Swagarika Jaharlal Giri

+1-7657671419| swagarika95@gmail.com | linkedin.com/in/swagarika-giri/ | Google Scholar

Background

Fourth-year PhD student in Computer Science with over seven years of experience in Machine Learning and Deep Learning research, specializing in biological and biomedical applications. My thesis focuses on leveraging Large Language Models (LLMs), Protein Language Models (PLMs), Retrieval-Augmented Generation (RAG), GraphRAG, and other deep learning techniques to improve the interpretability and prediction of protein function. Former Full-Stack Developer at Publicis Groupe, with over 2.5 years of experience in developing and maintaining production-grade web applications and services.

Purdue University, West Lafayette, IN Jan 2022 - Ongoing
Ph.D. in Computer Science *CGPA: 3.76/4*
Advisor: Dr. Daisuke Kihara

Indian Institute of Technology Patna Jun 2017 – Jun 2019
M. Tech in Computer Science and Engineering *CGPA: 9.31/10*
Advisor: Dr. Sriparna Saha

Maulana Abul Kalam Azad Unniversity of Technology, West Bengal Jun 2013 – May 2017
B. Tech in Computer Science and Engineering *CGPA: 9.11/10*

Publication

S. J. Giri, U. Pandey, J. H. Park, and D. Kihara, "Translating a GO term list to human readable function description using GO2Sum," **Methods in Molecular Biology, 2025**

S. J. Giri, U. Pandey, J. H. Park, and D. Kihara, "NaviGO: An Interactive Tool for Gene Ontology Functional Analysis with Free Text GO Summaries," **Methods in Molecular Biology, 2025**

S. J. Giri, N. Ibtehaz, L. Kurgan, and D. Kihara, "Identifying Hidden Moonlighting Proteins and Domains," **Methods in Molecular Biology, 2025**

S. J. Giri, N. Ibtehaz, and D. Kihara, "GO2SUM: Generating Human readable functional summary from GO terms," **Nature Publishing Journal (NPJ) Systems Biology and Applications, 2024**

Yury V. Bukhman, ..., **S.J.G**, D. Kihara, ..., and Ron Stewart, "Chromosome level genome assembly of the Etruscan shrew Suncus etruscus," **Scientific Data, 2024**

P. Panda, **S. J. Giri**, L. Sherman, D. Kihara, and U. Aryal, "Proteomic analysis of unicellular cyanobacterium Crocosphaera subtropica ATCC 51142 under extended light and dark growth," **Journal of Proteome Research, 2025**

P. Panda, **S. J. Giri**, L. Sherman, D. Kihara, and U. Aryal, "Proteomic changes orchestrate metabolic acclimation of a unicellular diazotropic cyanobacterium during light dark cycle and nitrogen fixation states," **Scientific Reports (Minor Revision), 2025**

S. J. Giri, P. Dutta, P. Halani, and S. Saha, "MultiPredGO: Deep Multi-Modal Protein Function Prediction by Amalgamating Protein Structure, Sequence, and Interaction Information," **IEEE Journal of Biomedical and Health Informatics, 2020**

S. J. Giri and S. Saha, "Multi-View Gene Clustering using Gene Ontology and Expression-based Similarities," **IEEE Congress on Evolutionary Computation (CEC), Glasgow, UK, 2020**

Research Experience

Qualcomm (Patent Filed) May 2024 - Aug 2024
Manager: Arvind Santhanam and Amr Martini

- **Travel Planner**: Investigated and developed a Large Language Model-based planner that generates a travel itinerary based on given dates and constraints. The planner makes API calls to various services, including flights, cabs, and distance calculations, to create an optimized travel plan. Additionally, it evaluates the feasibility of the proposed itinerary.

