Marjan Hosseini

⚲ University of Connecticut, Computer Science & Engineering Dept., 369 Fairfield Way, Storrs, CT 06268

✉ marjan.hosseini@uconn.edu

Google Scholar   ⌗ Marjan-Hosseini   ⬛ marjhosseini   ⊕ Website

Professional Summary

Ph.D. candidate in Computer Science and Engineering with 10+ years of experience in computational biology, machine learning, and large-scale data analysis. Hands-on background in modeling high-throughput sequencing data (e.g., nanopore), feature selection, and probabilistic/Bayesian ML. Experienced in end-to-end project execution: data preprocessing, model development, evaluation, and clear communication of results to collaborators with diverse technical backgrounds.

Core Skills

Programming & Software Engineering: Python, R, Matlab, Java, Spark, Git/GitHub, Linux (shell scripting), C/C++ (working knowledge).

Machine Learning & Statistics: Deep learning, supervised/unsupervised learning, feature selection, statistical modeling, Bayesian inference, probabilistic graphical models, model evaluation and benchmarking.

Computational Biology: High-throughput sequencing analysis (ONT, PacBio, Illumina), variant calling, haplotype assembly, polyploid genomics, biological signal processing, omics data integration.

ML/AI Frameworks: PyTorch, TensorFlow, scikit-learn, PyMC3, CUDA-accelerated workflows.

Algorithms & Optimization: Graph algorithms, combinatorial, dynamic programming, large-scale data processing.

Collaboration & Leadership: Interdisciplinary teamwork, mentoring students, scientific writing, clear communication with both technical and non-technical collaborators, workshop lecturing and training.

Education

Ph.D. in Computer Science and Engineering — **University of Connecticut**, USA   Exp. 2026
  Advisor: Prof. Derek Aguiar
  Focus: Computational biology, Bayesian machine learning, deep learning for sequencing data.
  Dissertation proposal defended; thesis defense expected in the coming weeks.

M.Sc. in Computer Science and Engineering — **Politecnico di Milano**, Italy   Jul 2018
  Thesis: "Feature Selection for Microarray Classification Problems"
  GPA: **110/110 (Honors – summa cum laude)**

B.Sc. in Computer Engineering (Software) — **Shomal University**, Iran   Feb 2006
  Graduated in Computer Engineering (Software).

Selected Industry & Research Experience

Department of Computer Science and Engineering, University of Connecticut — Storrs, USA   Aug 2019 – Present

Graduate Research & Teaching Assistant

  - Design and implement machine learning and Bayesian models for high-throughput sequencing and biological time-series data (e.g., nanopore translocation times).

  - Build data processing pipelines for large biological datasets, from raw data cleaning to model training and evaluation.

  - Collaborate with interdisciplinary teams (bioinformatics, biology, and engineering) to define requirements and interpret results.

  - Work on a joint project with deCODE Genetics on non-B DNA structures, including automating preprocessing and large-scale analysis pipelines for very large sequencing datasets.

- Communicate findings through publications, conference presentations, and reports.
- Mentor undergraduate and junior graduate students on research workflows, experimental design, and coding best practices.

DEIB, Politecnico di Milano — Milan, Italy   Dec 2018 – Aug 2019
Research Fellow
- Developed GPU performance prediction models and feature selection methods for large-scale cloud applications (EU H2020 ATMOSPHERE project).
- Implemented machine learning pipelines to evaluate and optimize application performance under resource constraints.
- Built tools for mapping between ontologies and data schemas (SPRINT project), enabling semantic interoperability across complex data sources.

K4Tech, R&D Department — Milan, Italy   Aug 2018 – Nov 2018
R&D Expert in Machine Learning and Signal Processing
- Designed and implemented ML and digital signal/image processing algorithms for smart home appliances and video analytics.
- Worked on applications including face recognition, fire/smoke detection, and virtual makeup using classical and deep learning techniques.

