Omer Raza

Machine Learning & Software Engineer

Email: raza12@purdue.edu | Website: https://www.mdotomer.com

Professional Summary

Machine Learning Engineer with 5+ years of experience in industrial ML applications and research, including LLMs, GANs, GNNs, and XGBoost. Skilled in **MLOps** and cloud services such as AWS with a robust **Software Engineering** background in web, mobile, and app development using React.

Skills & Technologies

- **Machine Learning**: GAN, Diffusion, LLM, XGBoost, GNN, CNN, Computer Vision
- **Deep Learning Frameworks**: PyTorch, TensorFlow
- **Cloud Computing**: AWS
- **Web & Mobile App Development**: React, Electron
- **Programming Languages**: Python, JS, C++, Java
- **Database**: SQL (intermediate), Pandas
- **Data & Image Processing**: Sklearn transform, OpenCV
- **Optimization**: Linear Programming, 2nd order Newtonian, Neural network hyperplane optimization

Experience

RA Machine Learning role at **CIVS** on **Multimillion dollar grant steelmaking projects** **Sep 2023 - present**

- **Silicon content prediction via ML** in blast furnace – **Paper accepted in MDPI Materials**
  - **Comprehensive** and **adaptable** pipeline (data processing and modeling) to predict silicon % in iron
  - **90% Accuracy - Beta** testing in **partner steel plant**.
  - **Energy efficiency** due to foresight of silicon trends
- **Hearth erosion modeling – Paper underway**
  - Designed and quickened **first principle** models (**forward heat transfer process**)
  - Formulated and Optimized the inverse process for calculating refractory profile
  - **Graph Neural Network** and **ROMs** further fasten optimization from **3-5 minutes to 20 seconds**
  - Can **improve safety warnings** and lengthen **campaign life – beta testing soon**

RA Machine Learning (remote) at **CUHK** and **HKUST** universities **Sep 2022 - Mar 2023 & Mar - Jul 2024**

- **First author** and **co-author** of multiple papers (details in publications)

Machine Learning Engineer at **Lalamove (Multinational) - Hong Kong** **Feb 2022 - May 2023**

- Improved **image localization & detection models** for **accurate market penetration count**
  - within **10% human count** or **90% accuracy**
  - more than **70,000 USD savings** annually
- Developed **image detection models, OCR & clustering models**. This facilitated:
  - **Prompt approval** of **bonus remuneration** to commissioned drivers
  - **Minimal staff intervention** and **associated costs.**
- Established data pipeline, modeling and deployment flow for **risk assessment models** to **predict fraud**
  - **Prototype** established with **0.95 F1 score** and **enhances** existing **rule-based model**
  - Handles **tens of thousands of transactions** per day
  - Blocks an **additional 20%** of **problematic transactions**

Received the highest bonus in the year (4 people among 150 in the office)

Machine Learning & Software Engineer at **Gense Technologies (Startup) - Hong Kong** **Nov 2020 – Dec 2021**

- Designed **LSTM & Conv1D** models for **guided breathing waveform pattern classification**.
  - **Auto Prompts** user for a retest on the spirometer for **better final diagnosis at User App side**
- Trained & Improved **EIT Amplitude Image Classifier** by engineering two distinct iterative improvement approaches
  - **Auto Prompts** user for an **EIT retest** to **prevent diagnosis based on corrupted EIT signals**
- Developed & Deployed **Gense Mobile App on PlayStore (React native)**
  - Manager **user accounts,** live test and **data sync with server** & **visualize data** and statistics

Software Engineer (Remote Commission) at **BeardBee - Hong Kong** **Feb 2020 - Mar 2020**

- Produced **maintainable** and **portable single code base** for **web, desktop** and **mobile** platform of a charging system

Publications

FIRST: Efficient Trustworthy Distillation Paper – Co-Author **Accepted in EMNLP 2024**

- Pipeline for efficient LLM training for better **calibration** and **accuracy**.
- **Distillation** with **trustworthy maximization process** whilst using only a **fraction of training data**
- **2.3% Accuracy gain & 10 % miscalibration reduced** compared with other distillation method

SR TGAN: Temporal Smoke Removal paper – Equal Contribution First Co-Author **Accepted in IEEE EMBS BHI 2024**

- Improved smoke removal by **incorporating sequential information** from temporal frames
- Separate **attention, convolution & LSTM mechanisms in CycleGAN** base skeleton
- **Better smoke removal** than **benchmark DeSmoke CycleGAN** model based on the **JNBM & FADE metrics**.

Mutual Information for EIT to CT Aligned Transformation – First Author **Accepted in IEEE EMBC July 2023**

- Researched during tenure at **Gense Technologies**
- **Modified CycleGAN** arch to convert low resolution time based EIT images to high resolution **structurally aligned** CT images
- **Normalized Mutual Information (NMI)** gain from **0.2600 to 0.2621, (p<0.0001)** from vanilla approach

Silicon Content Prediction in Blast Furnace via ML **– First Author** **Accepted in MDPI Materials 2025**

- Comprehensive & generalized **data processing pipeline** for large scale industrial data
    - **Adaptive** and **comprehensive** formatting, clean up, EDA (semi-auto) to process variable adaptive anomaly detection, multicollinearity check, target leakage, bias, lag correlation, feature engineering and standardization
- **Hyperparamater optimization** & **model selection** to forecast silicon content with **90 % accuracy**
- Pipeline yields an improved accuracy of **at least 5%** compared to typical standard processing and clean up

Mutual Information prior enhances EIT reconstruction – First Author **(Preprint - Results Pending: IEEE JBHI)**

- An **extension journal paper** of IEEE EMBC conf paper
- Incorporates modified Cyclegan to infer prior in upstream pipeline for **improved reconstruction**
- **NMI gain from 0.45 to 0.49** from vanilla approach

Surgical Blender journal paper – Second-Author **(Results Pending: Journal CIBM)**

- Explored various data synthesis of surgical images to **improve downstream tasks** such as training models to segment organs and surgical tools in unknown surgical areas
- Improved **JNBM and FADE** score on transformative tasks and improved **Dice score** on segmentation
- Potential to alleviate the **time** and **cost burden** for **acquiring real surgical images** in training tasks like smoke removal

Education & Accolades

Masters in Computer Science from **Purdue University (4.0 GPA)** **2023-2025**

Bachelors in **Engineering (Computer Science)** from **The University of Hong Kong (HKU)** **2016-2020**

- **First Class Honors** & **Deans' Honor List** in **2016-2017 & 2019-2020**
- **HKU Foundation Scholarship 2016-2020** (for **outstanding** undergrads)
- **Young Tsun Dart Scholarship 2017** (reserved for **only one student** in a particular year of study)

Certifications – Machine Learning

Machine Learning Specialization from **Coursera - Certificate** **Sep-Dec 2016, Nov 2017 & Dec 2019**

- 5-part specialization with Certification

References

Available upon request. Unpublished work subject to consent of other authors

- **Google Scholar Profile:** https://scholar.google.com/citations?user=oPbPzGgAAAAJ&hl=en
- **Coursera Machine Learning Certificate:** https://coursera.org/share/ae5cbf8d757883c352ba6933527bbded
- **Github:** https://github.com/MdotO?tab=repositories