PRATIK DUTTA

Senior Research Scientist,
State University of New York, Stony Brook,
Stony Brook Cancer Center.

150 N Country Rd, Apt A2, Port Jefferson
New York 11777, USA
pratik.dutta@stonybrook.edu
Github *Google Scholar* *My Webpage*

Ongoing Research Grants and Projects

Awarded Grants

2025–2029 **Developing Novel Deep-Learning Based Methods for Deciphering Non-Coding Gene Regulatory Code**, *NIH/NLM (R01LM013722)*.
Role: Co-Investigator (Co-I)
Amount: $1,417,645
Period: August 2025 – June 2029
Principal Investigator: Prof. Ramana V. Davuluri
Funding Agency: National Library of Medicine (NLM)

Grants Under Review

2026–2027 **Genomic Foundation Models for Illuminating Microbial Dark Matter and Predicting Community Function**, *DOE (Genesis Phase I Proposal)*.
Role: Thrust Lead (Senior Personnel)
Lead PI: Prof. Ramana V. Davuluri
Amount: $750,000 (Total Phase I)
Status: Submitted
Funding Agency: Department of Energy (DOE)

2026–2029 **Regulome-R: An Aging-Aware Reasoning Engine for Mechanistic Interpretation of Non-Coding Variants in Neurodegeneration**, *R21 Trailblazer Proposal*, 1R21AG103597-01.
Role: Principal Investigator (PI)
Amount: $626,718
Status: Submitted / Pending Sponsor Review
Funding Agency: National Institutes of Health (NIH)

2026–2030 **Foundation models for functional annotation of short non-coding variants in personal genomes**, *NIH (Grant #: FP00015167)*.
Role: Co-Investigator (Co-I)
Amount: $2,691,876
Period: December 2026 – 2030
Principal Investigator: Prof. Ramana V. Davuluri
Status: Submitted / Pending Sponsor Review

2026–2028 **Deep-learning framework for accurate and interpretable non-coding short genetic variant effect prediction on the AnVIL platform**, *NIH (Grant #: FP00015369)*.
Role: Co-Investigator (Co-I)
Amount: $390,831
Period: December 2026 – 2028
Principal Investigator: Prof. Ramana V. Davuluri
Status: Submitted / Pending Sponsor Review

2026–2028 **Genomic foundation model for human virome**, *NIH (Grant #: FP00014910)*.
Role: Co-Investigator (Co-I)
Amount: $311,331
Period: July 2026 – 2028
Principal Investigator: Prof. Ramana V. Davuluri
Status: Submitted / Pending Sponsor Review

Academic Experience

Department of Biomedical Informatics, State University of New York, Stony Brook.

March 2023 –Present : **Senior Research Scientist**.
- Developed **HViLM (Human Virome Language Model)**, a DNABERT-2-based model for *viral pathogenicity and host-infectivity prediction*; Introduced the **HVUE benchmark** (NCBI & BV-BRC) to standardize virome model evaluation. Conducted **interpretability analysis** revealing molecular mimicry patterns—demonstrating how attention mechanisms can discover biological principles (e.g., viral immune evasion strategies) without explicit supervision.
- Developed **DeepVRegulome** by leveraging **DNABERT** to predict **small somatic mutations** impacting non-coding functional sites(**splice sites and 700 transcription factor binding sites (TFBS)**) in cancer genomes using whole-genome sequencing data. Developed **an interactive dashboard** (`https://davuluri-lab-brainved.streamlit.app/`) for real-time exploration.
- Improved pretraining techniques for foundational genomic language models like **DNABERT-2**, which leverages **flash attention** and **optimizes memory efficiency** for faster and larger sequence processing.

April 2021 –March 2023 : **Postdoctoral Research Associate**.
- **Developed DNABERT-Enhancer**, a fine-tuned genomic language model for enhancer identification; identify **candidate SNPs with potential loss-of-function effects** in enhancer regions, enabling functional validation and enhancer-based therapies.
- **Developed DeepMOIS-MC**, a multi-omics integration framework leveraging **Generalized Canonical Correlation Analysis (GCCA)** for outcome-guided molecular subtyping.
- Currently developing models integrating **genomic, epigenomic, and structural data** to enhance genomic foundation models.

