

Mahbuba Tasmin

COMPUTATIONAL BIOLOGY, BIOLOGICAL FOUNDATION MODELS, AND PREDICTIVE BIOLOGY

University of Massachusetts Amherst · mtasmin@umass.edu · 413-479-9565

[Website](#) · [GitHub](#) · [LinkedIn](#) · [Google Scholar](#)

Profile

Ph.D. candidate developing interpretable and robust machine-learning methods for biological sequences, mutation-effect prediction, and predictive biology. Experienced in biological foundation-model adaptation, protein and genomic sequence modeling, benchmark development, model evaluation, multimodal biological data integration, and GPU/HPC pipelines. **Dissertation proposal planned for early Fall 2026. Available for full-time research roles with flexible start dates throughout 2027.**

Research Interests

Interpretable machine learning for biological sequences; protein foundation models; antimicrobial resistance; mutation-effect prediction; structure-aware modeling; multimodal biological data; robust evaluation under data scarcity.

Education

Sep 2022 – Present	University of Massachusetts Amherst Ph.D. Candidate in Computer Science, advised by Prof. Anna G. Green. GPA: 3.9/4.0, Research Concentration: Computational Biology, Machine Learning. Dissertation title: <i>"Integrating Biological Signal to Improve Generalizability and Interpretability of Machine Learning in Genomics"</i>	Amherst, MA
Sep 2022 – May 2025	University of Massachusetts Amherst M.S. in Computer Science, GPA: 3.9/4.0. Research aligned with predictive and interpretable modeling of antibiotic resistance.	Amherst, MA
Jan 2016 – Dec 2019	North South University B.S. in Computer Science and Engineering, <i>Summa Cum Laude</i> , GPA: 3.89/4.0. Concentration in Artificial Intelligence and Algorithms.	Dhaka, Bangladesh

Employment and Academic Appointments

Sep 2023 – Present	Graduate Research Assistant, Sequence Analysis and Genomics (SAGE) Lab Lead and co-lead research in interpretable biological sequence modeling, antibiotic resistance forecasting, protein foundation-model benchmarking, multimodal biological data integration, and evolutionary augmentation.	University of Massachusetts Amherst
Jan 2023 – May 2026	Head Teaching Assistant / Teaching Assistant, COMPSCI 520 Theory and Practice of Software Engineering. Served four semesters as Head TA and additional semesters as TA; supported classes of more than 140 students, designed instructional materials, coordinated teaching staff, oversaw grading, and maintained GitHub Classroom and Gradescope infrastructure.	University of Massachusetts Amherst
May 2023 - Aug 2023	Course Developer, COMPSCI 520 Redesigned assignments and lab materials, established GitHub Classroom workflows, and developed automated grading infrastructure.	University of Massachusetts Amherst
Mar 2022 – Jul 2022	AI Engineer Developed Detectron2-based instance-segmentation, OCR, and image-processing pipelines, along with NLP/CV tools for automated product identification and fashion-trend analysis.	NITEX Solutions Ltd., Dhaka, Bangladesh
Jul 2020 – Feb 2022	Software Engineer, AI & ML Developed image-spoofing detection, contextual recommendation systems, workflow automation with Camunda, and production ML deployments in distributed environments.	M2SYS Technology, Dhaka, Bangladesh

Selected Publications

Preprints and Manuscripts Under Review

1. **Tasmin, M.**, Barethiya, S., Wang, Y., Kang, L., Chen, J., and Green, A. G. "FARM: Forecasting Antibiotic Resistance in *Mycobacterium tuberculosis* Using Biophysics and Machine Learning." *bioRxiv*, 2026. [BIORxiv](#). Manuscript under review at

Proceedings of National Academy of Sciences (PNAS).

2. **Tasmin, M.**, Mohanty, S., Kulkarni, S., Farhat, M. R., and Green, A. G. “BIG-TB: A Benchmark for Prediction and Interpretability of Sequence-Based Machine Learning Using *Mycobacterium tuberculosis* Genomes.” *bioRxiv*, 2026. [BIORxiv](#). Manuscript submitted; under revision following peer review.

Peer-Reviewed Journal Articles & Book Chapters

1. Green, A. G., **Tasmin, M.**, Vargas Jr., R., and Farhat, M. R. “The structural context of mutations in proteins predicts their effect on antibiotic resistance.” *eLife*, 14:RP109450, 2025. [DOI](#).

2. Yang, Z., Yao, Z., **Tasmin, M.**, Vashisht, P., Jang, W. S., Ouyang, F., Wang, B., McManus, D., Berlowitz, D., and Yu, H. “Unveiling GPT-4V’s hidden challenges behind high accuracy on USMLE questions: Observational Study.” *Journal of Medical Internet Research*, 27:e65146, 2025. [DOI](#).

3. **Tasmin, M.**, et al. “Assessment of Deep Learning Models for Human Activity Recognition on Multivariate Time Series Data and Non-targeted Adversarial Attack.” In *Advances in Intelligent Systems Research and Innovation*, Studies in Systems, Decision and Control, vol. 379, Springer, 2022. [DOI](#).

4. **Tasmin, M.**, Nag, P., Hoque, Z. T., et al. “Non-Newtonian effect on heat transfer and entropy generation of natural convection nanofluid flow inside a vertical wavy porous cavity.” *SN Applied Sciences*, 3:299, 2021. [DOI](#).

Peer-Reviewed Conference and Workshop Papers

5. **Tasmin, M.** and Green, A. “Beyond Sequence-only Models: Leveraging Structural Constraints for Antibiotic Resistance Prediction in Sparse Genomic Datasets.” *ICLR 2025 MLGenX Workshop*, 2025. [OPENREVIEW](#).

