Pratik Katte

pratikkatte7@gmail.in, pkatte@ucsc.edu

Portfolio ★ LinkedIn ★ Github

WORK EXPERIENCE

Corbett-Detig Lab, Santa Cruz, CA ***Oct 2023 – Present***

Graduate Student Researcher | ***Phylogenetics, Reinforcement Learning, Generative Modeling***

- Tree-sequence Analysis Using LLMs – Developing an interactive framework combining large language models (LLMs) with the tskit library and a Taxonium-inspired visualization to simplify tree-sequence analysis and broaden access to evolutionary insights. [github]
- Dashboard for Wastewater Pathogen Surveillance – Built a real-time dashboard (with Turakhia Lab, UCSD) integrating WEPP outputs for SARS-CoV-2 lineage tracking from wastewater. Incorporated Taxonium and JBrowse for genome-level visualization. [github]

Niramai Health Analytix, Bangalore, India ***Jan 2020 – Jun 2023***

Senior Research Engineer | ***Thermal Imaging, Breast Cancer Diagnosis, Chest-X ray Analysis for COVID-19***

- Led the development of machine learning models within a desktop application (NTCT) aimed at assisting technicians in the efficient capture of breast thermal images for cancer screening purposes. This application is currently deployed in over 100 hospitals worldwide.
- Played a lead role in developing an AI-based desktop application designed to screen individuals for COVID-19 symptoms, including elevated temperature and shortness of breath, while enforcing adherence to COVID-19 protocols. Successfully screened over 1 million individuals across India.
- Trained U-Net based deep learning model to perform lung segmentation on WhatsApp-compressed chest X-ray images within the Xray-Setu project. Notably, this effort resulted in a substantial 6% enhancement in the AUC score of the lung abnormality classifier.
- Awards: “On Spot Award” and “Working beyond boundaries Award” for the performance on Niramai Fever Test product.

Rewind, Mumbai, India ***Sep 2018 – Nov 2019***

Co-Founder | ***A failed attempt in building an ambitious venture to solve the waste problem.***

- Co-founded a venture Rewind that capitalized on established waste pickers, paper and metal shops, and recyclers network to create a platform enabling the sale of diverse scraps, including electronics and furniture.
- Operated in stealth mode for one year, collaborating with 10 waste pickers and 2 recyclers to collect electronics and furniture totaling $1000 in value.

L.V. Prasad Eye Institute, Hyderabad, India ***Jun 2018 – Aug 2018***

Research Intern | ***Conversational Chatbot, Decision Tree, Image based Ocular Evaluation***

- Designed and developed a conversational chatbot using a decision tree algorithm to streamline booking an appointment with a doctor and educate patients about eye diseases. Deployed on a website serving more than 4500 visitors per day.
- Developed an algorithm to quantify the health of an eye using fuzzy c-means clustering algorithm.

Prakshep, Bangalore, India ***Jun 2017 – Jan 2018***

Data Science Intern | ***Geospatial Satellite Imagery, Spatial Modelling, Geographic Information system.***

- Explored unsupervised machine learning techniques such as Gaussian Mixture Model, along with employing deep learning models like U-Net, for the segmentation of forested and deforested areas within satellite imagery.
- Applied spatial modeling techniques by developing a Geographic Information System (GIS) using GeoServer and the GDAL Python library. This system efficiently processed geospatial data related to crop harvests, enabling farmers to engage in spatial modeling for informed decision-making regarding seed selection, ultimately contributing to improved and optimized yields.

EDUCATION

University of California Santa Cruz, California, USA ***Sep 2023***

PhD student in Biomolecular Engineering and Bioinformatics

Relevant Coursework: Mathematics of the mind (BME-218), Computational Genomics and Systems Biology.

University of Mumbai, Mumbai, India ***Jun 2015***

Bachelor in Engineering, Information Technology

Relevant Coursework: Intelligent System, Image Processing, Computer Graphics and Virtual Reality, Big Data Analytics

PUBLICATIONS

Katte, Pratik, et al. "Automated thermal screening for covid-19 using machine learning." MICCAI Workshop on Medical Image Assisted Blomarkers' Discovery. Cham: Springer Nature Switzerland, 2022.

Sabyasachi S., **Pratik K.,** et. al. 'Abstract PS2-44: Diagnosing COVID-19 From Images of Chest X-rays Communicated Via WhatsApp.' UKIO Congress (2022).

Patil, Vivek, **Pratik Katte**, and Abhay Patil. 'Restoration of Images Using Only Noisy Data.' International Journal of Research and Analytical Reviews (IJRAR) 6.1 (2019).

CONFERENCE TALKS

- 'Niramai Fever Test: Automated Screening for COVID Symptoms', Wolfram Technology Conference – 2021 [Link]
- 'Machine learning for COVID-19 detection', Data Science Conference, Europe – 2021 [Link]

PROJECTS

Uncertainty Estimation of a Phylogenetic Tree [Link]

- Exploring the use of cutting-edge Generative Flow Networks (GFlowNets) to estimate uncertainty in phylogenetic trees. This project aims to address the limitations of traditional methods by learning the complex distribution of evolutionary relationships and providing probabilities for all possible outcomes in a phylogenetic tree.
- Currently working on initial proof-of-concept implementation of the GFN for phylogenetic tree reconstruction and understanding its efficacy for uncertainty estimation tasks.

Biofilm Regulation by small RNA in Vibrio cholerae - Gene Expression Analysis [Link]

- Vibrio cholerae is responsible for cholera, the role of small RNAs(sRNAs) remains inadequately understood. Investigating the role of small RNA molecules in communicating with other colonies holds promise for gaining a deeper understanding of the root causes of infectious diseases caused by human pathogens.
- Using the OTTR seq dataset of Vibrio Cholerae and tRAX software tool, we perform gene expression analysis between the stationary and biofilm state.

StructHunt [Github]

- As a part of a hackathon at UCSF, we created a tool designed to track the publication of new research papers detailing integrative biomolecular structures in bioRxiv and medRxiv. This tool enables us to swiftly capture and incorporate this valuable new data into the RCSB Protein Data Bank.
- I was responsible for the implementation of context retrieval from LLM embeddings, utilizing the FAISS library for efficient indexing and similarity searches.

TECHNICAL SKILLS

- Bioinformatics Tools: UCSC Genome Browser, BLAST, DESeq2, Bowtie, HISAT, Samtools.
- Programming Language: Python, C#, C++, R.
- Frameworks/Libraries/Services: TensorFlow, PyTorch, PyTorch Geometric, Keras , scikit-learn, ReactJS, NodeJS, AWS.