Advisor : **Prof. Ramana V. Davuluri,** Department of Biomedical Informatics, Stony Brook Cancer Center, Stony Brook University.

Industrial Experience

Strand Life Sciences, Bangalore, India.

Aug 2020 – Mar 2021: **Research Intern**.
Developed a fine-tuned BioBERT model for phenotype extraction from clinical text, mapping free-text symptom descriptions to 15,000+ Human Phenotype Ontology (HPO) terms with hierarchical ranking for rare disease diagnosis

Advisors : **Dr. Vamsi Veeramachaneni,** *CSO*, Strand Life Sciences, India and **Dr. Rajesh Sundaresan**, *Professor*, Department of Electrical Communication Engineering, IISc, Bangalore.

Education

Jan 2016– Dec 2020 : **PhD(Visvesvaraya Research Fellow), Computer Science & Engineering**, *Indian Institute of Technology*, Patna.
Thesis Title: *Computational Approaches Leveraging Protein Interaction Information for Biomedical Tasks.*

Advisor : **Dr. Sriparna Saha**, *Associate Professor*, Computer Science & Engineering, IIT Patna.

2013–2015 : **Master of Engineering, Information Technology**, *Indian Institute of Engineering Science & Technology*, Shibpur(*Formerly* Bengal Engineering and Science University, Shibpur).

Advisor : **Dr. Hafizur Rahaman**, *Professor*, Department of Information Technology and School of VLSI, IIEST Shibpur.

2009–2013 : **Bachelor of Engineering, Computer Science & Technology**, *Indian Institute of Engineering Science & Technology*, Shibpur(*Formerly* Bengal Engineering and Science University, Shibpur).

Research Experience

Oak Ridge National Laboratory (ORNL), United States

November 2024 – Present

Interpretable AI Frameowrks for Spatial Transcriptomics.
Leading benchmarking and uncertainty quantification studies for deep learning models in spatial transcriptomics, integrating transcriptomic profiles with histological images for tumor microenvironment and disease biomarker analysis.

Collaborators: **Dr. Tirthankar Ghosal** *(Scientist, NLP/AI and HPC, National Center for Computational Sciences, Oak Ridge National Laboratory; Affiliated Faculty, University of Tennessee Knoxville)*

Broad Institute of MIT and Harvard

August 2020 – Jan 2021

Label-free Prediction of 3D Fluorescent Images.
Developing a deep Learning model to do label-free prediction of Cell Painting stains. More specifically, given the "brightfield" channel that is acquired together with the 5 Cell Painting channels, the task is to predict each of the 5 channels.

Advisors: **Dr. Shantanu Singh**(*Senior group leader, Imaging Platform*) and **Dr. Anne Carpenter** (*Institute Scientist, Senior Director of the Imaging Platform*), Broad Institute of MIT and Harvard

SystemOnSilicon Corporation(USA)

July 2019 – Dec 2020

Developing a Cloud-based Interactive AI Platform for Digital Health Analytics.
Designed & developed a cloud-based interactive AI platform for real-world applications in healthcare, agriculture, and personalized nutrition. I led AI and Data Science/Analytics R&D for creating novel ML/DL powered solutions for precision and smart agriculture, personalized food recommendation, and developing a conversational AI platform for wellness guidance .

Collaborators: **Raj Ray***(Founder and CEO*, SystemOnSilicon Corporation, USA) and **Dr. Dave Ray***(Co-founder and CSO*, SystemOnSilicon Corporation; Assistant Professor, Alcorn State University)

Publications

Preprints

2026 **Pratik Dutta**, Jack Vaska, Pallavi Surana, Rekha Sathian, Max Chao, Zhihan Zhou, Han Liu, and Ramana V Davuluri. Hvilm: A foundation model for viral genomics enables multi-task prediction of pathogenicity, transmissibility, and host tropism. *bioRxiv*. Cold Spring Harbor Laboratory, 2026.

2025 **Pratik Dutta**, Matthew Obusan, Rekha Sathian, Max Chao, Pallavi Surana, Nimisha Papineni, Yanrong Ji, Zhihan Zhou, Han Liu, Alisa Yurovsky, et al. Deepvregulome: Dnabert-based deep-learning framework for predicting the functional impact of short genomic variants on the human regulome. *arXiv preprint arXiv:2511.09026*, 2025.

