

Raunak Kumar

B.Tech, Electronics and Communications Engineering
Indian Institute of Technology Roorkee
Email: raunak_k@ece.iitr.ac.in

Mobile: +91-80156-23648
Github: github.com/0xGu4rdI4N
LinkedIn: linkedin.com/in/raunak-kumar
Website: 0xgu4rdi4n.github.io

EDUCATION

- Indian Institute of Technology** Roorkee, India
B.Tech — Electronics and Communication Engineering 2023 – 2027
 - **CGPA:** 8.82/10
 - **Relevant Coursework:** Fundamentals of AI/ML, Data Science, Biosignal & Image Analysis, Numerical Methods, Bioinformatics, High Throughput Sequencing, Physics-101

RESEARCH INTERESTS

- Developing AI Methods to understand Cellular State Transitions

RESEARCH EXPERIENCE

- P² Lab** On-site
Undergraduate Researcher March 2026 – August 2026
Tabular In-Context Learning for Antimicrobial Peptide profiling
 - Developed a gradient-free, tabular in-context learning framework using **TabPFN** and 330 sequence-derived descriptors, completely bypassing computationally heavy 3D structure prediction to profile multi-activity antimicrobial peptides.
 - Achieved state-of-the-art multi-label performance on the ESCAPE benchmark (82k+ peptides, 5 labels), increasing macro AP from **72.1% to 77.8%**, matching or outperforming prior deep multimodal baselines across all 5 targets simultaneously.
 - Demonstrated strong out-of-domain generalization with an **+11.2 AP gain** on remote homologs (< 30% sequence identity), showing that as few as 10 physicochemical scalars recover 91% of predictive performance for rapid experimental screening.
- Singla Lab** On-site
Undergraduate Researcher May 2025 – Present
Template-free Subtomogram Alignment & Classification
 - Converted a template-free alignment hypothesis of cryo-ET subtomograms into a End-to-End computational framework. Developed Fourier-space translational refinement and Lie algebra-based rotational optimization on Riemannian manifolds.
 - Implemented distributed GPU alignment routines, reducing runtime by **20×** and enabling **4×** larger batch processing. Subsequently derived an analytical gradient-based optimization procedure to further reduce the runtime.
- Communication Systems and Machine Learning Lab** On-site
Undergraduate Researcher January 2025 – June 2025
Cognitive Workload Classification using TDA
 - Investigated whether EEG signals encode cognitive workload through persistent topological features beyond conventional spectral statistics.
 - Extracted persistent-homology-based features and integrated them with spectral and statistical descriptors to classify cognitive workload levels.
 - Designed and implemented an end-to-end pipeline for EEG preprocessing, multimodal feature extraction, model training, and evaluation.

EXPERIENCE

- Mandrake Bioworks** Remote
AI Research Intern October 2025 – December 2025
De novo Protein Design
 - Developed a de novo protein design pipeline for reverse transcriptases targeting prime editing, combining RFdiffusion backbone generation with ProteinMPNN inverse folding to generate 10,000 candidate sequences.
 - Designed a UMAP-based filtering and segmentation pipeline to organize candidates by sequence-family representation and prioritize promising designs.
 - Adapted Gemma-4B for protein language modeling by expanding its vocabulary with amino-acid tokens and fine-tuning via LoRA on paired natural-language-protein data using causal language modeling.

- **Amgen Scholars Program – IIIT Hyderabad**
Amgen Scholar

On-site
July 2026 – August 2026

Bayesian Optimization for Ultra-Large Virtual Screening of EphA2 Ligands

- Independently designed and validated bayesian optimization framework for virtually screening a 100M small-molecule library, recovering **>90%** of the top-500 candidates while docking only **~1%** of the library
- Achieved a docking throughput of **34.1 ms/molecule**, enabling **~1M** molecules to be screened in 25–33 hours per receptor on a single GPU node.
- Developed a hierarchical screening pipeline integrating ensemble docking and molecular dynamics to refine candidates from the large-scale search to **40 physics-refined candidates** and **5 high-confidence EphA2 binder candidates** for experimental validation.

TECHNICAL SKILLS

- **Systems, Programming & Performance:** Python; Linux systems programming; CUDA; memory management and performance optimization; multithreaded and distributed execution; Git, Docker
- **Machine Learning & Representation Learning:** Diffusion models; geometric deep learning; physics-informed neural networks (PINNs); topological data analysis; curriculum learning; large-scale and distributed training
- **Scientific Computing & HPC:** CUDA acceleration; large-scale training on HPC clusters; Slurm-based workflows; profiling, batching, and pipeline optimization for distributed workloads

PUBLICATIONS AND MANUSCRIPTS

- **Coarse composition suffices: tabular in-context learning for multi-activity antimicrobial peptide profiling:** A. Pal*, **Raunak Kumar***, D. Solanki*, P. Pareek, J. Singh, J. Singla. (*Equal contribution). *bioRxiv* preprint; under review at *Bioinformatics* (Oxford University Press). Link: <https://www.biorxiv.org/content/10.64898/2026.08.27.747591v2>
- **Template-free Fourier Space alignment of subtomograms:** Manuscript in preparation for submission to *Structure*

HONORS AND AWARDS

- **All India Rank 1,359, JEE Advanced 2023:** Secured All India Rank 1,359 among approximately 180,000 candidates who qualified for the examination, placing in the **top 0.8%** nationwide.
- **All India Rank 1831, JEE Main 2023:** Secured All India Rank 1,831 among approximately 1.5 million candidates, placing in the **top 0.12%** nationwide.
- **Best Presentation Award, Asia Amgen Scholars Symposium 2026:** Selected to represent the Indian Amgen Scholars cohort at The University of Tokyo and received the Best Presentation Award among 75 international scholars for my research on Bayesian optimization for ultra-large virtual screening of EphA2 ligands.
- **NSEP State Topper 2022:** Recognized among the **top 1%** of candidates in the state in the National Standard Examination in Physics, a national-level physics olympiad qualifying examination in India.
- **NTSE Scholar:** Awarded the National Talent Search Scholarship by the Government of India, ranking among the **top 0.1%** of over 1.5 million candidates nationwide. Recognized for exceptional aptitude in Science and Mathematics.
- **Micron Memory Makers Hackathon — 1st Place (2025):** Developed the top-performing and most lightweight model in the competition, optimized for edge deployment with a significantly reduced memory footprint

PROJECTS

- **PINN for Temperature Field Interpolation:** Formulated and trained a physics-informed neural network for temperature super-resolution by enforcing heat-transfer PDE constraints during training. Demonstrated improved physical consistency and generalization compared to purely data-driven interpolation, particularly in low-data environments
- **Model Compression for Edge Deployment:** Studied the accuracy–capacity trade-offs in deep CNNs by applying structured pruning and quantization to VGG16 and ResNet50 using TensorFlow Lite. Achieved approximately 90% model size reduction with less than 1% accuracy degradation on held-out test sets
- **MaskGAN for Shadow Detection and Removal:** Designed a conditional GAN framework for shadow removal by jointly learning shadow masks and illumination-corrected reconstructions. Combined adversarial and reconstruction losses to stabilize training and improve robustness across varying shadow geometries

VOLUNTEER EXPERIENCE

- **Core Member — InfoSecIITR:** Conducted cybersecurity workshops; authored CTF challenges (BackdoorCTF 2024, n00bCTF 2025); As a cybersecurity team, reached global top-30 overall in CTF's
- **Cadet — National Cadet Corps (NCC):** participated in leadership and discipline training through regular drills, camps, and community service