

EDUCATION

- **Birla Institute of Technology and Science, Pilani** Hyderabad, India
Bachelor of Engineering in Computer Science; CGPA: 8.87/10 *Aug 2017 – May 2021*

EXPERIENCE

- **University College London** London, the UK
Research Assistant (Supervisor: Prof. Christine Orengo) *Jun 2021 - Apr 2022*
 - **CATHe: Detection of Remote Homologues** [Advisor: Prof. Burkhard Rost, TUM]
 - * Developed a DL tool, CATHe, that could identify remote homologues better than HMM-based approaches. This tool attained an **accuracy of 91%** using an ANN trained on **ProtT5** protein sequence embeddings
 - * Pulled in 9.8 million homologues from Pfam into CATH using CATHe, to expand the database by **6.5%** [\[code\]](#)
 - **AlphaFold Model Organism Analysis**
 - * Classified domains not in CATH from the 21 model organisms, into CATH superfamilies using CATHe
 - * Validated CATHe predictions using Foldseek, TMalign, and SSAP on **AlphaFold** predicted structures
 - **TTS Enzymes Study** [Advisors: Prof. Janet Thornton, EBI & Prof. Anne Osbourn, JIC]
 - * Worked on understanding the TriTerpene Synthases (TTS) by performing an **HDBSCAN** clustering on their sequence embeddings. This was done to group together proteins with similar mechanisms and products

Bachelor's Thesis Student (Supervisor: Prof. Christine Orengo) *Jan 2021 - May 2021*

 - **CATH Structurally Similar Groups Classification** [Advisor: Prof. Burkhard Rost, TUM]
 - * Classified protein sequences present in CATH v4.3 into structurally similar groups using DL techniques
 - * Achieved an **F1-Score of 85%** with an ANN model trained on **ProtBert** embeddings of the sequences [\[code\]](#)
 - **PETase Melting Point Prediction** [Advisor: Dr. Florian Hollfelder, Cambridge University]
 - * Designed a DL model to predict the melting point from protein sequence, using the BRENDA database
 - * This model made binary predictions to judge whether a sequence could be used for degrading PET [\[code\]](#)
- **SciLifeLab: Stockholm University** Stockholm, Sweden
Research Assistant (Supervisor: Prof. Arne Elofsson) *Jun 2020 - Dec 2020*
 - Constructed a dataset by extracting pairwise features such as gaussian DCA, mutual information, and cross-entropy from the multiple sequence alignment of 25,000+ protein chains, for studying **inter-residue distance prediction**
 - Designed a **U-Net** model which attained a **PPV of 59.74%**, comparable to the state-of-the-art trRosetta [\[code\]](#)
- **University of Ljubljana** Ljubljana, Slovenia
Research Assistant (Supervisor: Prof. Peter Peer) *Jun 2020 - Dec 2020*
 - Curated the largest, most comprehensive dataset of COVID-19 chest x-ray images by far, to diagnose COVID-19
 - Proposed an **ensemble CNN model of MobileNetv2 & VGG 19**, and attained **99.84% accuracy** [\[code\]](#)
- **Carnegie Mellon University** Pittsburgh, the USA
Research Intern (Supervisor: Dr. Min Xu) *Jul 2020 - Sep 2020*
 - Applied CV algorithms to classify bio-molecules simulated in **cryo-ET tomograms** at three different SNR values
 - Devised an end-to-end model with a **denoising Convolutional Auto-Encoder** and a **3D SqueezeNet** [\[code\]](#)
- **BITS Pilani, Hyderabad Campus** Hyderabad, India
Undergraduate Researcher *Aug 2018 - Dec 2020*
 - **Cysteine Modification Prediction** (*Supervisor: Dr. Debashree Bandyopadhyay*)
 - * Scraped the Protein Data Bank to develop a dataset for the detection of four different cysteine modifications, and extracted features such as buried fraction, rHpy, secondary structure, motifs, enzyme class, and pKa
 - * Designed a **deep neural network with residual blocks** and tested it on domains of unknown functions
 - * Built a web server for the model using Django and deployed it to the internet using Heroku [\[web-server\]](#) [\[code\]](#)
 - **Serine Repeat Antigen and Egress** (*Supervisor: Prof. Vidya Rajesh*)
 - * Devised multiple hypotheses to understand the negative kinetic regulator action of Serine Repeat Antigen 5 (SERA5) in the egress mechanism of red blood cells, pertinent to the malaria parasite *Plasmodium falciparum*
 - * Conducted **molecular docking** simulations using Schrödinger for in-silico validation of these hypotheses

Teaching Assistant *Aug 2018 - Dec 2019*

 - **Discrete Structures for Computer Science (CS F222)**: Prepared weekly competitive coding assignments, and conducted monthly algorithm coding tutorials for a class of 120+ students
 - **Mechanics, Oscillations & Waves (PHY F111)**: Participated in preparing quizzes and conducted weekly problem solving sessions for a batch of 50+ students