2025 Pallavi Surana, **Pratik Dutta**, Nimisha Papineni, Rekha Sathian, Zhihan Zhou, Han Liu, and Ramana Davuluri. Tsprom: Deciphering the genomic context of tissue specificity. *bioRxiv*, pages 2025–10. Cold Spring Harbor Laboratory, 2025.

2025 Rekha Sathian, **Pratik Dutta**, Ferhat Ay, and Ramana V Davuluri. Genomic language model for predicting enhancers and their allele-specific activity in the human genome. *bioRxiv*, pages 2025–03. Cold Spring Harbor Laboratory, 2025.

Journal Articles(Accepted)

2024 Pallavi Surana, **Pratik Dutta**, and Ramana V Davuluri. *TransTEx: novel tissue-specificity scoring method for grouping human transcriptome into different expression groups*. ***Bioinformatics***, volume 40. Oxford Academic, 2024.

2023 Liwei Yang, **Pratik Dutta**, Ramana V Davuluri, and Jun Wang. *Rapid, High-Throughput Single-Cell Multiplex In Situ Tagging (MIST) Analysis of Immunological Disease with Machine Learning*. ***Analytical Chemistry***. ACS Publications, 2023, **(Impact Factor: 8.008)**.

2023 Yanrong Ji, **Pratik Dutta**, and Ramana Davuluri. *Deep Multi-Omics Integration by Learning Correlation-Maximizing Representation Identifies Prognostically Stratified Cancer Subtypes*. ***Bioinformatics Advances***, volume 3, page vbad075. Oxford University Press, 2023.

2022 Kanchan Jha, Sriparna Saha, and **Pratik Dutta**. *Incorporation of gene ontology in identification of protein interactions from biomedical corpus: a multi-modal approach*. ***Annals of Operations Research***, pages 1–19. Springer, 2022.

2022 Priyadarshini Dasgupta, Lisa Kuhn, Ephraim Massawe, Mason Williams, Julian Perrone, **Pratik Dutta**, and Debarshi Roy. *Impact of environmental and socio-economic stressors leading to unequal distribution of COVID-19 incidences in the state of Louisiana*. ***Environmental Quality Management***. Wiley Online Library, 2022.

2021 **Pratik Dutta**, Sriparna Saha, and Sukanya Naskar. *A multi-objective based PSO approach for inferring pathway activity utilizing protein interactions*. *Multimedia Tools and Applications*, volume 80, pages 30283–30303. Springer, 2021.

2021 **Pratik Dutta**, Aditya Prakash Patra, and Sriparna Saha. *DeePROG: Deep Attention-based Model for Diseased Gene Prognosis by Fusing Multi-omics Data*. ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***, pages 1–1. IEEE Computer Society, 2021, **(Impact Factor: 3.71)**, **(h5-index: 47)**.

2021 Sanjay Sarkar, Archie Taylor, **Pratik Dutta**, Meghna Bajaj, Justin Nash, Martha Ravola, Sofia Ievleva, Cardarius Llyod, Praise Ola, Brenita Jenkins, et al. *Health disparity and COVIID-19: A retrospective analysis*. ***Health Science Reports***, volume 4, page e345. Wiley Online Library, 2021.

2020 **Pratik Dutta**, Sriparna Saha, Sanket Pai, and Aviral Kumar. *A Protein Interaction Information-based Generative Model for Enhancing Gene Clustering*. ***Scientific Reports***, volume 10, pages 1–12. Nature Publishing Group, 2020, **(Impact Factor: 4.122)**, **(h5-index: 178)**.

2020 **Pratik Dutta**, Piyush Mishra, and Sriparna Saha. *Incomplete multi-view gene clustering with data regeneration using Shape Boltzmann Machine*. *Computers in Biology and Medicine*, page 103965. Elsevier, 2020, **(Impact Factor: 3.434)**.

2020 Swagarika Jaharlal Giri, **Pratik Dutta**, Parth Halani, and Sriparna Saha. *MultiPredGO: Deep Multi-Modal Protein Function Prediction by Amalgamating Protein Structure, Sequence, and Interaction*. ***IEEE Journal of Biomedical and Health Informatics(IEEE Transactions on Information Technology in Biomedicine)***. IEEE, 2020, **(Impact Factor: 5.180)**.

