

Sutanu Bhattacharya, Ph.D.

Assistant Professor
Department of Computer Science
Auburn University at Montgomery
USA

URL: <http://sutanubh1.github.io/>

Mobile: +1 334 444 5256
E-Mail: sbhatta4@aum.edu

Education

| | |
|---|-------------|
| Ph.D. in Computer Science and Software Engineering Auburn University, USA Advisor: Dr. Debswapna Bhattacharya (Now at Virginia Tech) Dissertation Title: New computational and data-driven methods for protein homology modeling | Summer 2021 |
| M.Tech in Distributed and Mobile Computing Jadavpur University, India | June 2014 |
| B.Tech in Information Technology Bengal Institute of Technology, India | June 2011 |

Employment

Assistant Professor, Department of Computer Science, Auburn University at Montgomery, USA, 16 August 2022 - ongoing.

Assistant Professor, Department of Computer Science, Florida Polytechnic University, USA, September 2021 – 15 August 2022.

Assistant Professor, Department of Computer Science and Engineering, Chaibasa Engineering College, India, July 2014 – Aug 2017.

Research Interest

My primary research interests are **computational biology** and applied machine learning with a particular focus on AI-powered molecular modeling. I am also interested in applications of data science in bioinformatics. My recent projects are focused on developing data-driven **machine learning** and **optimization algorithms** to elucidate complex relationships between macromolecular sequences, structures, functions, interactions, and pathways. Protein molecules are essential chemicals in our body. A protein is folded into a unique three-dimensional (3D) shape, which defines the functionalities of the protein. Determining the 3D structure of a protein is considered one of the hardest challenges in biochemistry in the last 50 years. However, the experimental determination of protein 3D structures is very time consuming and expensive, resulting in a very large gap (more than 1000 times) between the number of available protein sequences and the number of available protein 3D structures (aka templates) in the Protein Data Bank (PDB). Therefore, there is a need to develop computational approaches to predict the protein 3D structure from its amino-acid sequence, which is still an open problem. My primary research interest revolves around developing computational approaches to enlarge the scope of modellable proteins.

My secondary research interest is **Sentiment Analysis** for low-resourced languages using **deep learning**. Pretrained language models have been shown to perform well when there is a lot of data. This consequently leaves out a huge percentage of the world's languages as they are under-resourced. My research interest revolves around applying deep learning to train multilingual language models, which are viable on low-resource languages with small training data.

Honors and Awards

- (10) My group **Ranked 1** (out of 33 teams) in AfriSenti-SemEval: Sentiment Analysis for African Language (Task 12) at the 17th International Workshop on Semantic Evaluation, Canada, 2023.
- (9) **Best Poster** (3rd place) award at the 19th Annual MCBIOS Conference, Dallas, USA, 2023.
- (8) Work has been selected for **Highlight Talk** in the 12th ACM-BCB conference (Virtual), Aug 1-4, 2021.
- (7) **Fellowship award** for ISMB/ECCB 2021 conference (Virtual), July 25 – 30, 2021.
- (6) **Best Poster** (1st place) award in the 11th ACM-BCB conference (Virtual), Sept 21-24, 2020.
- (5) Work has been selected for **Highlight Talk** in the 10th ACM-BCB conference, Niagara Falls, NY, Sept 7-10, 2019.
- (4) Paper got published as the **Front Cover Article** in the July 2019 issue of Proteins: Structure, Function, and Bioinformatics journal.
- (3) **Young Research Excellence Award** (2nd place) at 16th Annual MCBIOS Conference, Birmingham, USA, 2019.
- (2) **Travel Award** to attend 16th Annual MCBIOS Conference, Birmingham, USA, 2019.
- (1) **National Scholarship** at Jadavpur University, India [2012 – 14].

Students' Achievements

- (4) Kevan Baker, my first graduate student, successfully defended his **MS thesis** at Florida Polytechnic University in Spring 2023. He has joined Auburn University as a PhD student.
- (3) Nathaneil Hughes, my undergraduate research assistant, received **Travel Award** to attend ACM TAPIA conference at Dallas, USA, 2023.
- (2) Kevan Baker received **Best Poster** (3rd place) at 19th Annual MCBIOS Conference, Dallas, USA, 2023.
- (1) Kevan Baker, my first graduate student, received **Travel Award** to attend 19th Annual MCBIOS Conference, Dallas, USA, 2023.

