





Abhishek Agarwal

+91 7453046635 • abhiag.719@gmail.com

 abhiag719
 ultimateabhi719

 medium.com/@ultimateabhi
 scholar.google.com

Professional Experience

2046 LLC

Chief Science Officer

Hyderabad, India

2023–present

- Led a team of three people managing schedules, performance reviews and daily tasks reporting
- Led the development and deployment of three new bioinformatics pipelines, making processing of bio-assays simple and user-friendly. The pipelines process TeraByte sized biological assays on client facing servers
- Responsible for integration of LLMs and Deep Learning models into pipelines

Almonk Technologies

Senior Machine Learning Consultant

Bengaluru, India

2022–2023

- Image to Latex Optical Character Recognition:
 - Built a novel transformers based OCR system for mathematical equations in PyTorch
 - Achieved 87% BLEU score on image data with noisy background
 - Researched and designed improvements to reduce inference time of the NeuralNet by a factor of 2-10

The Jackson Laboratory

Postdoctoral Associate

Massachusetts, USA

2020–2022

- Attention Mechanism for Biological Assays:
 - Lead the effort to build a nextflow pipeline to create the first dataset of matching mouse and Human DNA features (enhancers) using Deep Learning.
 - Achieved 83% accuracy in predicting enhancer-promoter links. The model used LSTM attention and 1-D convolution to predict expensive ground truth assay outputs from inexpensive assays like ATAC-seq
- Deep Learning Model for COVID:
 - Designed the architecture for using 2D CNN model on patient CT scan images
 - Reduced the data requirement for CT scans by a factor of 5, by using transfer learning from 2D CNNs to 3D models
 - Reduced false positives by 20% using the 3D architecture

University of Illinois, Urbana-Champaign

Postdoctoral Associate

Illinois, USA

2018–2019

- Dimensionality Reduction for Gene Expression Data:
 - Invented an online matrix factorization algorithm, with proven stochastic convergence guarantees, for computing basis elements representative of the underlying gene expression signature
 - Published in NeurIPS 2019, the algorithm reduced computation requirements by a factor of 100 while speeding up convergence

Education

University of Minnesota, Minneapolis

Ph.D, Electrical and Computer Engineering

Thesis: Data Estimation and Recovery for Distributed Storage Systems

USA

2013–2018

Indian Institute of Technology, Kanpur

B.Tech-M.Tech. Dual Degree, Electrical Engineering

India

Projects

- **Neural Machine Translation (NMT) for English to German**
 - Implemented the “Attention is all you need” paper from scratch in Pytorch
 - Also implemented EncoderDecoder Model using transfer learning from BERT LLMs
 - Improved performance of NMT to a BLEU score > 0.2 by adding weighted entropy loss function and NER (named entity recognition) block
 - Deployed the LLM EncoderDecoder model using bentoml
- **Face Recognition**
 - Developed a Siamese Network for Face-Recognition using transfer learning on ResNet-50
 - Trained the model on multiple GPUs using DDP (Distributed Data Parallel)
 - Achieved 90% accuracy on the ATT test dataset
- **Land Usage Pattern Recognition**
Omdena Douala Chapter, Volunteer Project
 - Identified and Scraped datasets for land usage and land cover for the West African landscape
 - Developed and Deployed a U-Net model for aerial image segmentation achieving a pixel accuracy of 77% for 10 classes
 - Worked with a global team of 20 people to lead the project to completion

Skills and Achievements

- Solved a **75 year** old open problem in statistical theory (group testing) during PhD
- **Key Skills:** Programming, Algorithms, Mathematics, Data Science, Machine Learning, Deep Learning, Kaggle, Statistics and Probability, Research, Computer Vision, Natural Language Processing (NLP), Sentiment Analysis, Recommender Systems
- **Technologies:** Matlab, R Programming, SQL, git version control, Azure, AWS, GCP, Sagemaker, Docker, Azure Data Factory, Distributed Training, CI/CD, Distributed Computing, Data Products, Linux, Bash Shell Scripting
- **ML Tools & Techniques:** Pytorch, Tensorflow, Pytorch Lightning, Pandas, Numpy, Matplotlib, Seaborn, scikit-learn, NLTK, MLOPS, GPT, Llama2, supervised learning - decision trees, random forests, ensemble methods - bagging and boosting, unsupervised learning - k-means clustering
- **Volunteering:** Isha Foundation Save Soil Media Team (2022-)