients returned from the Convolutional Neural Network to enable optimization of ligand poses to more ideal positions.
- **Gnina Contributions:** Contributed to the **Gnina** project by creating an option for merging empirical and deep learning potentials for optimization of ligand docked pose via the BFGS algorithm to more sterically plausible positions. **GNINA 1.0 published in the Journal of Cheminformatics. Link**
- **RNA Secondary Structure Prediction:** Worked on a new ML architecture, that uses a push down automata to generate strings according to a set of rules or grammar (like balancing of parenthesis). Showed better performance than vanilla deep learning architectures like the GRU for RNA Secondary Structure Prediction. **Code**

Internships

- **GE Healthcare:** Worked on segmentation of the vertebrae column using volumetric (3D) MRI body scan images and 3D Convolutional Networks. Achieved Dice Coefficient values above 0.85
- **Pixxel:** Worked at Pixxel, a space-tech startup at BITS Pilani, on developing algorithms that could predict amount of crop harvest at the end of Rabi and Kharif crop seasons in India using multispectral Landsat imagery.
- **Lazada (An Alibaba Company):** Worked on improving customer experience on an e-commerce website by removing duplicate product listings by matching title using NLP tools and string manipulation. Also worked on matching similar product listings from different languages using google translate.

Projects

- **Alzheimers Detection:** Developed a data pre-processing pipeline for 3D T1 weighted brain MRI scans to be fed into a Convolutional Neural Network for diagnosis Alzheimer's Disease. The repository has 25 stars and 14 forks on github. **Code**
- **Cancer Biopsy:** Built a CNN classifiers to diagnose breast cancer based on HE stained histhopathological biopsy slide images. Furthermore developed Class Activation Maps (CAMS) for these classifiers to gain insights. **Code**
- **Fake News Detection:** Features such as tf-idf, punctuation count, readability and PCFG were extracted from the LIAR dataset. The performance of various linear classifiers such as Naive Bayes, Logistic Regression and SVM were compared for the task. Done as the Information Retrieval course project. **Code**

Technical Skills

Languages: Java, Python, C/C++, R, Matlab
Deep Learning tools: Tensorflow, Keras, Pytorch
Bioinformatics tools : RDkit, OpenBabel, Biopython, Prody, Pymol
Utility Tools: Git, Linux, MS office, Photoshop