Peer-reviewed publications (+ undergraduate student, ++ graduate student)

In-preparation

(1) K. Baker++, N. Hughes+, **S. Bhattacharya**, “GoFold: a real-time protein folding program”, 2023. (Selected for an oral presentation at the 19th Annual MCBIOS Conference, Dallas, USA, March 15-17, 2023)

Book Chapter

(1) **S. Bhattacharya**, R. Roche, M. H. Shuvo, B. Moussad, D. Bhattacharya, “Contact-assisted threading in low-homology protein modeling”, *Methods in Molecular Biology*, by **Springer Nature**, 2023. (*Impact Factor: 10.71*)

Journals

(9) R. Roche, **S. Bhattacharya**, M. H. Shuvo, D. Bhattacharya, “rrQNet: protein contact map quality estimation by deep evolutionary reconciliation”, **Proteins: Structure, Function, and Bioinformatics**, <https://doi.org/10.1002/prot.26394>, (2022). (*Impact Factor: 3.756*)

(8) **S. Bhattacharya**, R. Roche, B. Moussad, D. Bhattacharya, “DisCovER: distance- and orientation-based covariational threading for weakly homologous proteins”, **Proteins: Structure, Function, and Bioinformatics**, (2021). [**First method to utilize inter-residue orientations to boost protein threading performance**], [**Fellowship Award for 29th ISMB/ECCB**]. (*Impact Factor: 3.756*)

(7) Kryshtafovych et al. (2021). Modeling SARS-CoV2 proteins in the CASP-commons experiment. **Proteins: Structure, Function, and Bioinformatics**, (2021). (*Impact Factor: 3.756*)

(6) **S. Bhattacharya**, R. Roche, M. H. Shuvo, D. Bhattacharya, “Recent advances in protein homology detection propelled by inter-residue interaction map threading”, **Frontiers in Molecular Biosciences**, 8, 377 (2021). (*Impact Factor: 4.620*)

(5) R. Roche, **S. Bhattacharya**, D. Bhattacharya, “Hybridized distance- and contact-based hierarchical structure modeling for folding soluble and membrane proteins”, **PLOS Computational Biology**, 17(2): e1008753, (2021). [**Highlight Talk**]. (*Impact Factor: 4.428*)

(4) A. McGehee, **S. Bhattacharya**, R. Roche, D. Bhattacharya, “PolyFold: An interactive visual simulator for distance-based protein folding”, **PLOS ONE**, 15(12): e0243331 (2020). [**Best Poster Award**]. (*Impact Factor: 2.740*)

(3) M. H. Shuvo, **S. Bhattacharya**, D. Bhattacharya, “QDeep: distance-based protein model quality estimation by residue-level ensemble error classifications using stacked deep residual neural networks”, **ISMB Proceedings, Bioinformatics**, 36(S1): i285-i291 (2020). (*Impact Factor: 5.610*)

(2) **S. Bhattacharya**, D. Bhattacharya, “Evaluating the significance of contact maps in low-homology protein modeling using contact-assisted threading”, **Nature Scientific Reports**, 10(1), 1-13 (2020). (*Impact Factor: 4.379*)

(1) **S. Bhattacharya**, D. Bhattacharya, “Does inclusion of residue-residue contact information boost protein threading?”, **Proteins: Structure, Function, and Bioinformatics**, 87(7): 596-606 (2019). [**Front Cover Article**], [**Highlight Talk**], [**Top Downloaded Paper** of 2018-2019 by WILEY]. (*Impact Factor: 3.756*)

Conference

(1) N. Raychawdhary, N. Hughes+, **S. Bhattacharya**, G. Dozier, C. Seals, “A Transformer-Based Language Model for Sentiment Classification and Cross-Linguistic Generalization: Empowering Low-Resource African Languages”, **IEEE AIBThings**, USA, 2023.

Workshop

(1) N. Hughes+, K. Baker++, Aditya Singh+, Aryavardhan Singh+, T. Dauda+, **S. Bhattacharya**, “Bhattacharya_Lab at SemEval-2023 task 12: A Transformer-based language model for sentiment classification for low-resource African languages: Nigerian Pidgin and Yoruba”, In Proceedings of the 17th International Workshop on Semantic Evaluation (**SemEval-2023**), pages 1502-1507, Toronto, Canada, July 2023. Association for Computational Linguistics. [**Ranked 1** out of 33 groups for Nigerian Pidgin and **Ranked 5** out of 33 groups for Yoruba]

Abstracts

(3) R. Roche, **S. Bhattacharya**, D. Bhattacharya, “Hybridized distance- and contact-based hierarchical structure modeling for folding soluble and membrane proteins”, **BCB '21**: Proceeding of the 12th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, USA, August 2021, Pages 1 (2021).

