Sharjeel Mustafa
LinkedIn GitHub Portfolio
Email : mustafa7@uwindsor.ca

Technical Skills

Languages: Python, SQL, Java, C/C++, JavaScript
Frameworks: Pandas, NumPy, PyTorch, Flask, Matplotlib, SpaCy, JUnit, PyTest, Selenium, Scikit-Learn
Technologies: Git, Neo4j, GraphQL, REST, Linux, Docker, CUDA

Education

M.Sc. Computer Science, Artificial Intelligence 05/2025 – 12/2026
- *University of Windsor*, *Windsor, ON* GPA: 97.2/100

Awards: Vector Scholarship in AI (17,500), Ontario Graduate Scholarship (15,000)

B.Sc. Computer Science, Software Engineering 09/2020 – 08/2024
- *University of Windsor*, *Windsor, ON* GPA: 91.7/100

Awards: Gold LEAD Medal, Outstanding Scholars Medal, Rising Star Co-op Award

Experience

- **Graduate Researcher** – *University of Windsor, Windsor, ON* 05/2025 – Present
  - Designed **graph ML pipelines** for link prediction, transductive classification, and **graph representation learning** using structured and unstructured data with PyTorch Geometric
  - Investigated group-level interaction modeling on **biological data**, contributing to IEEE workshop paper [1]
  - Implemented node-, edge-, and group-level models using hypergraphs with **k-fold cross validation** to simplify prediction tasks under heterophily with potential applications to recommender systems
- **Research Assistant** – *University of Windsor, Windsor, ON* 01/2022 – 08/2024
  - **Data Engineering**: Developed a rule-based pipeline to extract text from over **30,000** semi-structured documents using Python, Pandas, and SpaCy, achieving over **90% median accuracy**.
  - **Automation**: Developed a program using Python, PyTesseract, REST APIs, and SpaCy to automate migration of 1000s of article metadata to online database, improving input time by **94%**
- **R&D Data Scientist** – *Swift Medical, Toronto, ON* 05/2023 – 12/2023
  - Wrangled and analyzed over **180 million** rows of time-series **healthcare data** using **SQL, Python, Pandas, AWS, Snowflake, and Matplotlib**, contributing to a research paper
  - Contributed to a multimillion-dollar initiative by creating a pipeline for Relation Extraction methods, such as **few-shot prompting LLMs for 12,000 clinical texts**, improving performance by **96%**
  - Validated models and extracted insights by creating ROC curves and analyzing trends in computed metrics like **AUC, F1, and balanced accuracy scores**, contributing to a paper
  - Handled end-to-end projects and presented analysis findings to a **cross-functional team** to aid with data-driven decision-making and prototyping of product features
- **Software Developer** – *Connecting With Technology, Windsor, ON* 05/2022 – 08/2022
  - Designed and developed a **Neo4j graph data model** for social networks, served by a **GraphQL API**, streamlining graph traversals compared to previous relational model and **REST API**
  - Developed an asynchronous **CRUD** backend with **JavaScript, Fastify, and GraphQL** following MVC architecture to operate efficiently on highly connected graph data

Projects

- **Cross-Script Transfer Learning** Python, PyTorch, CNN, CUDA, CV
  Built an **end-to-end pipeline** for **multilingual character recognition** using CNN, ResNet, and VGG models; applied **transfer learning** with script-aware layer **fine-tuning** to exploit cross-script similarity, enabling efficient **CUDA**-accelerated training and improved generalization
- **Hybrid Sentiment Analysis** Python, PyTorch, NLP, NLTK, RNN
  Developed a **multiclass sentiment analysis** pipeline, applying **model hybridization** over **unstructured twitter data**, using **word embeddings**, classifiers, and **RNNs**, with hybrid LSTM–CNN–SVM achieving **97.3% accuracy**

Research

1. M. Hashemi, **S. Mustafa**, A. Ngom, L. Rueda. "Heterophily-aware hypergraph neural networks for cell type prediction using ligand–receptor-informed single-cell RNA-seq data," in Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM), DLB2H Workshop, 2025. (accepted)

2. M. Hashemi, **S. Mustafa**, A. Ngom, L. Rueda. ""HeteroGraphNet: A ligand–receptor informed, heterophily-adapted graph neural network for cell type prediction in scRNA-seq data," in Proc. IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM), HP4MoDa Workshop, 2025. (accepted)

3. R. Fraser, M. Wynn, H. Mohammed, R. Gupta, D. Ramachandram, **S. Mustafa**, L. Goldstone, J. Rameriz-GarciaLuna and A. Cassata, "A retrospective analysis of deep tissue injury prevalence and incidence using a large-scale wound care database in long-term care settings across North America," presented at NPIAP National Pressure Injury Advisory Panel, Dallas, TX, USA, 2025, Poster.

4. M. Hashemi, **S. Mustafa**, A. Ngom, and L. Rueda, "HeteroGraphNet: Advancing cell type prediction in scRNA-seq with enhanced graph neural network for heterophilic structures," presented at ISMB Intelligent Systems for Molecular Biology Conf., Montreal, Canada, 2024, Poster.

5. R. Gupta, **S. Mustafa**, L. Goldstone, J. Allport, H. Mohammed, and R. Fraser, "Predicting delayed healing chronic wounds using big data and AI-based objective healing index," presented at SAWC Symposium on Advanced Wound Care, Las Vegas, NV, USA, 2024, Poster.