Omer Raza

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

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)

Research & Employment Experience

RA Machine Learning role at renowned **CIVS** in **Purdue University** **Sep 2023 - present**

- **Silicon content prediction via ML** in blast furnace – **Paper accepted in MDPI Materials**
    - A general yet **comprehensive** and **adaptable** data processing and modeling pipeline for industrial processes
    - **90% Accuracy & Beta testing** in **partner steel plant** soon.
    - **Energy efficiency** because of good silicon predictability implies immense economic impact.
- **Hearth erosion modeling – Paper underway**
    - Created **first principle** models (**forward heat transfer process**) and engineered optimization flow (**inverse optimization process)** for calculating **profile of refractory bricks in the hearth** of a blast furnace
    - **ML state models and ROMs** used fasten computation from **5-10 minutes per 1 forward iteration to 10 seconds**
    - Can improve safety warnings for hearth wear and help **productivity of blast furnace**

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

- **First author** and **co-author** of 3 papers (details in publications below)

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** for real-time license plate number detection for subscription autorenewal. 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**
    - Expected to handle **tens of thousands of transactions** per day
    - **Enhanced** existing **rules-based model**

  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**

- Developed LSTM and Conv1D models for **guided breathing waveform pattern classification**.
    - **Auto Prompts** user for a retest on the spirometer for **better final diagnosis**
- Developed & deployed **EIT Amplitude Image Classifier** using 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 (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

Tech Intern (Software Engineer) at **Lalamove - Hong Kong** **Jun 2019 - Aug 2019**

- **Refactored** & **debugged** a **microservice** for order grouping and route formation
- Deployed to India & Thailand

Publications & Preprint

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 Conference paper – 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
- Results, **JNBM & FADE metrics**, show that temporal incorporation helps **achieve better smoke removal** in live surgical videos

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 special edition 2024**

- 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 available & 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 – Co-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

Projects – Machine Learning

Image Feature Transform Project – **Personal Study** **2020**

- **Latent space transformation** encodings versus **direct image transformations** were studied and compared
- **Data scraping** – **automated comics panel downloads script** and **NN based face detection** to obtain plethora of expressions
- **Interesting insight - Latent space** transformations/interpolations provide **stabler generative images** than **direct transform**

3D Texture GAN Project – Academic FYP Bachelors **2020**

- **Final Year Bachelors Project** to produce **3D textured car meshes** based on 2D texture input for cars
- **Replicated "Learning to Generate Textures on 3D Meshes" paper** with limited success

Certifications

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

- 5 part specialization with Certification

Skills & Technologies

- Machine Learning: GAN, Diffusion, LLM, XGBoost, GNN
- Deep Learning Frameworks: PyTorch, TensorFlow
- Cloud Computing: AWS (Amazon Web Services)
- Web Development: React, Electron, JavaScript
- Programming Languages: Python, Javascript, C++, Java
- Database: SQL (intermediate)
- Data & Image Processing: Sklearn transforms & OpenCV
- Optimization: LP, 2nd order, SGD & Neural network hyperplane optimization

References

any other references or pending results docs are available upon request and 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