(2) **S. Bhattacharya**, D. Bhattacharya, “How Effective is Contact-assisted protein threading?”, **BCB'19**: Proceedings of the 10th **ACM** International Conference on Bioinformatics, Computational Biology and Health Informatics, September 2019, Pages 553 (2019).

(1) **S. Bhattacharya**, D. Bhattacharya, “Contact-assisted protein threading: an evolving new direction”, **BCB'19**: Proceedings of the 10th **ACM** International Conference on Bioinformatics, Computational Biology and Health Informatics, September 2019, Pages 536 (2019).

Conference Talks / Invited Talks

(7) “A real-time protein folding program with enhanced alignment capability”, at California State University, San Marcos, 2023.

(6) “GoFold: a real-time protein folding program”, 19th MCBIOS conference, Dallas, USA, 2023.

(5) “A distance- and orientation-based covariational protein threading”, at Florida Polytechnic University, USA, 2022.

(4) “DisCovER: distance- and orientation-based covariational threading for weakly homologous proteins”, 29th ISMB/ECCB conference, Virtual Event, July 25-30, 2021. (**Fellowship Award**)

(3) “How Effective is Contact-assisted Protein Threading?”, 10th ACM-BCB Conference, Niagara Falls, NY, Sept 7-10, 2019. (**Highlight Talk**)

(2) “Does contact information powered by deep learning boost protein threading?”, Student Symposium, Auburn University, USA, April 9, 2019.

(1) “Does inclusion of residue-residue contact information boost protein threading?”, 16th Annual MCBIOS Conference, Birmingham, USA, March 28- 30, 2019 (**Won 2nd place in student oral presentation**).

Posters

(7) K. Baker++, N. Hughes+, **S. Bhattacharya***, “GoFold: a real-time protein folding program”, 19th **MCBIOS** conference, Dallas, USA, 2023. (3rd prize in **Best Poster Award** category)

(6) N. Hughes+, K. Baker++, Aditya Singh+, Aryavardhan Singh+, T. Dauda+, **S. Bhattacharya***, “Bhattacharya_Lab at SemEval-2023 task 12: A Transformer-based language model for sentiment classification for low-resource African languages: Nigerian Pidgin and Yoruba”, 3rd **Annual Celebration of Undergraduate Research and Creative Activity**, Auburn University at Montgomery, 2023.

(5) **S. Bhattacharya**, R. Roche, D. Bhattacharya, “DisCovER: distance- and orientation-based covariational threading for weakly homologous proteins”, 29th **ISMB/ECCB** conference, Virtual Event, July 25-30, 2021. (**Fellowship Award**)

(4) A. McGehee, **S. Bhattacharya**, R. Roche, D. Bhattacharya, “PolyFold: An interactive visual simulator for distance-based protein folding”, 11th **ACM-BCB** Conference, Virtual Event, Sept 21-24, 2020. (**Best Poster Award**)

(3) **S. Bhattacharya**, D. Bhattacharya, “Contact-assisted protein threading: an evolving new direction”, 2019 Graduate Engineering Research Showcase, Auburn University, USA, November 7, 2019.

(2) **S. Bhattacharya**, D. Bhattacharya, “Contact-assisted protein threading: an evolving new direction”, 10th **ACM-BCB** Conference, Niagara Falls, NY, Sept 7-10, 2019. [[pdf of the poster](#)]

(1) **S. Bhattacharya**, D. Bhattacharya, “Does inclusion of residue-residue contact information boost protein threading?”, 16th Annual **MCBIOS** Conference, Birmingham, USA, March 28-30, 2019. [[pdf of the poster](#)]

Grants

(3) NSF proposal, titled “CAP: Cortex-Inspired Cross-Modal Deep Learning and Application to Bioimage Analysis”, with a budget of around \$400K, is currently under review (role: Co-PI).

(2) Institutional Grant: Received \$7,500 grant from Auburn University at Montgomery, 2023-25 (role: PI).

(1) Summer Research Grant of three months’ salary from Florida Polytechnic University, 2022 (role: PI).

