Amitesh Badkul

@ Email | LinkedIn | GitHub | Website

Education

Graduate Center, City University of New York — New York City, USA
PhD in Computer Science — *August 2024 - Present*

Birla Institute of Technology and Science, Pilani — Hyderabad, India
Master of Science in Chemistry — *August 2018 - June 2023*
Bachelor of Engineering in Electrical and Electronics Engineering

Publications

- ***Amitesh Badkul***, Li Xie, Shuo Zhang, and Lei Xie. (2024). "eMOSAIC: Multi-modal Out-of-distribution Uncertainty Quantification Streamlines Large-scale Polypharmacology", **Nature Machine Intelligence** [Status: Under Revision]
- ***Amitesh Badkul***, Vamsi Inturi, Sudha Radhika. (2024), **Journal of Medical Engineering & Technology**, "Comparative study of DCNN and image processing based classification of chest X-rays for identification of COVID-19 patients using fine-tuning". [DOI: 10.1080/03091902.2024.2438158]
- ***Amitesh Badkul***, Sonakshi Mishra, Srinivas Prasad K. (2024)., "A Comparative Study of DeepLabCut and Other Open-Source Pupillometry Data Analysis Algorithms – Which to Choose?", **Machine Graphics and Vision** [Status: Accepted, Under Production]
- ***Amitesh Badkul***, Li Xie, Shuo Zhang, and Lei Xie. (2023). "TrustAffinity: accurate, reliable and scalable out-of-distribution protein-ligand binding affinity prediction using trustworthy deep learning", [NeurIPS 2023 Workshop] &[AAAI 2024 Workshop]
- Tian Cai, Li Xie, Shuo Zhang, Muge Chen, Di He, ***Amitesh Badkul***., ... and Lei Xie. (2023)., "End-to-end Sequence-Structure-Function Meta-learning Predicts Genome-Wide Chemical-Protein Interactions for Dark Proteins", **PLoS Computational Biology** [DOI: 10.1371/journal.pcbi.1010851]

Experience

Research Foundation of City University of New York — New York City, USA
*Graduate Research Assistant (Supervisor: **Dr. Lei Xie**)* — *July 2024 – Present*

- Developing mixture-of-experts model for tackling highly data imbalance in PTM Classification.

Weill Cornell Medicial College — New York City, USA
*Visiting Scholar (Supervisor: **Dr. Lei Xie**)* — *Aug 2022 – March 2024*

- Developed a deep learning framework for predicting protein-ligand binding affinity, incorporating uncertainty. Achieved a Pearson correlation of **0.92** and MAE of **0.25** in OOD settings, surpassing state-of-the-art methods.
- Engineered and Integrated Multi-Task learning-based algorithms with Protein Language Models (DISAE and ESM-2) for enhanced chemical protein binding affinity prediction. Utilized deep learning techniques such as Transformers, RNNs, and CNNs, resulting in a Pearson correlation of **0.81** and MAE of **0.56**.
- Successfully implemented and fine-tuned a residue-residue contact classification model with a DISAE and MLP achieving an accuracy of **98.81%**.

Arizona State University — Tempe, USA
*Summer Research Intern (Supervisor: **Dr. Ashif Iquebal**)* — *Jun 2022 – Aug 2022*

- Analyzed large chemical datasets to identify the optimal dataset emphasizing compounds with hydrogen bonding.
- Optimized a generative model using LSTM and GRU architectures for self-healing compound generation.
- Trained a generative model yielding **98.43%** valid compounds, with a **higher ease** of synthesis on average. [Poster]

Birla Institute of Technology and Science Pilani — Hyderabad, India
*Undergraduate Research Assistant (Supervisor: **Dr. Srinivas Prasad K**)* — *Jun 2021 – Jun 2022*

- Employed architectures such as MobileNet, ResNet, and EfficientNet to achieve accurate rat pupil measurements.
- Achieved minimal deviation from ground truth, outperforming other state-of-the-art image processing algorithms.

*Undergraduate Research Assistant (Supervisor: **Dr. Durba Roy**)* — *Aug 2021 - Feb 2022*

- Simulated a water box cube using Molecular Dynamics for 20 nsec and analyzed positional and energy data.

- Devised algorithms for Mean Square Displacement (MSD) and Diffusion Coefficient in water systems; investigated Rhodobacter Sphaeroides' Reaction Center. [GitHub] [Blog 1] [Blog 2].

Undergraduate Research Assistant (Supervisor: ***Dr. Sudha Radhika****)* *Feb 2021 - April 2022*

- Optimized pretrained models (ResNet, MobileNet, etc) for CXR classification, achieving **97%** accuracy. Enhanced accuracy by **2%** with a CXR enhancement algorithm.
- Developed a novel dataset from CXR using statistical descriptors post wavelet transform. Achieved **97.46%** accuracy with XGBoost and Random Forest models. [GitHub]

Sensordrops Networks, IIT Kharagpur Kharagpur, India

Research Intern (Supervisor: ***Dr. Sudip Misra****)* *Dec 2020 – Feb 2021*

- Developed a Graph Neural Networks (GNNs) based algorithm for Contact Tracing of COVID-19 patients.
- Created a novel Twitter dataset for training and testing. Used Twitter metadata as features and deployed the GNN model. Obtained accuracy of **92.31%**.

Million Sparks Foundation Noida, India

Summer Intern *Jun 2020 – Aug 2020*

- Refactored and refined of existing JavaScript code leading to increased efficiency and usability.
- Developed and enhanced educational content, **benefiting over 20+ teachers** and elevating learning outcomes.

Projects

CYP3A4 Inhibition Classification | *GitHub*

- **Curated** and **cleaned** the datasets for improved accuracy of machine learning models.
- **Deployed** machine learning algorithms such as Random Forest, and XGBoost. **Achieved** an accuracy of 77%.

CXR Bit Plane Classification | *GitHub*

- **Implemented** CXR classification on bit plane sliced CXRs using MobileNet model.
- **Obtained** a highest accuracy of 95% validating the fact that all layers of a CXR are equally important.

Cdh23EC1 Analysis | *Blog*

- **Conducted** thorough and **extensive** analysis of the data obtained from the MD Simulation of Cdh23EC1 protein.
- **Programmed** functionalities for calculation of various properties - MSD, Radius of Gyration and more.

Skills

- **Programming Languages:** Python, MATLAB, R, Verilog, LaTeX, C, Bash, Java, Javascript, HTML, CSS
- **Software Skills:** BLAST, NAMD, VMD, OpenCV, EMU8086, LTSpice, Microsoft Office Suite, Adobe Suite, AutoCAD
- **Languages:** Hindi (Native), English (Professional)
- **Libraries:** PyTorch, GPyTorch, PyTorch Geometric, TensorFlow, Keras, OpenCV, RDkit, Scikit-Learn, Biopython, Numpy, Pandas, Matplotlib, Seaborn, DeepLabCut, Networkx, DGL, Pywt, Deepchem

Service

- **Reviewer**, International Conference on Learning Representations (ICLR) *2024*
- **Reviewer**, International Journal of Computational Biology and Drug Design *2023*
- **Student Member**, IEEE Organization, BITS-Pilani, Hyderabad *2019-2022*

Awards & Achievements

- **Scholarship for Practice School - 1**: Industry exposure program, given to the students with excellent performance (highest grade - 'A') in the industry provided by BITS Pilani. *2020*
- **Scholarship for Higher Education (SHE):** Recipient of Scholarship for Higher Education provided by the Indian Government for excellence in academics Higher Secondary School Board examination, given to the top 1% of students, held in month of March 2018 in India. *2018-2022*