6. **Tasmin, M.**, et al. “Comparative Study of Classifiers on Human Activity Recognition by Different Feature Engineering Techniques.” *IEEE 10th International Conference on Intelligent Systems*, pp. 93–101, 2020. [DOI](#).

Extended Abstracts

7. **Tasmin, M.**, “Multi-Dimensional Aspect Analysis of Text Input through Human Emotion and Social Factors.” *ACM International Joint Conference and Symposium on Pervasive and Ubiquitous Computing (UbiComp)*, pp. 1779–1781, 2018. [DOI](#)

Technical Skills

Programming & Tools	Python, Bash, PyTorch, Scikit-learn, Pandas, Java, LaTeX
Machine Learning	Deep learning, regression and classification, high-dimensional modeling, model evaluation and experimental design, baseline selection, cross-validation, ablation studies
Foundation Models	Protein and genomic language models, ESM, large pretrained model adaptation, parameter-efficient tuning (LoRA), fine-tuning, zero-shot evaluation, model probing, embedding engineering, PCA-based representation compression
Computational Biology	Biological sequence modeling, functional genomics, protein structure analysis, sequence alignment, variant annotation, DNA-to-protein translation, proteomics
Interpretability & Reliability	SHAP, feature attribution, causal-variant recovery, model diagnostics, robustness evaluation, error analysis
Computing Infrastructure	Linux, Docker, SLURM/HPC, GPU computing, distributed multi-GPU training, memory-mapped datasets, reproducible pipelines
Scientific Communication	Scientific writing, manuscript development, benchmark and dataset release, research presentations, interdisciplinary collaboration

Selected Predictive Biology Research Contributions

2024–Present	BIG-TB Led benchmark design and protein-model analyses for approximately 17,000 <i>Mycobacterium tuberculosis</i> isolates across 11 antibiotics, including defensible train/test splits, classical and deep-learning baselines, interpretability evaluation, causal-variant recovery, and robustness analysis.	UMass Amherst
2024–Present	Biophysics-Aware Resistance Forecasting Developed interpretable models integrating protein structure, Rosetta energetics, evolutionary features, and protein-language-model signals to predict mutation-level antibiotic resistance under data scarcity.	UMass Amherst

2025–Present	Evolutionary Protein Data Augmentation Designed and evaluated homolog-aware and biologically constrained augmentation strategies for mutation-sensitive prediction, including pretrained protein language-model scoring, transfer-learning experiments, and controlled ablations across sparse biological datasets.	UMass Amherst
---------------------	---	---------------

Software and Data Resources

2024–Present	BIG-TB Benchmark Open-source pipelines for preprocessing, training, cross-validation, robustness analysis, and interpretability evaluation across classical ML, deep neural networks, and biological foundation-model representations.	GitHub
2023–Present	Structure-Aware Variant Analysis Toolkit Research code for mapping mutations to protein structures, quantifying spatial clustering, and supporting structure-informed prediction and mechanistic interpretation.	GitHub
2024–Present	Resistance Forecasting in <i>Mycobacterium tuberculosis</i> Reproducible data-preparation, multimodal feature-engineering, model-training, and temporal-evaluation workflows for prioritizing antibiotic-resistance variants of uncertain significance.	GitHub

Talks and Posters

- **MLCB 2025**, Machine Learning for Computational Biology: spotlight/lightning talk on BIG-TB benchmark design and interpretability.
- **Harvard Program in Quantitative Genomics Conference 2024**: poster, “Protein Structure-Informed Regularized Linear Model Outperforms ESM for Predicting Antibiotic Resistance in *Mycobacterium tuberculosis*.”

Research Funding

- Graduate researcher supported by a UMass interdisciplinary research award for multimodal antibiotic-resistance forecasting, 2024. Principal Investigator: Anna G. Green.

Honors, Awards, and Sponsored Participation

- Selected as a fully funded graduate participant, Tapia Conference, September 2026.
- Fully funded selected participant, CRA-WP Grad Cohort for Women, 2023.
- Sudha and Rajesh Jha Scholarship, UMass Amherst, 2023.
- Fully funded Student Volunteer and ACM Student Travel Grant recipient, ACM UbiComp, Singapore, 2018, in connection with an accepted paper.
- Merit-Based Financial Aid, North South University, 2017–2019.

Public Engagement and Research Outreach

- **Featured Graduate Researcher, AI at UMass public-engagement campaign**, University of Massachusetts Amherst, 2026. Appeared in a university-produced outreach video explaining how computational biology and machine learning can support antibiotic-resistance prediction and the development of more effective disease treatments. *Watch video | AI at UMass campaign*

Research Mentoring

- **Shakir Sahibul**, M.S. student, Fall 2024: mentored a transformer-based antibiotic-resistance prediction project; subsequently joined Amazon.
- **Shuqi Hong**, M.S. student, Fall 2025–Summer 2026: mentored synthetic and evolutionary augmentation research for protein resistance modeling, resulting in a co-authored full-length manuscript submitted to MLCB 2026.

Professional and University Service

Peer Review

- Reviewer, *Bioinformatics Advances*, 2026.
- Reviewer, ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies Posters Track (COMPASS 2026).
- Reviewer, MLCSB: Machine Learning in Computational and Systems Biology at ISMB, 2026.
- Reviewer, ACM CHI Late-Breaking Work, 2023 and 2025.
- Reviewer, Machine Learning In Computational Biology 2026.

Academic and Community Leadership

- Elected Graduate Representative for Ph.D. students, Faculty Senate / CICS, UMass Amherst, 2025–2026; represented graduate concerns in faculty meetings and participated in faculty-candidate evaluation.
- Chapter Chair, North South University ACM Student Chapter, 2017–2019; led technical competitions, workshops, research training, and outreach, including ACM SIGCHI schools and an inter-university girls' programming contest.