Teaching, mentoring, and Advising Activities

Teaching Experience

Auburn University at Montgomery (Fall 2022 – now)

| Academic Term | Course | Enrollment |
|----------------------|--|-------------------|
| Fall 2023 | CSCI 6000 Data Structures and Algorithms | 142 |
| Fall 2023 | CSCI 3400 Data Structures | 11 |
| Fall 2023 | CSCI 3600 Fund Algorithm Design and Analysis | 14 |
| Summer 2023 | CSCI 6170 Advanced Network Systems (<i>Online</i>) | 74 |
| Spring 2023 | CSCI 6170 Advanced Network Systems | 33 |
| Spring 2023 | CSCI 3400 Data Structures | 13 |
| Spring 2023 | CSCI 3600 Fund Algorithm Design and Analysis | 17 |
| Fall 2022 | CSCI 6170 Advanced Network Systems | 100 |
| Fall 2022 | CSCI 3400 Data Structures | 17 |
| Fall 2022 | CSCI 3600 Fund Algorithm Design and Analysis | 18 |
| Fall 2022 | CSCI 2200 Discrete Structures | 11 |
| Fall 2022 | CSCI 4350 Network Systems (<i>Online</i>) | 8 |

Florida Polytechnic University (Fall 2021 – Summer 2022)

| Academic Term | Course | Enrollment |
|----------------------|---|-------------------|
| Spring 2022 | COP 4935 Senior Design II | 55 |
| Spring 2022 | COP 5531 Advanced Algorithm Design and Analysis | 8 |
| Fall 2021 | COP 4934 Senior Design I | 60 |

Chaibasa Engineering College, India (July 2014 – August 2017)

| Course | Enrollment |
|--------------------------------|-------------------|
| Discrete Mathematics | 45 |
| Data Structures | 45 |
| Automata | 45 |
| Introduction to Computing | 90 |
| Computer Organization | 45 |
| Algorithms Design and Analysis | 45 |

As a Graduate Teaching Assistant at Auburn University (Fall 2017 – Summer 2021)

COMP 1210 Fundamentals of Computing I
 CPSC 1223 Data Structures (*Online*)
 COMP 5970/6970 Computational Intelligence & Adversarial Machine Learning
 CPSC 1213 Introduction to Computer Science I (*Online*)
 CPSC 1223 Introduction to Computer Science II (*Online*)

MS Thesis Supervised

(1) “GoFold: a real-time protein folding program”, Kevan Baker, Florida polytechnic University, Department of Computer Science, Spring 2023. Now: Ph.D. student at Auburn

University, AL, USA.

Undergraduate Students Supervised

- (4) Nathaneil Hughes (BS – Spring 2024 expected)
- (3) Aditya Singh (BS – Spring 2025 expected)
- (2) Aryavardhan Singh (BS – Spring 2025 expected)
- (1) Tharalillah Dauda (BS – Fall 2023 expected)

Undergraduate Peer Mentor Supervised

- (3) Nathaneil Hughes, CSCI 3400 Data Structures, Fall 2023
- (2) Om Patel, CSCI 3400 Data Structures, Spring 2023
- (1) Riya Patel, CSCI 3600, Fund Algorithms Design and Analysis, Spring 2023

Services

University and Departmental Services

Auburn University at Montgomery

- 2023: Instructional Designer Search Committee for College of Sciences
- 2023: Poster Judge at Science Fair

Department of Computer Science, Auburn University at Montgomery

- 2022 - Present: Curriculum Committee
- 2022 - Present: Curriculum Quality Control Committee
- 2022 - Present: Faculty Search Committee
- 2022 - Present: Admission Committee

Florida Polytechnic University

- 2021 - 22: Undergraduate Curriculum Committee

Professional

Journal Reviewer

- Public Health Reviews (*Impact Factor of 5.5*)
- Bioinformatics (*Impact Factor of 6.9*)
- BMC Bioinformatics (*Impact Factor of 3.3*)
- Nature Scientific Reports (*Impact Factor of 4.6*)

Conference/Workshop Reviewer

- IEEE-BIBM
- SemEval Workshop

Outreach

- Head Judge at the GEARSEF (Greater East Alabama Regional Science and Engineering Fair) for K12 on Feb 24, 2023.
- Judge at the Alabama Science and Engineering Fair (ASEF) for K12 at Auburn University on April 1, 2023.
- Judge at the 19th MCBIOS conference, Dallas, 2023.