manas.mahale@bcp.edu.in | manasmahale.xyz

Manas Moreshwar Mahale

SKILLS

Computational Chemistry | Python | Machine Learning | Metabolomics | Git | JavaScript | Bash | AWS

EXPERIENCE

Pangea Botanica, Berlin — *AI & ML Intern*

Sept 2022 - June 2023

- Established codebase for Computational Metabolomics module on AWS.
- Improved performance of Mass Spectrometry Data based Structural Similarity Prediction Models for Natural Products using Siamese Neural Network.
- Developed state-of-the-art Computational Prediction Models for PK endpoints and Ligand-Target Prediction.

ChemBio Discovery, Boston — *Computational Chemistry Consultant*

June 2022 - July 2022

- Created Schrödinger workflow automation for computational validation of Covalent Docking Pipeline.
- Automated protein alignment and homology modeling.

The Coutinho Lab, BCP — *Undergraduate Researcher*

Sept 2021 - Present

- Worked on Chemical language models for De Novo & Fragment Based Drug Design.
- Developed and Integrated frequently used computational chemistry workflows.
- Worked on Reasonable Chemical Data Split Algorithm, Memory Efficient Docking and Accelerated MD simulations.

EDUCATION

Bombay College of Pharmacy, University of Mumbai — *B. Pharmacy*

July 2020 - 2025, Mumbai

8.0 CGPA

Publications

- Martis, E. A. F., **Mahale, M.**, Choudhary, A., & Coutinho, E. C. (2023). Chapter 7 — Understanding protein-ligand interactions using state-of-the-art computer simulation methods. In K. Roy (Ed.), Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development (pp. 181-203). Academic Press. https://doi.org/10.1016/B978-0-443-18638-7.00015-3

- Gupta, N., Mohan, C. D., Shanmugam, M. K., Jung, Y. Y., Chinnathambi, A., Alharbi, S. A., Ashrafizadeh, M., **Mahale, M.**, Bender, A., Kumar, A. P., Putti, T. C., Rangappa, K. S., Zhang, X., Ahn, K. S., & Sethi, G. (2023). CXCR4 expression is elevated in TNBC patient derived samples and Z-guggulsterone abrogates tumor progression by targeting CXCL12/CXCR4 signaling axis in preclinical breast cancer model. Environmental Research, 232, 116335. https://doi.org/10.1016/j.envres.2023.116335

Projects

Causal Language Modeling objective for Single Step Retrosynthesis

- Trained a GPT-2 model on SMIRKS based Reaction dataset.
- Conducted a comparative study on CLM objectives with Masked LMs.

Scaffold Generative Pretraining

- De Novo Scaffold Generation model trained on concatenated scaffold dataset created from MOSES dataset.
- Constrained yet diverse scaffold generation using context and few shot learning.
- Fine-tuned the model on Privileged Medicinal Chemistry structures for Target specific Molecular Library generation.

FragmentBERT : Fragment Based Drug Discovery using Masked Language Models

- BERT and RoBERTa based CLMs trained on Substructure of the compound as SMILES rather than tokenized SMILES characters.
- Benchmarked against SMILES tokenizer based Molecular Transformer, Byte Pair Encoding and CharFormer.
- Massively parallelized the output generation process using Multithreading.

PyEVANS : Python implementation of EigenValue ANalySis (EVANS)

- End-to-end implementation of the paper into python.
- Model pipeline delivered the feature matrix from structural features and rules defined in the paper. - Benchmarked the downstream QSAR/QSPR performance to the SOTA models.

References

Dr. Andreas Bender, CTIO, Pangea Botanica Ltd.
Prof. Evans Coutinho, Professor of Pharmaceutical Chemistry, BCP, University of Mumbai.