2019 **Pratik Dutta**, Sriparna Saha, and Saurabh Gulati. *Graph-based Hub Gene Selection Technique using Protein Interaction Information: Application to Sample Classification*. ***IEEE Journal of Biomedical and Health Informatics(IEEE Transactions on Information Technology in Biomedicine)***, volume 23, pages 2670–2676. IEEE, 2019, **(Impact Factor: 4.217)**.

2019 **Pratik Dutta**, Sriparna Saha, Saraansh Chopra, and Varnika Miglani. *Ensembling of Gene Clusters utilizing Deep Learning and Protein-protein Interaction Information*. ***IEEE/ACM transactions on computational biology and bioinformatics***. IEEE, 2019, **(Impact Factor: 2.896)**.

2017 **Pratik Dutta** and Sriparna Saha. *Fusion of expression values and protein interaction information using multi-objective optimization for improving gene clustering*. ***Computers in Biology and Medicine***, volume 89, pages 31–43. Elsevier, 2017, **(Impact Factor: 3.434)**.

Accepted Conference Publication

2025 Agampreet Saini, Supragya Gandotra, Abhijit Kumar, **Pratik Dutta**, and Tirthankar Ghosal. Artificial intelligence for spatial transcriptomics: A scoping review of architectures and models. In ***NeurIPS 2025 Workshop for Imageomics***: *Discovering Biological Knowledge from Images Using AI*, 2025.

2025 Nimisha Papineni, **Pratik Dutta**, Max Chao, Orbin Acanto, Rekha Sathian, Pallavi Surana, and Ramana Davuluri. Augmenting dnabert embeddings with multimodal dna features for improved regulatory sequence interpretation. In *Machine Learning in Computational Biology*, pages 294–303. PMLR, 2025.

2024 Zhihan Zhou, Yanrong Ji, Weijian Li, **Pratik Dutta**, Ramana V Davuluri, and Han Liu. *DNABERT-2: Efficient Foundation Model and Benchmark For Multi-Species Genomes*. In ***The Twelfth International Conference on Learning Representations (ICLR)***, 2024.

2021 **Pratik Dutta**, Nupur Shah, and Sriparna Saha. *A Multi-Objective Optimization-based Clustering Approach for CORD-19 Scholarly Articles*. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1393–1398. IEEE, 2021.

2021 Sriram Pingali, Shweta Yadav, **Pratik Dutta**, and Sriparna Saha. *Multimodal Graph-based Transformer Framework for Biomedical Relation Extraction*. In ***Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021***, pages 3741–3747, 2021.

2020 **Pratik Dutta** and Sriparna Saha. *Amalgamation of protein sequence, structure and textual information for improving protein-protein interaction identification*. In ***Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics(ACL)***, pages 6396–6407. Association for Computational Linguistics, 2020, **(h5-index: 135)**, **(*Core ranking: A**)**.

2019 **Pratik Dutta** and Sriparna Saha. *A Weak Supervision Technique with a Generative Model for Improved Gene Clustering*. In ***IEEE Congress on Evolutionary Computation (CEC)***, pages 2521–2528. IEEE, 2019, **(h5-index: 70)**, **(*Core ranking: B*)**.

2018 **Pratik Dutta**, Sriparna Saha, and Agni Besh Chauhan. *Predicting Degree of Relevance of Pathway Markers from Gene Expression Data: A PSO Based Approach*. In ***International Conference on Neural Information Processing***, pages 3–14. Springer, 2018, **(*Core ranking: A*)**.

2016 Ripla Roy Chowdhury, Chandan Bandyopadhyay, **Pratik Dutta**, and Hafizur Rahaman. *A Boolean Expression based Template Matching Technique for Optical Circuit Generation*. In ***Proceedings of the International Conference on Advances in Information Communication Technology & Computing***, page 36. ACM, 2016.

2015 **Pratik Dutta**, Chandan Bandyopadhyay, and Hafizur Rahaman. *All Optical Implementation of Mach-Zehnder Interferometer based Reversible Sequential Counters*. In *VLSI Design (VLSID)*, pages 232–237. IEEE, 2015.

