Mehdi FOROOZANDEH SHAHRAKI

PhD Student | Research Assistant

+1 (778) 228 2027 @ mehdiforoozandehsh@gmail.com @ mfa76@sfu.ca 8888 University Dr., Burnaby, B.C. V5A 1S6
GitHub Google Scholar Linkedin ResearchGate Personal Website

I am a PhD student in the Computing Science department at SFU. My research interests are machine learning, and genomics. I am a graduate research assistant in the Computational Biology lab supervised by Dr. Maxwell Libbrecht.

EDUCATION

2023 - Present **Ph.D. Computing Science**, Simon Fraser University, *CGPA : 4.08 /4.33*,
Courses *: Advanced Natural Language Processing, Special Topics in Computational Biology, Human Molecular Genetics*

2020 - 2022 **MSc. Computing Science**, Simon Fraser University, *CGPA : 4.0 /4.33*,
Courses *: Machine Learning, Statistical Machine Learning, Design and Analysis of Algorithms, Problem-based Learning in Bioinformatics* | **Thesis Title :** Evaluating reproducibility of segmentation and genome annotation (SAGA) algorithms

2015 - 2019 **BSc. Cell and Molecular Biology- Microbiology**, University of Tehran, *GPA (Final Two Years) : 3.68 /4*,
Selected Courses *: Biostatistics, Molecular Genetics, Epigenetics, Virology, Evolutionary Biology, Developmental Biology*

PUBLICATIONS

2025 Foroozandeh, Mehdi, *et al.* **"CANDI : self-supervised, confidence-aware denoising imputation of genomic data"** (Preprint : `https://doi.org/10.1101/2025.01.23.634626`).

2024 Foroozandeh, Mehdi, *et al.* **"Robust chromatin state annotation."** Genome Research 34.3 (2024) : 469-483.

2022 Foroozandeh, Mehdi, *et al.* **"A computational learning paradigm to targeted discovery of biocatalysts from metagenomic data : a case study of lipase identification"** Biotechnology and Bioengineering :21-130.R1 (2022).

2020 Foroozandeh, Mehdi, *et al.* **"A generalized machine-learning aided method for targeted identification of industrial enzymes from metagenome : A xylanase temperature dependence case study."** Biotechnology and Bioengineering (2020).

2020 Foroozandeh, Mehdi, *et al.* **"MCIC : Automated Identification of Cellulases From Metagenomic Data and Characterization Based on Temperature and pH Dependence."** Frontiers in microbiology 11 (2020).

EXPERIENCE

January 2025
July 2025

Computational Biology Intern, DIAGNOSTICS DIVISION, Roche
- At Translational Research Bioinformatics Team
- Research on methods for Omics data analysis
- Research on liquid biopsy biomarker discovery

omics | biomarker discovery | oncology

September 2020
Present

Graduate Research Assistant, DEPARTMENT OF COMPUTING SCIENCE, Simon Fraser University
- At Libbrecht Computational Biology Lab
- Research on designing self-supervised foundation models of the human epigenome.
- Research on reproducibility of chromatin state annotations.
- Research on drug resistance (DR) prediction in bacteria.

machine learning | self-supervised learning | reproducibility | probabilitic models | chromatin states | regulatory genomics | epigenomics | drug resistance

January 2019
September 2020

Undergraduate Research Assistant, INSTITUTE OF BIOPHYSICS AND BIOCHEMISTRY, University of Tehran
- At Complex Biological Systems and Bioinformatics (CBB) lab, led by Dr. Kavousi.
- Research on machine learning-based prediction of enzyme properties.
- Research on targeted discovery of novel enzymes from metagenomic sources.

Bioinformatics | Machine Learning | Metagenomics | Enzymes

PROJECTS

FOUNDATION MODELS FOR EPIGENOMICS

2023 - PRESENT

github.com/mehdiforoozandeh/EpiDenoise

Designing self-supervised learning frameworks to train foundation models of the human genome and epigenome. These models generate compact representations that are used for various downstream tasks, including the imputation and denoising of assays such as histone modifications, among others.

self-supervised learning | attention | foundation models | imputation | denoising | transformers | epigenomics | genomics

REPRODUCIBILITY OF GENOME ANNOTATIONS

2021 - 2023

github.com/mehdiforoozandeh/SAGAconf

This study investigates the use of reproducibility as an evaluation method for genome annotations generated by SAGA algorithms. These algorithms, a type of unsupervised probabilistic graphical model, are often built on structures like Hidden Markov Models and are used to cluster genomic sequence data. However, measuring their effectiveness is not straightforward, as is the case with most unsupervised learning algorithms. To overcome this challenge, our research applies reproducibility analysis to evaluate these algorithms.

reproducibility | genome annotations | probabilistic modeling | unsupervised learning | epigenomics | genomics

DRUG RESISTANCE PREDICTION USING MACHINE LEARNING

2020 - 2021

github.com/mehdiforoozandeh/DRML

In related projects, I designed and optimized several machine learning and deep learning frameworks to understand and predict anti-microbial drug resistance from genomic data. This involved extracting custom features from bacterial genomes, comparing machine learning algorithms, and implementing Bayesian hyper-parameter optimization.

deep learning | Bayesian optimization | drug resistance

PREIDCTION OF ENZYMATIC PROPERTIES USING MACHINE LEARNING AND EENZYME MINING FROM METAGENOME

2019-2020

github.com/mehdiforoozandeh/MeTarEnz   github.com/mehdiforoozandeh/MCIC

This research involves developing machine learning models to predict enzyme properties using amino acid sequence data. I created automated software for enzyme discovery from metagenome data. Using these tools, we discovered and mined novel enzymes for specific applications from metagenomic sources.

enzyme mining | metagenomics | machine learning | automation