K4Tech, Internship in Machine Learning — Milan, Italy   Apr 2017 – Aug 2017
Machine Learning Intern
- Developed ML algorithms in Matlab/C for computer vision applications, including face detection/recognition and surveillance.
- Built and tested prototypes for real-time video analytics and monitoring systems (e.g., Pantograph Monitoring System).

Selected Technical Projects

Modeling Non-B DNA Structure in Nanopore Data — UConn, CSE Dept.   2023 – Present
- Developed statistical and deep learning models to detect non-canonical DNA structures from nanopore translocation time signals.
- Performed large-scale data preprocessing and feature engineering from raw sequencing signals.
- Collaborating with deCODE Genetics to scale and automate the analysis pipeline for large sequencing data.

Probabilistic Polyploid Haplotype Assembly (pHapCompass) — UConn, CSE Dept.   2022 – 2025
- Developed probabilistic haplotype assembly algorithms for diploid and polyploid genomes that explicitly model and propagate read assignment ambiguity.
- Designed graph-theoretic algorithms to perform statistical inference and uncertainty quantification over an exponential space of possible haplotype phasings.
- Built a simulation workflow for realistic auto- and allopolyploid genomes and DNA-seq data.
- Released four benchmark datasets and an open-source implementation of

Bayesian Reconstruction of Excised mRNA — UConn, CSE Dept.   2021 – 2025
- Built a Bayesian modeling framework for reconstructing excised intron structures and testing differential splicing.
- Implemented inference and evaluation pipelines in Python using probabilistic programming.

Feature Selection for Microarray Classification — Politecnico di Milano   2016 – 2018
- Designed and evaluated distributed feature selection algorithms based on distance correlation for high-dimensional gene expression data.
- Improved classification performance and interpretability of microarray-based models.

Cloud Application Performance Modeling (ATMOSPHERE) — Politecnico di Milano   2018 – 2019

- Combined Bayesian optimization and ML models to predict performance of GPU-intensive applications under different resource allocations.
- Reduced the number of required production samples while maintaining accurate performance estimates.

LEADERSHIP, TEACHING & SERVICE

Leadership & Mentoring
- Mentored undergraduate and junior graduate students on research projects in computational biology and ML.
- Contributed to organizing and running research group activities and collaborations.

Teaching
- Teaching Assistant, Computer Programming, University of Connecticut (2020).
- Lecturer, Siemens Workshop on Neural Networks and Deep Learning, Milan (2017).
- Teaching Assistant, Computer Networks and Internet Engineering, Shomal University (2005).

Professional Service
- Co-organizer, ASE 2025 Program Workshop (2025).
- Reviewer, Nature Communications (2025) — manuscript on non-B DNA structures.
- Program Committee Member, International Workshop on Semantics and the Web for Transport (2021–2023).
- External Reviewer, SAGE Open Journal (2021).
- Sub-reviewer, RECOMB (Computational Molecular Biology) conference (2019).

Selected Honors
- Predoctoral Fellowship Award, University of Connecticut (2022, 2023, 2025).
- Research Fellowship, Politecnico di Milano (2018).
- Graduated with honors in Computer Science, Politecnico di Milano (2018).

SELECTED PUBLICATIONS

Deep statistical modelling of nanopore sequencing translocation times reveals latent non-B DNA structures 2023
M. Hosseini et al.
Bioinformatics (ISMB proceedings).

Bayesian Reconstruction and Differential Testing of Excised Introns 2025
M. Hosseini, D. McConnell, D. Aguiar.
Bioinformatics (accepted).

pHapCompass: probabilistic haplotype assembly for diploid and polyploid genomes with uncertainty quantification 2025
M. Hosseini et al.
Submitted to *RECOMB* (under review). Code and datasets available at GitHub.

A Distributed Feature Selection Algorithm Based on Distance Correlation with an Application to Microarrays 2018
A. Brankovic, **M. Hosseini**, L. Piroddi.
IEEE/ACM Transactions on Computational Biology and Bioinformatics.

LANGUAGES

ENGLISH:	Fluent	ITALIAN:	Intermediate
PERSIAN:	Native Speaker	ARABIC:	Elementary