Haider Zainuddin Ali

Hali4@mail.yu.edu | (551) 344-4769

alihaider20.github.io/ | https://www.linkedin.com/in/haiderzali/ | GitHub/AliHaider20

PROFESSIONAL EXPERIENCE

S&P Global | Machine Learning Joint Research Project *| Deep Learning | Geospatial Analysis*   **Remote, USA, May 2023 – Aug 2023**
- Scraped, engineered, and extracted relevant features of Global Precipitation Data from NASA's GDDP (HDF) files using Python and R
- Applied image analysis, removed watermarks, and analyzed and imputed NAN values using KNN.
- Applied progressive model building approach using models KMeans, **SRResnet**, SR-GAN, and **ESRGAN**, using **PSNR** metric
- Identified **ESRGAN** as the best model for preserving **60% image** quality when scaling NEX test data and monitored progress with MLFlow.

Kan Innovations | Deep Learning Developer *| Computer Vision | Medical Imaging*   **Mumbai, India, Jan 2022 – July 2022**
- Generated a 5000-image dataset by scraping foot images using Google image scraper, Albumentations library, and Pix2Pix GAN
- Captured images from Intel's IntelliSense (depth sensor) camera, maintaining **intrinsic and extrinsic calibrations**
- Constructed **3D voxel** foot meshes using **Open3D** library and **Blender** to help design insoles, reducing **40%** patient's foot deformity
- Implemented BlazePose model for real-time **pose estimation**, enhancing gaming interaction and reducing **20%** patient's foot deformity

Indian Servers | Deep Learning Intern *| Computer Vision | Leadership*   **Mumbai, India, July 2020 – Aug 2020**
- Mentored and worked with a team of three members on detecting Diabetes using Retina images, with augmentation techniques.
- Optimized custom model by tuning hyperparameters to achieve **73% F1 score**, also achieved **85% F1 score** using **Efficient Net**.

EDUCATION

Master of Science in Artificial Intelligence; GPA: 3.5   **May 2024**
Katz School of Science and Health at Yeshiva University, New York, United States
Coursework; Numerical Analysis, Computer Vision, Deep Learning, Statistics and Probability

SKILLS & CERTIFICATIONS

- **Programming and Tools:** Python, R, ChatGPT API, NLTK, Git, AWS, PySpark, PyTorch, OpenCV, LLM, Generative AI, MLFlow, Open3D
- **Data Science Techniques:** 3D Image processing, Camera calibration, Machine learning, Generative AI, Computer Vision, MLOps
- **Machine Learning:** XGBoost, SVM, LightGBM, PCA, t-SNE, DBSCAN, LSTM, VAE, VIT, MAE, AdaBoost, Transformers, GANs
- **Certifications: LLMOps, Text Embeddings (Deep Learning.AI), Data Science Master's program (GreyAtom)**

RELEVANT PROJECTS

University Chatbot using LLM *[github] | Large Language Models | Data Engineering*   **Sept 2023**
- Scraped, augmented, and created 60000 question-answer pairs manually and using the **NLPAug** tool, and **BERT**
- Employed **LlamaIndex** and **Langchain** with **FAISS** vector database to build a **RAG** pipeline
- Compressed Mistral Instruct model with DeepSparse for faster inference, applied Prompt Engineering for optimal responses
- Fine-tuned and merged including Microsoft's **Phi 1.5B**, Google's **Flan T5**, **GPT 2**, and **Mistral Instruct** achieving **70%** RougeLSum score

Cow teat Keratosis level identification *[github] | Data Augmentation | Generative AI*   **May 2023**
- Applied Image processing, extracted features using SURF, SIFT, HOG, and **AutoEncoders** leading to a 50% F1 score using SVC, and KNN.
- Balanced data using DreamBooth **Stable Diffusion** and achieved an **80% F1 score** using **GoogleNet**. Paper available here.

Cow stall number detector *[github] | Computer Vision | Object Detection*   **Feb 2023**
- Achieved better detections of Stall numbers by applying image processing techniques with multi-Otsu thresholding to augment images.
- Developed a custom CNN and loss which increased the accuracy by 30% with ResNet34 base, reaching an overall **80% IOU** score.
- Employed **MLflow** for model tracking and utilized **PyTorch Lightning** for efficient code development and deployment.

Flood Area Image Segmentation *[github] | Transfer Learning | Image Processing*   **Oct 2022**
- Utilized UNet, ResNext50UNet, and **DeepLabV3+** models to segment flood-affected areas using a Kaggle dataset.
- Achieved **95% IOU score** using distributed training resulted in building better Urban Planning, and Humanitarian Aid in affected areas.

Protein Muti-Label Image Classification – Jovian *[github] | Medical Imaging |*   **June 2022**
- Classified multiple proteins present in a single-cell image using the Human Protein Atlas dataset from Kaggle.
- Achieved **80% F1 score** in classifying ten types of proteins using the **ResNet-18** backbone model on single-cell images.

ACHIEVEMENTS

- Won 2 Python hackathons with **700** participants each and a Data Science competition at **GreyAtom** with **100** participants