Abhishek Agarwal

☐ +91 7453046635 • ✉ abhiag.719@gmail.com

in abhiag719   M medium.com/@ultimateabhi
○ ultimateabhi719   🔖 kaggle.com/curioushomosapien

Professional Experience

2046 LLC   **Hyderabad, India**
Chief Science Officer   *March 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

AImonk Technologies   **Bengaluru, India**
Senior Machine Learning Consultant   *Aug 2022–Feb 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 using linearized attention

The Jackson Laboratory   **Massachusetts, USA**
Postdoctoral Associate   *Jan 2020–June 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   **Illinois, USA**
Postdoctoral Associate   *June 2018–Dec 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   **USA**
Ph.D, Electrical and Computer Engineering   *June 2013–June 2018*
Thesis: Data Estimation and Recovery for Distributed Storage Systems

Indian Institute of Technology, Kanpur   **India**
B.Tech-M.Tech. Dual Degree, Electrical Engineering

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
- **Sentiment Analysis using BERT**
  - Developed a Sentiment Score prediction model on joint text and ratings data combining BERT with tabular features
  - The model combines several text inputs with ratings score to achieve 97% test set accuracy
  - With an initial basline accuracy of 40% our model provides a 140% improvement in accuracy
- **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 AT&T 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

Selected Publications

- A Agarwal, J Peng, O Milenkovic, "Online Convex Matrix Factorization with Representative Regions", NeurIPS 2019
- A Agarwal, S Jaggi, A Mazumdar, "Novel Impossibility Results for Group-Testing," ISIT 2018

For a full list see google scholar.

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, Extractive Question Answering
- Technologies: Matlab, R Programming, SQL, git version control, Azure, AWS, GCP, Sagemaker, Docker, Kubernetes, Distributed Training, Linux, R Programming Bash Shell Scripting, Kaggle
- ML Tools & Techniques: Pytorch, Tensorflow, Pytorch Lightning, Pandas, Numpy, Matplotlib, Seaborn, scikit-learn, NLTK, GPT, Llama2

- Volunteering: Isha Foundation Save Soil Media Team (2022-)