PERSONAL PROJECTS

- **ABLE:** Created a dataset and developed a **Bidirectional LSTM** network with **Attention** to classify protein sequences into the six enzyme classes and a negative class, using labels from their Enzyme Commission (EC) numbers [\[code\]](#)
- **Anti-HIV:** Developed a **Graph Neural Network** model that was able to predict using SMILES representations with an ROC-AUC score of 0.745, which chemical compound could be used to effectively inhibit HIV replication [\[code\]](#)
- **Player Detection and Tracking:** An OpenCV and MobileNetv2 ensemble network to detect, categorize, and track the players along with the ball in a football match video. Used for generating training statistics for football teams [\[code\]](#)

PUBLICATIONS AND PRESENTATIONS

- **Nallapareddy, V.**, et al. (2021). “DeepCys: Structure-Based Multiple Cysteine Function Prediction Method Trained on Deep Neural Network: Case Study on Domains of Unknown Functions Belonging to COX2 Domains.” *Proteins*, vol. 89, no. 7, pp. 745–61. DOI: [10.1002/prot.26056](https://doi.org/10.1002/prot.26056)
- **Nallapareddy M. V.**, and Dwivedula, R. (2021). “ABLE: Attention Based Learning for Enzyme Classification”. *Computational Biology and Chemistry*, 107558, ISSN 1476-9271, DOI: [10.1016/j.compbiolchem.2021.107558](https://doi.org/10.1016/j.compbiolchem.2021.107558)
- Liu, S., Ma, Y., Ban, X., Zeng, X., **Nallapareddy, M. V.**, Chaudhari, A. and Xu, M. (2020). “Efficient Cryo-ET Simulation of Macromolecules with Neighbors and SARS-CoV-2.” *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 80–87. DOI: [10.1109/BIBM49941.2020.9313185](https://doi.org/10.1109/BIBM49941.2020.9313185)
- Bordin, N., Sillitoe, I., **Nallapareddy, V.**, Littmann, M., Heinzinger, M., Rost, B., Orengo, C. (September 28, 2021). “Improvements in CATH-Gene3D v4.3 and its applications to diverse sets of proteins” *UK Bioinformatics, and Computational Biology Conference 2021*. [\[certificate\]](#)
- **Nallapareddy, V.**, et al. (January 15, 2020). “Evaluating SERA 5 Competition for Sub-1 Mediated Cleavage of SERA 6 in *Plasmodium Falciparum* and Its Role in the Kinetics of Egress Using In-Silico Tools.” *International Conference on Drug Discovery (ICDD) 2020*. [\[abstract\]](#) [\[poster\]](#)
- **Nallapareddy, M. V.** and Dwivedula, R. (April 27, 2021) “ABLE: Attention Based Learning for Enzyme classification”. *Students’ Research Conference (SRC) 2021*, IIT Kanpur. [\[certificate\]](#)
- **Nallapareddy, V.**, Bordin, N., Sillitoe, I., Heinzinger, M., Littmann, M., Orengo, C., Rost, B. “CATHe: Detection of remote homologues for CATH superfamilies”. In preparation
- Bordin, N., **Nallapareddy, V.**, Sen, N., Rauer, C., Heinzinger, M., Rost, B., Orengo, C. “Massively Expanding CATH with Predicted Structures from AlphaFold”. In preparation.

AWARDS & HONORS

- **BITS Pilani Merit Scholarship:** Awarded for being in the top 1% of my class
- **National Talent Search Scholarship:** Government of India merit scholarship awarded to the top 0.05%
- **Summer Schools & Conferences:** [ISCB ISMB/ECCB 2021](#), [Google India AI Summer School 2020](#), [EMBL PDB50](#)
- **First Runner Up at InnoHack 7.0 Pan-India Hackathon:** Built a machine learning system to detect maintenance issues in heavy machinery and integrated it with a dashboard built in MERN

SKILLS

- **Programming Languages:** Python, C/C++, Java, SQL
- **Python Libraries:** PyTorch, TensorFlow/Keras, DGL/DGL-LifeSci, ProtTrans, Biopython, Scikit-Learn, OpenCV, Numpy, Pandas, Matplotlib/Seaborn, Gensim, NLTK, and OpenAI Gym
- **Tools/Frameworks:** Git, HPC, Django, Heroku, Bash
- **Bioinformatics Tools:** MMseqs2, BLAST, Schrödinger Suite, PyMOL, ClusPro 2.0, SwissModel, DSSP, PROPKA, I-TASSER, Robetta, trRosetta, ClustalW, ViennaRNA, and ORF FINDER

COURSEWORK

- **Birla Institute of Technology and Science, Pilani:** Deep Learning, Artificial Intelligence, Information Retrieval, Data Mining, Data Structures & Algorithms, Object Oriented Programming, General Biology, Biological Chemistry, Introduction to Bioinformatics, Differential and Integral Calculus, Linear Algebra, and Probability & Statistics
- **Birkbeck, University of London:** BioComputing I, Molecular Basis of Life, Sequence Analysis & Genomics
- **Coursera:** [Introduction to the Biology of Cancer \(JHU\)](#), [Genomics \(UIUC\)](#), [AI for Medical Diagnosis \(deeplearning.ai\)](#), [Fundamentals of Reinforcement Learning & Sample-based Learning Methods \(UAlberta\)](#)