# Dipayan **Sarkar**

Siliguri, West Bengal, India
☎ +91 9735490241 • ✉ dipayansarkar26@gmail.com • dipayansarkar.com
Dipayan26 • Google Scholar • in dipayansarkar • ORCID
🤗 dipayan26

## Research Summary

Currently pursuing a PhD in bioinformatics and computational biology, focused on deep learning methods for biological sequence-based prediction and generative model development. Experience includes neural network training, transfer learning with pretrained protein language models, and analysis of large-scale biological datasets. Current work involves autoregressive generative modeling of biological sequences, with ongoing exploration of diffusion-based approaches for biological data.

## Work Experience

**University of North Bengal** **Siliguri, India**
*PhD Research Scholar* *2022–Present*
Department of Bioinformatics

- Designed and implemented deep learning architectures for sequence-based protein–protein interaction prediction.
- Developed attention-based hybrid models integrating transfer learning from pretrained protein language models, trained and evaluated on large-scale biological sequence datasets.
- Developed and deployed web-based platforms for model inference and protein interaction network visualization.

## Education

**University of North Bengal** **India**
*Ph.D. in Bioinformatics* *Expected March 2027*
Advisor: Dr. Chiranjib Sarkar
Focus: Deep learning-based frameworks for protein–protein interaction network prediction.

**University of North Bengal** **India**
*M.Sc. in Botany, 77.25% – First Class* *2021*
Specialization: Biochemistry

**Ananda Chandra College, University of North Bengal** **India**
*B.Sc. (Hons.) in Botany, 63.50% – First Class* *2019*

**Haldibari High School (H.S.)** **West Bengal**
*Higher Secondary (Class XII), 80.00% – First Class* *2016*
WBCHSE | Subjects: Physics, Chemistry, Mathematics, Biology, English, Bengali

**Haldibari High School (H.S.)** **West Bengal**
*Secondary (Class X), 91.58% – First Class* *2014*
WBBSE | Subjects: Mathematics, Physical Science, Life Science, History, Geography, English, Bengali

## Publications

**2025**: Sarkar, D., & Sarkar, C. AttnSeq-PPI: Enhancing protein-protein interaction network prediction using transfer learning-driven hybrid attention. *Biochimica et Biophysica Acta (BBA)—Proteins and Proteomics*, 141102. doi:10.1016/j.bbapap.2025.141102

**2022**: Sarkar, C.; Sarkar, D.; Parsad, R.; Mishra, D. Package `EGRNi`. **CRAN**. `cran.r-project.org/package=EGRNi`

## Skills

**Programming**: Python (PyTorch, HuggingFace), R (CRAN package development), C (basic)

**Deep Learning**: Transformers, attention mechanisms, sequence modeling, transfer learning with pretrained language models

**Generative AI**: Autoregressive sequence generation, diffusion-based approaches, probabilistic sequence modeling for biological data

**Scalable Training**: GPU-based training, large-scale sequence datasets, parallel training on HPC environments

**Bioinformatics**: Biological sequence analysis, sequence-only protein–protein interaction prediction, generative modeling of biological sequences

**Data & Viz**: NumPy, Pandas, Matplotlib, Seaborn

**Deployment**: Docker, Flask, FastAPI, Linux server deployment, model-backed web applications

## Software & Tools

**AttnSeq-PPI**: Deep learning-based protein–protein interaction prediction platform.
`compbiosysnbu.in/attnseqppi/`

**ARACoFusion-PPI**: Arabidopsis-specific PPI prediction and network analysis tool.
`aracofusion.compbiosysnbu.in/`

## Awards & Certificates

**2021**: Qualified **CSIR-NET (JRF)** in Life Sciences — June 2021, **All India Rank: 216**

**2022**: Qualified **GATE** in Life Sciences (XL)

## Grants & Computing Resources

**2025**: **IndiaAI Compute Initiative** — Awarded GPU cloud compute (NVIDIA L4) under the IndiaAI Mission for the project *Deep learning-based framework for protein-protein interaction network prediction* (Project ID: P1-S2025070964, Sep 2025 – Aug 2026). Funded under CSIR-UGC SRF Fellowship.

**2025**: **Google TRC (TPU Research Cloud) Award** — Granted free access to Google Cloud TPUs (v4, v5e, v6e) for machine learning research.