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

SELECTED PUBLICATIONS

- **Badkul, A.**, Xie, L. Adaptive Individual Uncertainty under Out-Of-Distribution Shift with Expert-Routed Conformal Prediction. EuRIPS Workshop Epistemic Intelligence in Machine Learning. *2025*
- **Badkul, A.**, Xie, L., Zhang, S., *et al.* eMOSAIC: Multi-modal Out-of-distribution Uncertainty Quantification Streamlines Large-scale Polypharmacology. Nature Machine Intelligence. Status: Accepted In Principle. *2025*
- **Badkul, A.**, Xie, L., Zhang, S., *et al.* TrustAffinity: accurate, reliable and scalable out-of-distribution protein-ligand binding affinity prediction using trustworthy deep learning. NeurIPS Workshop *2023*
- Cai, T., Xie, L., · · · ., **Badkul, A.**, *et al.* End-to-end Sequence-Structure-Function Meta-learning Predicts Genome-Wide Chemical-Protein Interactions for Dark Proteins. PLoS Computational Biology. *2023*

EXPERIENCE

Research Foundation of City University of New York — New York City, USA
*Graduate Research Assistant (Supervisor: **Dr. Lei Xie**)* — *July 2024 − Present*
- Built a novel MoE-based UQ system, achieving near-nominal coverage with tight, adaptive, informative intervals.
- Developed a three-stage pharmacokinetics-based end-to-end DL framework for drug concentration over time by combining a chemical foundation model, multi-task PK parameter learning, and a physics-informed Neural ODE for concentration delivering stable trajectories across compounds.
- Created DrugPTM-Bench, a large-scale, dose- & time-resolved dataset for drug-induced, cell type–specific PTM regulation with extreme class imbalance, providing baselines, and benchmarked imbalance-robust algorithms on the multi-omic data.

Weill Cornell Medical College — New York City, USA
*Visiting Researcher (Supervisor: **Dr. Lei Xie**)* — *Aug 2022 − March 2024*
- Developed eMOSAIC, a multi-modal framework to predict binding affinity with built-in uncertainty quantification to support polypharmacology and scalable virtual screening, with improved out-of-distribution (OOD) generalization.
- Reducing OOD error MAE by 23%, and performs virtual screening faster in 0.01s with $8\times$ enrichment factor.

Arizona State University — Tempe, USA
*Summer Research Intern (Supervisor: **Dr. Ashif Iquebal**)* — *Jun 2022 − Aug 2022*
- Curated a hydrogen-bond–focused chemical dataset and optimized LSTM/GRU generative models for self-healing compounds, achieving 98.43% valid molecules with improved ease of synthesis and 93.29% novelty.

Birla Institute of Technology and Science Pilani — Hyderabad, India
Undergraduate Research Assistant — *Feb 2021 − Jun 2022*
- Benchmarked open-source pupillometry, and DL-based methods, identifying lightweight MobileNet models as most compute-efficient while matching ground truth with error $\leq 2\%$ for accurate rat-pupil tracking.
- Simulated a water box cube using Molecular Dynamics for 20 nsec and analyzed positional and energy data.
- Built a wavelet-derived feature COVID-19 CXR pipeline that exceeded a DCNN baseline, while requiring orders of the 10^5–10^6x less compute and delivered a 2% gain in performance.

Sensordrops Networks, IIT Kharagpur — Kharagpur, India
*Research Intern (Supervisor: **Dr. Sudip Misra**)* — *Dec 2020 − Feb 2021*
- Built and deployed a GNN-based COVID-19 contact-tracing model on Twitter metadata, achieving 92.31% accuracy.

PROJECTS

- **Exploring Perturbation-Induced Morphological Shifts** — designed a CLIP-based drug to cell morphology pipeline achieving retrieval R@5 = 84.5%, R@10 = 96.5% ⭘
- **Parameter-Efficient Fine-Tuning (PEFT) for NLP** — implemented and benchmarked LoRA, Prefix, and BitFit approaches, matching full fine-tuning performance with <1% trainable parameters and reduced compute ⭘

SKILLS

- **Programming Languages:** Python, C++, SQL, Bash, Java, JAX
- **ML/DL Libraries:** PyTorch, PyTorch Geometric, DGL, TensorFlow, scikit-learn, NumPy, Pandas, Matplotlib
- **Frameworks & Tools:** Docker, Kubernetes, Git, Slurm
- **MLOps & Serving:** CI/CD, MLflow/Model Registry, Multi-GPU and distributed training, Monitoring (WandB)
- **Data & Cloud Platforms:** AWS (EC2/S3), SageMaker, GCP, Apache Spark, Hadoop, Jupyter