

## EDUCATION

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

## RESEARCH EXPERIENCE

- **University College London** London, the UK  
*Honorary Research Assistant (Supervisor: Prof. Christine Orengo)* Jun 2021 - Present
  - **Detection of Remote Homologues** [Advisor: Prof. Burkhard Rost, TUM]
    - \* Developed a DL tool for analysing and classifying new entries to the PDB database into CATH superfamilies
    - \* Achieved a performance of **91%** with an Artificial Neural Net paired with the ProtT5 embeddings [\[code\]](#)*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]
    - \* A DL model was designed 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 PETase experiments [\[code\]](#)
- **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 [\[web-server\]](#) [\[code\]](#)
  - **Prediction of N6-Methyladenosine (m6A) Sites** (Supervisor: Dr. Piyush Khandelwal)
    - \* Constructed a **Convolutional LSTM with Attention** to detect m6A sites at a single-base resolution in RNA using sequence motifs and secondary structure features, for studying the infectivity of SARS-CoV-2
    - \* Achieved a **precision of 94%**, outperforming the state-of-the-art DeepM6ASeq model by 6% [\[code\]](#)
  - **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
  - **Adaptive Bit Rate (ABR) Estimation in Video Streaming** (Supervisor: Dr. Paresch Saxena)
    - \* Advanced the state-of-the-art Pensieve framework by implementing the Follow then Forage method to increase exploration in the **Asynchronous Advantage Actor-Critic (A3C)** deep reinforcement learning network
    - \* Attained a 6% increase in the Quality of Experience (QoE) and a 3% decrease in the training time [\[code\]](#)

## PERSONAL & COURSE 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 which was able to predict which chemical compound could be used to effectively inhibit HIV replication, using their SMILES representations, with an ROC-AUC score of 0.745 [\[code\]](#)
- **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 [\[code\]](#)
- **RL Game Agents:** Implemented various reinforcement learning algorithms such as Q-Learning, Deep Q-Network, Temporal Difference (TD) Learning, SARSA, and Expected SARSA to solve game environments in OpenAI Gym [\[code\]](#)
- **Recommender System:** Built a movie recommender system using collaborative filtering (with and without baselines), Singular Value Decomposition, CUR and the Latent Factor model for the course *Information Retrieval* [\[code\]](#)

## PUBLICATIONS AND PRESENTATIONS

---

- **Nallapareddy, Vamsi** et al. "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*. 2021; 89: 745–761. doi: [10.1002/prot.26056](#)
- 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" in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Seoul, Korea (South), pp. 80-87. doi: [10.1109/BIBM49941.2020.9313185](#)
- **Nallapareddy, M. V.**, Bhat, S., and Rajesh, V. (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". Proceedings of *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, M. V.** and Dwivedula, R. (2020) "ABLE: Attention Based Learning for Enzyme classification". Under review at *The Journal of Computational Biology and Chemistry* [\[preprint\]](#)
- **Nallapareddy, M. V.**, Michael, H., Ian, S., Bordin, N., Maria, L., Orenge, C.A., Rost, B. "Detection of remote homologues for CATH superfamilies using deep learning". Manuscript in preparation

## AWARDS & HONORS

---

- Max Planck Institute for Informatics 2020 Research Scholar Fellowship
- **National Talent Search Examination (NTSE) Scholar:** Recipient of the prestigious merit scholarship provided by the Government of India
- **BITS Pilani Merit Scholarship:** Awarded for being in the top 1% of my class
- **Summer Schools & Conferences:** Google AI India Machine Learning Summer School 2020, HSE University Summer School on Machine Learning in Bioinformatics 2021, [ISCB ISMB/ECCB 2021](#)
- **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
- **Coursera:** [Introduction to the Biology of Cancer](#) by Johns Hopkins University, [Genomics: Decoding the Universal Language of Life](#) by University of Illinois at Urbana-Champaign, [AI for Medical Diagnosis](#) by deeplearning.ai, [Fundamentals of Reinforcement Learning & Sample-based Learning Methods](#) by University of Alberta