

Pushkar Ambastha

CONTACT INFORMATION	IIT Guwahati Guwahati, India	Portfolio GitHub LinkedIn p.ambastha@iitg.ac.in
EDUCATIONAL QUALIFICATIONS	Indian Institute of Technology (IIT) Guwahati , Guwahati, India (Nov '21 - Present) B.Tech - Bio Science and BioEngineering CGPA: 8.08/10 , Batch Rank: 4th/75 Vidya Sanskar School , Patna, India (June '19 - Mar '21) Higher secondary (12th grade), CBSE Board Percentage: 92.2% , School Rank: 2nd/200 Delhi Public School Patna , Patna, India (June '18 - Mar '19) Secondary school (10th grade), CBSE Board Percentage: 95.2% , School Rank: 7th/420	
INTERNATIONAL RESEARCH INTERNSHIPS	Calibrating Agent-Based Models for Tumor-Immune Interactions using Spatial Biopsy Data and Multi-Modal Pipelines with AgentTorch (June '23 - Present) Guide: Ayush Chopra (<i>Camera Culture Group, MIT Media Labs, MIT</i>) Massachusetts Institute of Technology (MIT) Media Labs <ul style="list-style-type: none">Developed methods to calibrate clinical Agent-Based Models (ABM) directly from biopsies to have a mean accuracy of 77% under the Spatial Agreement Measure (<i>SAM</i>) Metric, minimizing the number of biopsy samples taken.Designed a novel multi-modal calibrated ABM pipeline to apply gradient-based ABMs to simulate tumor-immune cell interactions. (<i>for Cytotoxic CD8+ T Cells in multiple carcinoma cases</i>)Extended a novel pip library <i>AgentTorch</i> (Pytorch framework to design, simulate, and optimize agent-based models) for the cancer immunotherapy domain. Extracting the parameters from the classification model with an accuracy of 90%. Optimizing Medical Segmentation with Integrated Med-SAM and Fast-SAM Models for Enhanced Accuracy in Multi-Modal Imaging (May '23 - Dec '23) Guide: Katie Link (<i>Hugging Face X Healthcare</i>), Hugging Face, NY Report <ul style="list-style-type: none">Developed novel models derived from the cumulative performance and extrapolation of Segment Anything Model (<i>SAM</i>), Medical SAM (<i>Med-SAM</i>), and Fast-SAM.Created a streamlined process that reduced the time it takes to analyze images by 68% and decreased the model size by 82% compared to the Vanilla SAM model.The results observed in Pathology and CT modalities gave an average improvement of 0.48 in mean Intersection of Union (<i>mIOU</i>) and 0.42 in Dice Score Coefficient (<i>DSC</i>). Adaptive Biomedical Segmentation: Enhancing model Explainability through Domain Shift Analysis (Nov. '22 - Feb. '23) Guide: Tushar Kataria (<i>Computer Vision Group</i>), University of Utah Report <ul style="list-style-type: none">Analysed Domain Shift in biomedical image segmentation models as a critical insight into Model Explainability. (for both binary and multiclass semantic segmentation instances)Fine-tuned U-Net, DeepLabV3 model on Dataset like GlaS from MICCAI (2015), CRAG, CPM15 to observe domain dependency of models on the dataset, created a pipeline to improve Image masks mIOU and Dice Score.The results achieved for metrics like mIOU Loss, Pixel Accuracy, Jaccard Loss, and Dice Score met the standards of prior field experiments, like 0.96 for CRAG and 0.89 for GLAS (<i>mIOU</i>).	
SCHOLASTIC ACHIEVEMENTS	<ul style="list-style-type: none">Inter IIT Tech Meet 11.0 : Secured Gold Medal in building QA Chatbot by <i>DevRev.ai</i>, 2023Kaggle 4X Expert: Reached the Expert tier in all the categories of Kaggle, 2023Bronze Medal (85th/1100): Open Problems – Single-Cell Perturbations, Kaggle, 2023Bronze Medal (99th/1025): HuBMAP – Hacking the Human Vasculature, Kaggle, 2023IITG.ai Hackathon: Secured 6th/25 rank in hackathon conducted by IITG.ai club, 2022Convolve Hackathon: Secured 28th/231 rank in hackathon among 6 IITs by <i>Cisco</i>, 2022All India Rank 140 out of 100,000 candidates in National Entrance Screening Test (NEST), 2021Qualified for the Pre-Regional Mathematics Olympiad in PRMO, 2019	
COMPUTATIONAL SKILLS	Programming languages : C, C++, Python, R; MATLAB Frameworks : PyTorch, Tensorflow, Keras Machine Learning/Data Analysis : Deep Learning, including CNNs, RNNs, VAEs, GANs, Bayesian NNs, Transformers	

