Sumit Tarafder

sumittarafder@vt.edu sumittarafder.github.io github.com/SumitTarafder

Research Interests

Topics	3D structure prediction, refinement or quality estimation of protein, RNA and associated complexes
Areas	Structural bioinformatics, deep learning

Education

Virginia Tech **Aug. 2022 – Present**
Ph.D. student in Computer Science *Blacksburg, VA*

Bangladesh University of Engineering & Technology (BUET) **Jan. 2022**
M.Sc. in Computer Science and Engineering *Dhaka, Bangladesh*

Bangladesh University of Engineering & Technology (BUET) **Feb. 2017**
B.Sc. in Computer Science and Engineering *Dhaka, Bangladesh*

Experience

Virginia Tech **Aug. 2022 – Present**
***Graduate Research Assistant**, Bhattacharya Lab* *Blacksburg, VA*

- Focused on projects related to the prediction and quality estimation of RNA 3D structures and protein-RNA complexes.
- Recently implemented research projects are available in *GitHub*.

Virginia Tech **Aug. 2025 – Spring. 2026**
Graduate Teaching Assistant *Blacksburg, VA*

- **CS 2114 Software Design & Data Structures (Spring 2026)**
  - Responsible for conducting lab sessions, grading & project evaluations
- **CS 5805 Advanced Machine Learning (Fall 2025)**
  - Responsible for assignment design & grading, project proposal & report evaluation (in-class sessions).
- **CS 5914 TS: Warehouse Scale Computing (Fall 2022)**
  - Graded multiple paper review and summary writings for individual students and conducted office hours.

United International University **May 2017 – Oct. 2021**
***Lecturer**, Department of Computer Science and Engineering* *Dhaka, Bangladesh*

- I was the instructor and coordinator for a wide range of undergraduate-level theory and practical courses.

Projects

***RNAbpFlow*: Flow matching for RNA 3D structure prediction** *September 2024 - February 2025*
Tools: PyTorch Lightning, Flow matching algorithm []

- Developed SE(3)-equivariant generative model for RNA 3D structure sampling
- Nucleobase center representation for all-atom RNA modeling
- Applied flow matching for efficient sample generation
- Sequence- and base-pair-conditioned model without the use of evolutionary information.

***lociPARSE*: RNA 3D structure quality estimation** *July 2023- April 2024*
Tools: PyTorch, IPA, OpenStructure []

- Developed lociPARSE, a locality-aware RNA 3D scoring method
- Implemented invariant point attention for accurate structure assessment
- Curated a large-scale training set to train and evaluate the method
- Created lDDT-based scoring to evaluate RNA structures without superposition

***Figbird*: Gap filling in genome assembly** *April 2020- September 2021*
Tools: Expectation-Maximization (EM) algorithm, Bowtie2 []

- Developed Figbird, a C++ tool for gap filling in genome assemblies
- Implemented probabilistic modeling instead of graph-based approaches
- Applied partially aligned and unmapped Illumina read pairs to maximize sequence usage
- Utilized insert-size-aware read distribution for gap filling