Purdue University - Kihara lab Jan 2022 - Ongoing
Advisor: Dr. Daisuke Kihara

- **IntegGO – Integrating heterogeneous data sources for GO annotation (Ongoing).** Developing a Q-Former-based model that takes protein features as input to generate GO annotations and functional descriptions. To handle missing modalities, the system integrates a GraphRAG built on a knowledge graph derived from diverse protein data sources. Additionally, to address outdated database limitations, it incorporates Lit-RAG, which retrieves up-to-date evidence from biomedical literature.
- **Lit-RAG: Leveraging Biomedical Literature for Protein Function Prediction :** A Retrieval-Augmented Generation (RAG) framework for protein function prediction, leveraging a large biomedical literature database of over 40 million papers. Utilized BM25 and FAISS-based search techniques to retrieve relevant documents. Achieved state-of-the-art performance in predicting functions for novel proteins and GO terms.
- **GraphRAG-PFP: Leveraging protein knowledge graphs and the capabilities of GraphRAG for protein function prediction:** Designed and implemented a GraphRAG-based system to predict protein functions by integrating sequence, structure, and network data through graph-based retrieval and multi-hop reasoning.

- **GO2Sum: Generating Human Readable Functional Summary of Proteins from GO Terms :** Developed, designed, and implemented a novel Large Language Model (LLM)-based summarizer model for better interpretability of protein function, with significant improvements over the state-of-the-art SciFive.
- **Gene Ontology(GO) Prediction and Enrichment for Etruscan shrew Suncus etruscus:** Generated protein functions for 25,000 genes for the world's smallest mammal using well-known protein function prediction models like PFP, Phylo-PFP, and ESG. Additionally, conducted GO enrichment analysis using NaviGO.
- **Gene Ontology(GO) Prediction and Enrichment for Crocosphaera subtropica:** Generated protein functions for 5,000 genes for Cyanobacterium during different light-dark conditions using well-known protein function prediction models like PFP, Phylo-PFP, and ESG. Additionally, conducted GO enrichment analysis using NaviGO.

Purdue University Aug 2023 - Dec 2023

Advisor: Dr. Dan Goldwasser

- **Crafty or Faithful? The Mixed Bag of Synthetic Prompts!:** Designed and developed a novel method that leverages LLMs' generative abilities for synthetic prompt generation, emphasizing the importance of topic relevance and complexity for enhancing model performance.

Purdue University Jan 2023 - May 2023

Advisor: Dr. Aniket Bera

- **Walk this Way: Using Gait Analysis to Detect Parkinson's Disease:**Led a team of three in designing, developing, and implementing a tool for detecting Parkinson's disease from gait analysis. Exploring feature extraction methods and transfer learning across multiple datasets for improved diagnosis and disease monitoring.

Indian Institute of Technology Patna - AI-ML-NLP lab Jun 2017 - Jun 2019

Advisor: Dr. Sriparna Saha

- **MultiPredGO:** Designed, developed, and implemented a deep learning-based model that combines the features learned from protein structure, PPIN (protein-protein interaction), and FASTA sequence to predict gene ontology-based protein functions. Also, developed a full-stack web application version of the work and deployed it on AWS.
- **Multi-View Gene Clustering using GO and Expression-based Similarities:** Developed a novel approach to solve the problem of Multi-view gene clustering by formulating it as a Multi-objective optimization problem.

Indian Statistical Institute Kolkata May 2016 - Jul 2016

Advisor: Dr. Jija Das Gupta

- **Tool for Text Dependent Writer Recognition** Led a team of 3 in developing a handwriting recognition system for identifying writers on Indian bank cheques. Employed image processing techniques to extract features on 30K handwritten images to enhance system performance.

Work Experience

Graduate Teaching Assistant Jan 2022 – Present

Purdue University, West Lafayette, IN

- Teaching Assistant for CS:573 Data mining and CS:182 Foundations of Computer Science.
- Conducted PSO, office hours, and grading for a class with a large class size of approximately 870 students, respectively.

Full-Stack developer Jul 2020 – Nov 2021

Publicis Groupe *Client: CJ Affiliate*

- One of the core members for release planning, architechture design and tech-stack finalization for Data Transfer project.
- Development of Data transfer application that uses big data technologies like Scala, Spark, etc, and efficiently processes and validates the file by adding efficient header and transforms the data into required format i.e readable by REST-APIs.
- End-to-end development, testing, and deployment of the e-commerce plugin in the Shopify app store using AWS Lambda, Jest, React, AWS, and Jenkins.