2014 **Pratik Dutta**, Chandan Bandyopadhyay, and Hafizur Rahaman. *All optical implementation of Mach-Zehnder interferometer based reversible sequential circuit*. In ***18th International Symposium on VLSI Design and Test***, pages 1–2. IEEE, 2014.

2014 **Pratik Dutta**, Chandan Bandyopadhyay, Chandan Giri, and Hafizur Rahaman. *Mach-Zehnder Interferometer Based All Optical Reversible Carry-Lookahead Adder*. In ***IEEE Computer Society Annual Symposium on VLSI (ISVLSI)***, pages 412–417. IEEE, 2014.

Open-Source Software and Deployed Web Servers

2026 **DeepVRegulome (Python Package + Web Platform)**.
DNABERT-based deep-learning framework with 462 fine-tuned regulatory models for predicting the functional impact of short genomic variants. Deployed as a pip-installable package (`pip install deepvregulome`), with all models on HuggingFace, full source on GitHub, and an interactive Streamlit dashboard. Researchers worldwide can score any variant against 462 regulatory contexts in three lines of Python code.
Web app: `https://deepvregulome.streamlit.app/` **Code:** `DavuluriLab/DeepVRegulome`
Models: `huggingface.co/duttaprat/DeepVRegulome` **Preprint:** arXiv:2511.09026

2025 **DNABERT-Enhancer Web Platform**.
Interactive Streamlit platform for exploring enhancer predictions and SNP effects from the DNABERT-Enhancer-350 model.
Web app: `https://dnabertenhancer.streamlit.app/`

2024 **DNABERT-2 Foundation Model (ICLR 2024)**.
Co-developed and publicly released widely adopted multi-species genomic language model, available on HuggingFace and GitHub. 590+ citations as a backbone for downstream genomic AI applications.

2020 **MultiPredGO Web Server**.
Multi-modal protein function prediction tool combining protein structure and sequence.
Web server: `http://multipred.co.in/`

Tutorials and Academic Recognitions

2024 Guest lecture on **"Introduction to Multi-omics using AI/ML"** at ***The University of Chicago, Illinois***.

2020 Presenter of the tutorial **Multi-modality for Biomedical Problems: Theory and Applications** in ***IEEE World Congress on Computational Intelligence* (WCCI 2020)**, Glasgow, UK.

2018 **Session Chair of the session "Prediction"** in 25^{th} ***International Conference of Neural Information Processing* (ICONIP 2018)**, Siem Reap, Cambodia.

2018 Invited to conduct lab sessions in ***"Training Program on Machine Learning For Ocean Acoustics and Climate Data Analysis"***, during 22-36 October 2018 at **Defence R&D Organization-Naval Physical & Oceanographic Laboratory (DRDO-NPOL), Kochi, Kerala**.

2018 Invited to conduct lab sessions in ***"Faculty Development Program on Machine Learning and Applications"***, during 26-30 March 2018 at **AVB-Indian Institute of Information Technology and Management, Gwalior**.

2017 Invited as a keynote speaker in ***"Symposium on Research Trends in Machine Learning"***, held on 21^{st} December,2017 at **University Of Petroleum & Energy Studies (UPES), Dehradun**.

Fellowships & Awards

2025 **2nd place, CAMDA Challenge**, ISMB/ECCB 2025 (DNABERT-MB for antimicrobial resistance prediction).

2020 Receipt of the ***ACM India-IARCS travel grant*** to attend ***58th Annual Meeting of the Association for Computational Linguistics(ACL), 2020*** at Seattle, Washington, USA.

Jan, 2016 – Dec 2020 ***Visvesvaraya Fellowship*** of Ministry of Electronics and Information Technology (MeitY), Government of India, as a PhD research scholar in Indian Institute of Technology Patna.

2019 Receipt of ***Visvesvaraya Travel Grant*** to attend a international conference ***IEEE Congress on Evolutionary Computation, 2019*** at Wellington, New Zealand.

2018 Recipient of ***SciGenome Research Foundation (SGRF) GYAN Scholarship*** to participate ***Nextgen Genomics, Biology, Bioinformatics and Technologies-2018*** meeting at Jaipur India from 30^{th} September to 2^{nd} October 2018.