KEY PROJECTS

Classifying SARS-CoV-2 genomes using advanced NLP

(July '23 - Nov '23)

Course: Bio-informatics

[Poster](#)

- Developed DNABERT, a pre-trained bidirectional encoder that achieved state-of-the-art performance in predicting promoters, splice sites, and transcription factor binding sites with a **5%** increase in F1-score over traditional models, using minimal fine-tuning on small datasets.
- Implemented a [transformer-based pipeline](#) with **99%** accuracy for SARS-CoV-2 genome classification, outperforming traditional random forest and SVC models by **40%** in precision.
- Enhanced viral genome classification by integrating model probabilities, resulting in a **12%** improvement in the correct assignment of cDNA sequences to distinct viral classifications.
- Presented a poster at Conference Hall, IITG and achieved **1st/23** presenting groups

ProteoSynth - Automated Custom Protein Sequence Generator

(Jan '24 - May '24)

Course: Computational Biology

[GitHub](#)

- Developed Flask-based app generating 10,000 custom proteins with random amino acid sequences.
- Implemented custom options for sequence length, amino acid exclusion, and protein quantity, allowing users to generate up to 100 different protein sequences in a single request.
- Created functionality to output multiple .pdb files with sequences with downloadable .zip files.

GrooveSynth - Protein Active Site Structure Generator

(Jan '24 - May '24)

Course: Computational Biology

[GitHub](#)

- Developed a Flask-based app to analyze and visualize protein active binding sites, achieving a **20%** increase in the accuracy of ligand-binding predictions as a continuation of **ProteoSynth**.
- Developed an algorithm to identify and connect key amino acids in protein structures, enhancing understanding of ligand-binding mechanisms.
- Created a novel system to generate simplified Protein Data Bank (PDB) structures, reducing analysis time by **30%** and aiding drug design efforts.
- Secured **1st/75** position in the course with **AS** (Outstanding) grade for the project presentation.

PRODUCT DEVELOPMENT PROJECTS

Domain-specific Question Answering chatbot

(Dec. '22 - Feb. '23)

11th Inter IIT TechMeet, IIT Kanpur

[GitHub](#)

- Developed a question-answering system by retrieving the top candidate sentences from the corpus, used up to free collab resources, and *quantized* models to keep *latency* less than *1000 ms*.
- Developed question-answering pipeline using techniques like model distillation, sparsification, pruning, and fine-tuning the *DebertaV3-Base* model to decrease inference time and have a minimum loss in accuracy.

Cover Generation using all the OpenAI tools

(Jan. '23 - March '23)

IITG.ai Club, IIT Guwahati

[GitHub](#)

- Developed a [multi-modal pipeline](#) that converts audio/text input into images. Generated optimal transcripts for the podcasts and songs with *OpenAI Whisper* to create prompts.
- Designed pipeline with Latent Diffusion Models (*DALL-E*) to generate aesthetic cover images from created prompts using *ChatGPT/GPT-2* models, and deployed this on [Hugging Face Spaces](#).

Super Resolution Photographic Mosaic

(March '23 - April '23)

Coding Club, IIT Guwahati

[GitHub](#)

- Developed a Computer Vision pipeline that enhances images by super-resolution and stitching.
- Designed multi-model pipeline consisting mainly of Latent Diffusion Upscaler model for super-resolution and Image Stitcher for creating a panorama.

Re-colorisation of monochrome images using conditional GANs

(Aug. '22 - Sep. '22)

Coding Club, IIT Guwahati

[GitHub](#)

- Trained a conditional GAN model based on *U-Net* block with *Resnet18* backbone and devised Image Processing strategies for colorization of monochrome images.
- Deployed a [web app](#) using Streamlit library on HuggingFace for *COCO* dataset fine-tuned model.

RELEVANT COURSE WORK

Bioengineering (Computational):- Bioinformatics • Biological Data Analysis • Biophysics • Computational Biology • Systems Biology¹ • Research Methodology¹

Computer Science:- Data Structures and Algorithms • Computing Lab • Pattern Recognition and Machine Learning

Mathematics:- Linear Algebra • Basic Calculus • Complex Analysis

Bioengineering (Wet laboratory):- Genetic Engineering • Biochemistry • Microbiology • Cell and Molecular Biology • Analytical Biotechnology • Bioseparation Engineering • Immunology

MOOCs:- Machine Learning² • Deep Learning² • Computer Vision² • Natural Language Processing (NLP)² • Reinforcement Learning³

¹ Ongoing course ² Course undertaken from Coursera DL Spec. ³ Course undertaken from Kaggle