

Shahabeddin Sotudian

📍 Boston, USA

in LinkedIn

🐙 GitHub

🎓 Google Scholar

@ sotudian@bu.edu

Education

Ph.D. in Systems Engineering - Machine Learning

Advisor: Prof. Ioannis Ch. Paschalidis

Thesis: Robust Learning to Rank and Its Biomedical Applications

Outstanding Ph.D. Dissertation Award (Granted to only the top 5 BU dissertations)

Boston University

Sep 2018 - May 2023

M.S. in Systems Engineering - Data Science

Thesis: Fuzzy Methods in Machine Learning and Data Mining

GPA: 4.0/4.0 (Rank: 1st)

Tehran Polytechnic

Sep 2015 - Sep 2017

- **Coursework:** Learning from Data, Natural Language Processing, Online Learning, Machine Learning, Data Management, Artificial Intelligence and Expert Systems, Optimization Theory and Methods, Advanced Stochastic Modeling, Mathematical Programming, Principle of Simulation, Data Structures and Algorithms

Projects and Experience

🕒 Harvard University | Dana-Farber Cancer Institute

Postdoctoral Research Fellow

Feb 2023 - Present

Boston, USA

- Enhancing gene expression prediction from sequences through Large Language Model (LLM) utilization.
- Leading ongoing project to integrate DNA methylation, histone modifications, and fragmentation features for early cancer detection. Incorporating diverse data types like MeDIP, H3K4me3, and H3K27ac cfChIP to boost sensitivity.
- Analyzing the impact of genetic ancestry on Multicancer Early Detection (MCED) performance.
- Developed a multi-cancer classifier utilizing epigenomic data from 1,084 plasma profiles across 15 cancer types in 415 individuals. Achieved a remarkable AUC of 0.94.

🕒 Best Buy Co.

Machine Learning Research Intern

Jan 2022 - May 2022

Boston, USA

- Developed an automatic fall detection framework using deep learning-based time series methods. Improved the accuracy of existing models by up to 3%.
- Designed a hybrid model (voice and text) for mental health prediction using natural language processing techniques.
- Implemented the prediction pipeline (i.e., data generation and selection; data analysis; modeling and evaluation) for emergency dispatch prediction. Outperformed previous models.

🕒 Boston University

Research And Teaching Assistant

Sep 2018 - May 2023

Boston, USA

■ Learning to Rank & Recommendation Systems

• Distributionally Robust Learning to Rank

Jul 2020 - Oct 2021

- Developed a listwise distributionally robust multi-output regression ranking framework for document retrieval and drug recommendation. Outperformed SOTA ranking models on biomedical data sets by up to 15%.

• Cluster Ranking in Computational Protein Docking

Oct 2019 - Oct 2020

- Proposed a ranking scheme for output predictions of a protein docking server (ClusPro), improving its performance by 21-100%.

• Learning to Rank for Personalized Drug Recommendation

Jan 2019 - Jul 2020

- Proposed a robust elastic-net-based regression ranking algorithm to select the best therapeutic options given the molecular characteristics of a patients tumor.

■ NLP & Language Models

• Attention-Based Deep Net for Drug Recommendation

Aug 2021 - Oct 2022

- Developed Inversion Transformer-based Neural Ranking (ITNR), a Transformer-based model, to enhance drug response prediction using RNAseq gene expression profiles, drug descriptors, and drug fingerprints.
- Introduced a novel loss function based on Inversion and approximate permutation matrices.

• Transformers for Protein Structure Modeling

Spring 2021

- Developed and implemented a deep bidirectional transformer encoder model and processed data at the character level (i.e., individual amino acids) for protein contact prediction.

■ Fairness in Machine Learning

● COVID-19 Data Analysis and Prediction

Apr 2020 - Jul 2021

- Created four predictive models (hospitalization, ICU admission, mechanical ventilation, mortality) with over 93% accuracy using patient characteristics, clinical symptoms, laboratory results, and chest x-rays.
- Examined socioeconomic factors' impact and assessed racial fairness in algorithms.
- Collaborated with 10+ researchers from MGH and BMC for model design and validation.

● Impact of Social Needs on Healthcare Resource Utilization

Apr 2019 - Jan 2020

- Developed predictive models of resource utilization for missed appointments prediction, designed to minimize avoidable disparities in healthcare system.

⊙ Other Projects

Research Fellow

2017 - 2018

■ Fuzzy Methods in Machine Learning and Data Mining

● Fuzzy Clustering for Gene Expression Data Analysis

2017 - 2018

- Proposed a fuzzy possibilistic validity index for determining the number of clusters, as well as a clustering algorithm based on type-2 fuzzy logic and possibility theory for image segmentation and high-dimensional genomic data analysis.

● Hepatitis Expert System

2017 - 2018

- Designed a type-2 possibilistic fuzzy expert system for diagnosis of hepatitis.

Technical Skills



Honors & Awards

- Boston University Outstanding Ph.D. Dissertation Award

May 2023

- Summer Fellowship from Boston University Digital Health Initiative (DHI)

Summer 2019

- Deans Fellowship Award (Boston University)

2018-2019

- Tehran Polytechnic Distinguished IE student award

2015

Selected Publications

- "Liquid Biopsy Epigenomic Profiling for Cancer Subtyping." Nature Medicine, 2023.
- "Machine Learning for Pharmacogenomics and Personalized Medicine: A Ranking Model for Drug Sensitivity Prediction." IEEE/ACM TCBB, 2021.
- "Improved cluster ranking in proteinprotein docking using a regression approach." Computational and structural biotechnology journal, 2021.
- "Distributionally robust learning-to-rank under the Wasserstein metric." PLOS One 2023.
- "Early prediction of level-of-care requirements in patients with COVID-19." Elife, 2020.
- "Development and Validation of Predictive Models for COVID-19 Outcomes in a Safety-net Hospital Population." JAMIA. 2022.
- "ITNR: Inversion Transformer-based Neural Ranking for Cancer Drug Recommendations." AI Medicine, Under Review.
- "Social Determinants of Health and the Prediction of Missed Breast Imaging Appointments," BMC Health Services Research, 2022.