2015 Awarded under ***Students Reward Programme*** at the Annual General Meeting of **Global Alumni Association of Bengal Engineering and Science University(GAABESU).**

2015 Recipient of ***Student Fellowship*** for attending and presenting a paper in ***28th International Conference on VLSI Design***(VLSID-2015), held from 2nd January – 7th January 2015 at The Leela Palace, Bangalore, India.

2014 Recipient of ***Student Fellowship*** for attending and presenting a paper in ***18th International Symposium on VLSI Design and Test***(VDAT-2014), held from 16th July-18th July 2014 at PSG College of Technology, Coimbatore, India.

2013 – 2015 Recipient of ***GATE Scholarship*** from AICTE as a postgraduate student in IIEST Shibpur.

Academic Citizenship

Peer Review

2018 – Present **Regular Reviewer**.
Journals: IEEE JBHI, IEEE TCBB, Bioinformatics (Oxford), PLOS ONE, Computers in Biology and Medicine.
Conferences: ACL 2020, JCDL 2020, ISMB/ECCB, RECOMB.

Conference Service

2018 **Session Chair — "Prediction" Session**, *25th International Conference on Neural Information Processing (ICONIP 2018)*, Siem Reap, Cambodia.

Workshop Organization

April 2019 **Organizer — GIAN Workshop on "Introduction to Unsupervised Data Mining: From Batch to Stream Mining Algorithms"**, IIT Patna, India.

December 2016 **Organizer — GIAN Workshop on "Multi-objective Optimization"**, IIT Patna, India.

May 2016 **Organizer — GIAN Workshop on "Introduction to Natural Language Processing"**, IIT Patna, India.

October 2017 **Volunteer — #OpenGovDataHack 24-Hour Hackathon**, *Organized by National Informatics Centre (NIC), Government of India and IAMAI*, IIT Patna.

Professional Memberships and Service

2016 – 2020 **Executive Member, IEEE Student Branch**, IIT Patna, India.

Teaching and Instruction Experience

Course Instruction — Stony Brook University

Fall 2025 **BMI 503: Introduction to Programming for Biomedical Informatics**, *Instructor*, Stony Brook University.
Designed and delivered a hands-on Python curriculum for graduate students with diverse backgrounds. Developed interactive Jupyter notebooks covering data structures, genomics, multi-omics integration, and spatial transcriptomics. Live coding demonstrations and simplified case studies improved student comprehension on complex topics.

Fall 2024 & 2025 **BMI 501: Introduction to Biomedical Informatics**, *Instructor*, Stony Brook University.
Introduced core biomedical informatics concepts, connecting AI/ML methods to genomics, clinical data, and translational use cases through real-world examples.

Spring 2023-2026 **BMI 511: Translational Bioinformatics**, *Instructor*, Stony Brook Cancer Center.
Integrated AI-driven genomic analysis for cancer prognosis into the curriculum; guided students through end-to-end data interpretation pipelines using cancer genomics datasets.

Summer 2023 – 2026 **BMI Bootcamp**, *Instructor*, Department of Biomedical Informatics, Stony Brook University.
Delivered hands-on workshops on multi-omics, spatial transcriptomics, and genomic AI applications. Developed interactive tutorials adapted for students from diverse disciplinary backgrounds.

Guest Lectures and Workshops

2024 **Guest Lecture : "Introduction to Multi-omics using AI/ML"**, University of Chicago, Illinois, USA.
Adapted complex multi-modal data integration concepts using live coding demonstrations; post-session feedback showed marked improvement in student engagement.

2020 **Tutorial Presenter: "Multi-modality for Biomedical Problems: Theory and Applications"**, IEEE World Congress on Computational Intelligence (WCCI 2020), Glasgow, UK.

2018 **Lab Sessions: "Training Program on Machine Learning for Ocean Acoustics and Climate Data Analysis"**, DRDO-NPOL, Kochi, India.
Customized content for researchers and early-career professionals from non-CS backgrounds.

2018 **Lab Sessions: "Faculty Development Program on Machine Learning and Applications"**, AVB-IIITM Gwalior, India.