Publications

- **Sumit Tarafder**, Debswapna Bhattacharya. "*RNAbpFlow: Base pair-augmented SE(3)-flow matching for conditional RNA 3D structure generation*", 2026. DOI: 10.1101/2025.01.24.634669, Nature Methods [In press], IF: 32.1
- **Sumit Tarafder**, Debswapna Bhattacharya. "*PARSEbp: Pairwise Agreement-based RNA Scoring with Emphasis on Base Pairings*", 2026. DOI: 10.1101/2025.10.13.682106, Bioinformatics Advances [Accepted], IF: 2.8
- **Sumit Tarafder**, Debswapna Bhattacharya. "*lociPARSE: a locality-aware invariant point attention model for scoring RNA 3D structures*", Journal of Chemical Information and Modeling, Volume 64, Issue 22, Pages 8655–8664, November 2024. DOI: 10.1021/acs.jcim.4c01621, IF: 5.7
- **Sumit Tarafder**, Rahmatullah Roche, Debswapna Bhattacharya. "*The landscape of RNA 3D structure modeling with transformer networks*", Biology Methods and Protocols, Volume 9, Page 47, Issue 1, July 2024. DOI: 10.1093/biomethods/bpae047, IF: 2.8
- **Sumit Tarafder**, Xinyu Wang, Rahmatullah Roche, Debswapna Bhattacharya. "*Advances in Language-Model-Informed Protein–Nucleic Acid Binding Site Prediction*", Methods in Molecular Biology, Page 139-151, July 2025. IF: 1.3
- **Sumit Tarafder**, Mazharul Islam, Swakkhar Shatabda, Atif Rahman. "*Figbird: A probabilistic method for filling gaps in genome assemblies*", Bioinformatics, Volume 38, August 2022, Pages 3717–3724, Issue 15. DOI: 10.1093/bioinformatics/btac404, IF: 5.4
- **Sumit Tarafder**, Md. Toukir Ahmed, Sumaiya Iqbal, Md Tamjidul Hoque, M. Sohel Rahman. "*RBSURFpred : Modeling Protein Accessible Surface Area in Real and Binary Space using Regularized and Optimized Regression*", Journal of Theoretical Biology, Volume 441, January 2018, Pages 44 - 57. DOI: 10.1016/j.jtbi.2017.12.029, IF: 2.0
- Rahmatullah Roche, **Sumit Tarafder**, Debswapna Bhattacharya. "*Single-sequence protein-RNA complex structure prediction by geometric attention-enabled pairing of biological language models*", Cell Systems, 2025. DOI: 10.1016/j.cels.2025.101400, IF: 7.7
- Rahmatullah Roche, Bernard Moussad, Md Hossain Shuvo, **Sumit Tarafder**, Debswapna Bhattacharya. "*EquiPNAS: improved protein-nucleic acid binding site prediction using protein-language-model-informed equivariant deep graph neural networks*", NAR, volume 52, Page e27, Issue 5, January 2024. DOI: 10.1093/nar/gkae039, IF: 13.1
- Rahmatullah Roche, **Sumit Tarafder**, Parkesh Suseendran, Xiaofeng Wang, Debswapna Bhattacharya. "*Protein-Protein Interaction Site Prediction via EquiPPIS and Its Application in Studying Viral Replication*," *Protein Evolution: Methods and Protocols*, pages 225–235, November 2025. DOI: 10.1007/978-1-0716-4828-5_14. IF: 1.3

Conference Talks/Presentations

- Presented our work "*lociPARSE*" as a *poster* at *ISMB '24* (Montreal, CA)
- Presented "*lociPARSE*" as a full-length talk at *GLBIO conference* (Minnestota, USA)

Awards

Pratt Fellowship	Department of Computer Science, Virginia Tech. Awarded in *2024* and *2025*.
Travel Grant	Department of Computer Science, Virginia Tech. Supported participation at *ISMB 2024*.
Travel Grant	ISCB and GLBC. Supported presentation at *GLBIO 2025*.
Dean's List Award	University Merit Scholarship and Dean's List from *(BUET)*.

Referee Activities

- Journal reviewer for IEEE Transactions on Computational Biology and Bioinformatics (x1)
- Journal sub-reviewer for Nature Protocols 2025 (x1)
- Conference sub-reviewer for ISMB 2026 (x1), ISMB 2025 (x1), IEE BIBM 2025 (x1), IEE BIBM 2024 (x1)
- Workshop sub-reviewer for BIOKDD 2024 (x1)

Skills

Languages: Python, Java, C, C++, HTML/CSS
Tools: Git, Docker, Chimera, PyMOL, VS Code
Frameworks: PyTorch, PyTorch Lightning, Matplotlib, Scikit-learn, LangChain

References

Dr. Debswapna Bhattacharya
Associate Professor, Department of Computer Science
Virginia Tech
Ph.D. Research Advisor

Dr. Atif Hasan Rahman
Associate Professor, Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
M.Sc. Research Advisor