Senior React developer July 2019 – Apr 2020

Publicis Groupe *Client: MLC Australia*

- Worked on technologies like ReactJs, NodeJs, CSS, JavaScript to develop the Front-end application for MLC Insurance.

Graduate Teaching Assistant Jul 2017 – Jun 2019

Indian Institute of Technology Patna

- Teaching Assistant for CS341 - Operating Systems and CS564 - Introduction to Machine Learning.
- Coordinated the assignments, viva, and capstone project for class strength of more than 100 students.

Technical Skills

Machine Learning Models: LLMs, PLMs, RAG, GraphRAG, Q-Former

Machine Learning Tools: PyTorch, TensorFlow, Keras, NumPy, Pandas, PySpark, Scrapy

Languages: Python, Java, C/C++, SQL (PostgreSQL), JavaScript, HTML/CSS, R

Frameworks: React, Flask, Node.js, JUnit, Jest, Material-UI

Developer Tools: Git, AWS, GCP, Docker, VS Code, Visual Studio, PyCharm, IntelliJ, Eclipse

Libraries: NumPy, Pandas, Matplotlib, TensorFlow, Keras

Other: Computer Vision, Data Mining

Internship and Workshop

ACM summer school on Data Science Jun 2018 – Jul 2018
Organizer: ACM India and RBC DSAI, IIT Madras

Global Initiative of Academic Networks on Neural Machine Translation. Dec 2017
Advisor: Dr. Andy Way, School of computing, Dublin City University.

Third Summer School on Computer Vision, Graphics and Image Processing Jun 2016 - Jul 2016
Organizer: ISI Kolkata, India

Cultural and Social Activity

- Level 2 holder in Toastmasters at Purdue, a public speaking club under the Presentation Mastery pathway. Held roles such as Toastmaster, Table Topics Master, General Evaluator, and Grammarian. Jan 2022 – Present
- Volunteer at SKY@Happiness, a meditation, yoga, and spirituality club. Nov 2024 – Present
- Treasurer at Toastmasters at Purdue, a public speaking club. July 2023 – June 2024
- Food Secretary at Purdue University Tagore Society. July 2023 – June 2024
- Speaker at a two-day Faculty Development Program at Rajkiya Engineering College, Uttar Pradesh, India. 2020
- Technical Coordinator at Social Service@RCCIIT, an organization dedicated to providing students in rural areas with fundamental knowledge of science, technology, and basic computer skills. Sep 2013 - Jun 2017

Award and Recognition

- Received Scholarship from Purdue University to attend **Volunteer Training Program for SKY Happiness, Boone, North Carolina** (2024).
- Received the **Partnering for Client Impact** award at Publicis Sapient (2019).
- Ranked **2nd out of 20 students** in M.Tech CSE at IIT Patna.
- Received a travel grant to attend the CoDS-COMAD Conference (2019).
- Received a travel grant to attend the Grace Hopper Celebration India (2018).
- Achieved **All India Rank (AIR) 1011** in the Graduate Aptitude Test in Engineering (GATE) for Computer Science (2017).
- Secured **3rd out of 120 students** in B.Tech CSE at RCC IIT, Kolkata.
- Achieved 3^{rd} position for a Paper Presentation dealing with topic *Hybrid Apps* at RCC IIT, Kolkata.

Academic Responsiblities

- Reviewer **at International Conference on Machine Learning (ICML) and International Conference on Learning Representations (ICLR).**
- Reviewer at **IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB).**
- Leads the team of three developers who maintain Protein Function Prediction(PFP) servers PFP, Phylo-PFP, ESG, GO2Sum, and NaviGO.
- Leads the team of three developers who maintain 3D-Surfer which is a web-based tool from the Kihara Lab for fast and efficient protein surface shape similarity search and comparison using 3D representations.
- Mentor at **Toastmasters at Purdue.**

Referees

Dr. Daisuke Kihara
Prpfessor of Department of Biological Sciences and the Department of Computer Science *Purdue University, West Lafayette*

Dr. Sriparna Saha
Associate Professor, Computer Science and Engineering. *Indian Institute of Technology Patna*

Nilesh Khabra
Head of Engineering,CJ-Inida, CJ Affiliate *Publicis Groupe*