2017 **Keynote Speaker: "Symposium on Research Trends in Machine Learning"**, University of Petroleum & Energy Studies (UPES), Dehradun, India.

Teaching Assistantships, Indian Institute of Technology (IIT) Patna

2017 – 2020: **Undergraduate & Master's CS Courses**.

- CS564: Foundations of Machine Learning (Fall 2019, Fall 2018)
- CS342: Operating Systems Lab (Spring 2019, Spring 2018)
- CS345: Advanced Database Lab (Fall 2017, Fall 2016)
- CS101: Introduction to Computing (Spring 2016)

Student Mentoring and Supervision

2021–Present **Student Mentoring at Stony Brook University**, NY, USA.

- **Matthew Obusan** (MD/PhD student): Application of DeepVRegulome to pan-cancer (TCGA/ICGC) variant interpretation; co-author on DeepVRegulome (Genome Medicine, submitted).
- **Rekha Sathian** (PhD candidate): DNABERT-Enhancer; allele-specific enhancer activity prediction in human genome (co-first author, bioRxiv 2025).
- **Nimisha Papineni** (PhD candidate): Augmenting DNABERT embeddings with multimodal DNA features for regulatory sequence interpretation (MLCB 2025).
- **Sahanya Sivarajah** (Master's student): Application of DeepVRegulome to Autism Spectrum Disorder variant interpretation.
- **Samantha** (Master's student): Application of DeepVRegulome to Schizophrenia and Bipolar Disorder variants.
- **Sibi Satheeshkumar** (Undergraduate): Application of DeepVRegulome to Alzheimer's and Parkinson's disease ClinVar variants.

2016–2021 **Student Mentoring at IIT Patna**, India.

- **Swagarika Jaharlal Giri** (Master's student): Multi-modal protein function prediction integrating structure, sequence, and interaction (MultiPredGO, IEEE JBHI 2020).
- **Aditya Prakash Patra** (B.Tech): Deep attention-based multi-omics integration for disease gene prognosis (DeePROG, IEEE/ACM TCBB 2021).
- **Sriram Pingali** (B.Tech): Multimodal graph-based transformer framework for biomedical relation extraction (Findings of ACL-IJCNLP 2021).
- **Saraansh Chopra** (B.Tech) and **Varnika Miglani** (B.Tech intern): Ensemble deep learning with protein interactions for gene clustering (IEEE/ACM TCBB 2019).
- **Sanket Pai** and **Aviral Kumar** (B.Tech): Generative model for gene clustering using protein interaction information (Scientific Reports 2020).
- **Nupur Shah** (B.Tech intern): Multi-objective optimization-based clustering for CORD-19 scholarly articles (IEEE SMC 2021).

Referees

Prof. Ramana V. Davuluri
Professor, Department of Biomedical Informatics
Director, Bioinformatics Shared Resource,
Affiliated Faculty, Institute for
AI-Driven Discovery and Innovation
Stony Brook University, Stony Brook Cancer Center
✉ Ramana.Davuluri@stonybrookmedicine.edu

Dr. Ferhat Ay
Institute Leadership
Assoc. Prof. of Computational Biology,
Center for Autoimmunity and Inflammation,
Center for Cancer Immunotherapy,
La Jolla Institute for Immunology,
✉ ferhatay@lji.org

Dr. Tirthankar Ghosal
Scientist (NLP/AI and HPC),
National Center for Computational Sciences (NCCS)
Oak Ridge National Laboratory, United States
Affiliated Faculty, The UT-Oak Ridge Innovation Institute
University of Tennessee Knoxville
✉ ghosalt@ornl.gov

Rajarsi Gupta, MD, PhD
Assistant Professor,
Department of Biomedical Informatics,
Instructor of Clinical Pathology
Stony Brook University, Stony Brook, NY.
✉ Rajarsi.Gupta@stonybrook.edu

Dr. Sriparna Saha
Associate Professor, Department of
Computer Science & Engineering,
Indian Institute of Technology Patna, India.
✉ sriparna.saha@gmail.com

Dr. Debarshi Roy
Associate Professor,
Department of Biological Sciences,
Alcorn State University, Mississippi, USA.
✉ droy@alcorn.edu