Yash Mathur

mathury.1259@gmail.com ym59.github.io Delhi, India

PhD scholar in computational biology with **357 citations** and an **h-index of 5**, driving **AI-powered drug discovery** through generative deep-learning pipelines and multi-criteria optimization frameworks under Dr Md Imtaiyaz Hassan at Jamia Millia Islamia. Passionate about building robust, user-focused computational platforms that turn cutting-edge research into real-world solutions.

Skills

Proficient in languages including **Python, R, C++, and SQL**. Experienced in Machine Learning and AI with frameworks such as **TensorFlow, PyTorch, Keras, Scikit-learn, and CatBoost**, complemented by data science libraries like **Pandas, NumPy, Matplotlib, and Seaborn**. Skilled in computational biology and bioinformatics using tools like **Autodock Vina, RDKit, OpenBabel, DESeq2, PyMol, and Biopython**. Knowledgeable in web and GUI development with **Django, Flask, and PyQt**, and proficient in DevOps practices using **Git** and designing **REST APIs**.

Research Experience

Project Assistant | Dr. Md. Imtaiyaz Hassan's Lab, Jamia Millia Islamia, New Delhi | 2020-2023
Authored seven publications and developed two automated bioinformatics tools (InstaDock, PyPAn) that accelerate workflows by up to 100x in fields including cancer genomics and structure-based drug discovery.

Major Projects

InstaDock (v1 & v2) *Ongoing*
Automated molecular docking

- Developed InstaDock v1: Python/QuickVina-W pipeline for one-click batch docking and pose scoring. (DOI: 10.1093/bib/bbaa279) (***2021***) (250+ citations)
- Created InstaDock v2: Open-source cross-platform PyQt GUI integrating AI-based ADMET prediction (CatBoost + PyTorch + Scikit-learn). (Under review for publication)
- Automated end-to-end solution for all molecular docking requirements
- Available at www.hassanlab.in/instadock

Thesis Research *Ongoing*
AI-based molecular generator

- Implementing of a next-generation molecular structure generator.
- Creating a multi-criteria optimization framework to balance compound efficacy, pharmacokinetics, and safety profiles.

Hassanlab.in *2025*

- Built www.hassanlab.in backend with Python publication scraper and Tailscale-secured API
- Developed HTML/CSS/JS frontend and managed hosting, forms and overall site infrastructure beyond standard Wix functionality

PyPAn *2023*
A complete protein analysis suite (DOI: 10.2174/0929866529666220210155421)

- Created a PyQt application offering protein sequence and structural analyses
- Engineered a modelling advisor that recommends optimal strategies for 3D structure prediction directly from raw sequences.
- Implemented structure-refinement tools to minimize or compute protein energy from PDB files, enhancing model quality.
- Available at www.hassanlab.in/pypan**.**

Education

PhD in Computational Biology (Expected 2026)
Jamia Millia Islamia | New Delhi, India
Thesis title: "Employing Generative Deep Learning Models for Strategic Advancements in Drug Development"

Master of Science in Bioinformatics (2020)
Jamia Millia Islamia | New Delhi, India
Dissertation title: "Modelling Method Advisory Tool: a web-based module for protein modelling advise"

Bachelor of Science in Microbiology (2018)
University of Delhi | New Delhi, India