

Murto Hilali

Toronto, ON | 647-569-4616 | mhilali@uwaterloo.ca | [LinkedIn](#) | [GitHub](#) | [Website](#) | [Medium](#)

SUMMARY OF QUALIFICATIONS

Computational biology researcher and programmer with experience in building and deploying computational pipelines for biological data analysis. Led development of a benchmark for six protein-ligand docking models, optimizing distributed training on 11,500+ complexes and cutting runtime by 50%, and created an XGBoost classifier that achieved 0.91 AUC for predicting protein-protein interaction impacts. Published findings in Nature Genetics and presented results to over 45 researchers. Seeking to apply advanced ML and HPC skills to accelerate drug discovery and improve predictive modelling in a fast-paced biotech environment.

EXPERIENCE

QDX Technologies

10/2024 - 12/2024

Computational Chemistry Research Student

Singapore, SG

- Designed and executed a comprehensive benchmark comparison of 6 protein-ligand docking models (DiffDock, FABind, etc.) across 3 experimental datasets using 30+ metrics, revealing fundamental limitations in physical plausibility.
- Optimized distributed retraining pipeline (DiffDock-Pocket, FABind) on 11,500+ protein-ligand complexes, implementing early stopping mechanisms in PyTorch that reduced training time by multiple hours while maintaining reproducibility.
- Validated ~25 predicted poses using PoseBusters and PyMol, systematically uncovering steric clashes and misalignments to explore deep learning models' inability to produce physicochemically valid outputs.
- Evaluated 12+ large-scale databases (10K+ entries, several GB each) for binding affinity data leakage, identifying clean datasets that improved retrained model reliability by eliminating train-test contamination.

The Hospital for Sick Children

01/2023 - 12/2023

Bioinformatics Researcher

Toronto, ON

- Implemented an XGBoost classifier to predict structural impacts of de novo missense variants on protein-protein interactions, achieving a micro-average AUC of 0.91 via hyperparameter tuning and GPU parallelization on a 10,000+ variant dataset.
- Discovered 2 novel protein mechanisms implicated in Autism Spectrum Disorder by engineering features from 3D protein structures, exploratory data analysis, and SHAP value interpretation to identify non-obvious biological patterns.
- Developed and applied AlphaFold2-Multimer to predict Cerebral Palsy disease mechanisms, self-studied structural biology fundamentals, and generated data on 25+ protein complexes, surfacing multiple candidate interaction partners per protein as unvalidated hypotheses for future work.
- Co-authored a Nature Genetics publication by collaborating with 10+ researchers across 3 institutions and organizational stakeholders (SickKids, Holland Bloorview, University of Toronto).
- Presented research findings to 45+ scientists and executives at The Center for Applied Genomics, translating complex computational predictions into actionable biological insights.
- Achieved authorship credit on a scientific abstract presented at the International Cerebral Palsy Genomics Consortium.

OICR

01/2022 - 04/2022

Bioinformatics Workflow Developer

Toronto, ON

- Developed three production workflows from scratch in a distributed computing environment using WDL, SLURM, and the Cromwell engine, enabling faster analysis of cancer genomics data
- Enhanced four existing pipelines in OICR's production codebase through collaborative development on GitHub, Bitbucket, and Jira, applying structured development practices to improve pipeline reliability and reduce deployment time for a team of five developers
- Engineered a regression test suite with Jenkins CI/CD that checks execution speed and stochastic variance of bioinformatics algorithms, increasing testing coverage and catching performance regressions early

SKILLS

- **AI/ML:** Deep Learning (PyTorch), XGBoost, Distributed Model Training, Hyperparameter Optimization, Feature Engineering, Model Validation, SHAP Interpretability
- **Scientific Computing, Data & Infrastructure:** AlphaFold2-Multimer, Protein Structure Prediction, Molecular Docking, Structural Bioinformatics, PyMol, ChimeraX, HPC Clusters (SLURM), Docker Containerization, GPU Parallelization, Large-Scale Dataset Management (10K–1M+ samples), Linux/Unix
- **Programming & Tools:** Python (NumPy, SciPy, BioPython, Scikit-Learn, Seaborn), R, Bash, Git, Jupyter Notebooks

EDUCATION

University of Waterloo

09/2020 - 04/2026

Honours Bachelor of Science (BSc), Biotechnology/Economics

Waterloo, ON

- **GPA:** 3.2
- **Coursework:** Mathematical Modeling of Molecular Systems, Methods in Bioinformatics, Biostatistics & Experimental Design