TALHA ARSHAD

San Jose CA · talha@utexas.edu · (512) 698-7349 · https://talha-arshad.github.io · https://www.linkedin.com/in/tarshad

SUMMARY A data scientist with broad expertise in machine learning and experience in research and industry. Skilled at owning projects end-to-end, technical leadership and effective collaboration. **Status:** Permanent Resident.

SKILLS

- **LANGUAGES:** Python, SQL.
- **ML/DL FRAMEWORKS & LIBRARIES:** PyTorch, TensorFlow (Keras), NumPy, XGBoost, Scikit-learn.
- **DATA & VISUALIZATION:** Pandas, Matplotlib, Seaborn.
- **CLOUD & MLOPS:** Google Cloud Platform (GCP), Weights & Biases (Wandb).
- **ENVIRONMENT & TOOLS:** Linux, Bash scripting, Git.
- **MODELING:** machine learning, deep learning, statistics.

EDUCATION

PHD CHEMICAL ENGINEERING, THE UNIVERSITY OF TEXAS AT AUSTIN **2016**
RESEARCH: Mathematical modeling and computational optimization for manufacturing nanostructured devices.
MENG CHEMICAL ENGINEERING, CAMBRIDGE UNIVERSITY **2010**
BA (HONS) CHEMICAL ENGINEERING, CAMBRIDGE UNIVERSITY **2009**

EXPERIENCE

STAFF DATA SCIENTIST, INTEL CORPORATION (SANTA CLARA CA) **2020-Present**
SENIOR DATA SCIENTIST, INTEL CORPORATION (SANTA CLARA CA) **2018-2020**
Modeling and optimization of deep ultraviolet (DUV) lithography for memory chips:

- Pushed models and code to production and managed these through the development lifecycle for 7 lithography steps across 3 generations of memory products.
- Led the definition, implementation and analysis of multi-die experiments to compare the performance of alternate design configurations and enable 30% denser memory.
- Led training and mentored engineers on lithography source-mask optimization (SMO).
- Devised and implemented corrective measures on a lithography step to reduce model error from 10% to 1% and improve throughput (dose) by 20%.

SENIOR DATA SCIENTIST, GLOBALFOUNDRIES (MALTA NY) **2016-2018**
Modeling and optimization of extreme ultraviolet (EUV) lithography for logic chips:

- Built data pipelines to transform and visualize datasets at various stages of the modeling workflow.
- Created tests and analyzed results to investigate tradeoffs between optimization runtime and performance.
- Designed a project to compare error statistics between alternate photomask vendors and mentored a graduate intern through its execution.

PROJECTS

DEFECT INSPECTION FOR SEMICONDUCTOR WAFERS WITH DEEP LEARNING **2022**
Built a convolutional neural network (CNN) image classification model to replace manual visual defect inspection in semiconductor manufacturing. Defined an architecture based on literature, implemented a hyperparameter search and evaluated error distributions to achieve a >90% recall and precision across all 8 defect types on the validation dataset, including when up to 4 defect types are present simultaneously.

FAILURE PREDICTION OF HARD DRIVES WITH GRADIENT-BOOSTED TREES **2021**
Trained gradient-boosted decision tree (XGBoost) models to predict the failure of hard drives in a data center up to a week in advance based on health and usage metrics. Experimented with feature-selection and grid-searched hyperparameter values resulting in an F1-score of 0.93 on the validation dataset.

PUBLICATIONS, COURSES, ACCREDITATION

- **PUBLICATIONS:** 3 papers in high-impact journals on applied modeling and optimization.
- **STANFORD CS231n**: Deep Learning for Computer Vision.
- **COLUMBIA X 4721**: Machine Learning.
- **COURSERA**: SQL for Data Science, Data Structures & Algorithms I & II.
- **GOOGLE:** TensorFlow Developer Certificate. **2022**