

# Hossein Sharifi-Noghabi

Machine Learning Researcher — Applied Scientist  
hossein1990@gmail.com | Homepage | LinkedIn | Github

## HIGHLIGHT

---

- Expertise in developing and applying machine learning methods for real-world problems (six years of experience in employing machine learning in biology and medicine with five top-tier publications).
- Experienced in teamwork particularly when members have diverse backgrounds (four years of experience in collaborations in academia and one year of cross-team collaboration in industry).
- Interested and passionate about learning (More than 12 online certificates on different topics such as deep learning and reinforcement learning).
- Skilled in Python, R, Pytorch, Slurm, Bash, scikit-learn, Git, Matplotlib, and Pandas (five years of experience in deploying them for real-world problems).

## EDUCATION

---

<b>Ph.D. in Computer Science</b> <i>Simon Fraser University</i>	Sep. 2016 – Sep. 2021 <i>Burnaby, Canada</i>
<b>M.Sc. in Artificial Intelligence</b> <i>Ferdowsi University of Mashhad</i>	Sep. 2012 – Feb. 2015 <i>Mashhad, Iran</i>
<b>B.Eng. in Information Technology</b> <i>Sadjad University of Technology</i>	Sep. 2008 – Jul. 2012 <i>Mashhad, Iran</i>

## EXPERIENCE

---

<b>Machine Learning Researcher (Full-time employee)</b> <i>Borealis AI</i> <ul style="list-style-type: none"><li>• Designing robust models with out-of-distribution generalization capabilities.</li></ul>	Oct. 2022 – Present <i>Vancouver, Canada</i>
<b>AI for Life Resident (Full-time intern)</b> <i>Novartis</i> <ul style="list-style-type: none"><li>• Designed and developed a metric learning model for batch effect removal of SomaScan proteomics data.</li><li>• Developed a conditional denoising auto-encoder for imputation of missing proteins and removing the batch effect.</li><li>• Worked with Study Data Tabulation Model (SDTM) domain and medical history data.</li><li>• Implemented codes in Python, PySpark, and utilized Pytorch and Matplotlib.</li></ul>	Sep. 2021 – Sep. 2022 <i>Basel, Switzerland</i>
<b>Research Admin Support (Part-time intern)</b> <i>Princess Margaret Cancer Centre</i> <ul style="list-style-type: none"><li>• Led a project to propose guidelines on how to employ machine learning in drug response prediction. These guidelines investigate generalization of machine learning methods in making within-/cross-datasets predictions and provided solutions to improve generalization.</li><li>• Implemented codes in Python and R, and utilized Pytorch and scikit-learn.</li></ul>	Jun. 2020 – Apr. 2021 <i>Toronto, Canada</i>
<b>Research Assistant (Full-time student)</b> <i>Simon Fraser University</i> <ul style="list-style-type: none"><li>• Led multiple projects on improving the accuracy of drug response prediction in patients using biomedical omics data. I developed 3 methods based on multi-modal representation learning, transfer learning with input and output space adaptation, and semi-supervised domain generalization. On average, these methods improved the prediction accuracy by 9% compared to state-of-the-art methods.</li><li>• Implemented codes in Python and utilized Pytorch and Keras frameworks.</li><li>• Mentored undergraduate research assistants in Database and Data Mining Laboratory.</li><li>• Collaborated with life scientists, clinical fellows, and staff at the Vancouver Prostate Centre.</li><li>• In collaboration with Decipher Biosciences Inc., developed a novel method to predict metastasis (a binary outcome) in prostate cancer from early stages using biomedical omics data. This method was based on denoising autoencoders and transfer learning of unlabelled and labelled clinical samples and improved the prediction accuracy by 2% compared to state-of-the-art clinical studies. Codes were implemented in Python and R, and Tensorflow framework was utilized.</li></ul>	Sep. 2016 – Jul. 2021 <i>Burnaby, Canada</i>

## SELECTED PUBLICATION

---

- H. Sharifi-Noghabi, P. Alamzadeh Harjandi, O. Zolotareva, C. Collins, M. Ester, **Out-of-distribution generalization from labelled and unlabelled gene expression data for drug response prediction** *Nature Machine Intelligence* 3, 962–972 2021.
- H. Sharifi-Noghabi, S. Jahangiri-Tazehkand, P. Smirnov, C. Hon, A. Mammoliti, S. Kadambat Nair, A. Mer, M. Ester, B. Haibe-Kains, **Drug Sensitivity Prediction From Cell Line-Based Pharmacogenomics Data: Guidelines for Developing Machine Learning Models** *Briefings in Bioinformatics* 22(6), 1–14 2021.
- O. Snow, H. Sharifi-Noghabi, J. Lu, O. Zolotareva, M. Lee, M. Ester, **Interpretable drug response prediction using a knowledge-based neural network** *Proceedings of the 27th ACM SIGKDD* 2021.
- H. Sharifi-Noghabi, S. Peng, O. Zolotareva, C. Collins, M. Ester, **AITL: Adversarial Inductive Transfer Learning with input and output space adaptation for pharmacogenomics** *Bioinformatics* 36, i380-i388 (ISMB 2020).
- H. Sharifi-Noghabi, O. Zolotareva, C. Collins, M. Ester, **MOLI: multi-omics late integration with deep neural networks for drug response prediction** *Bioinformatics* 35 (14), i501–i509 (ISMB/ECCB 2019).
- H. Sharifi-Noghabi, H. Rajabi Mashhadi, and K. Shojaee, **A novel mutation operator based on the union of fitness and design spaces information for Differential Evolution**, *Soft Computing* (21) 6555–6562 2016.
- M. Mohammadi, H. Sharifi-Noghabi, H. Rajabi Mashhadi, and G. Hodtani, **Robust and stable gene selection via Maximum-Minimum Correntropy Criterion** *Genomics* (170) 83-87 2016.

For the complete list please visit my Google Scholar

## SKILL SET

---

- **Programming:** Python, R, MATLAB
- **Frameworks:** Tensorflow, Keras, Pytorch, PySpark
- **Machine learning:** Deep learning, Reinforcement learning, Transfer learning, Domain generalization and adaptation, Multi-task learning, Adversarial learning, Semi-supervised learning, Meta-learning, Multi-modal data integration
- **Libraries:** Pandas, Numpy, Scikit-learn, Matplotlib
- **Other:** High Performance Computing (Slurm), Git, Bash, Object-oriented programming in Python, Teamwork (Slack, MS Teams, Miro), Visualization (Photoshop, Lucidchart), and Latex

## SELECTED PROJECTS

---

- **Vision:** Applied Convolutional Neural Networks to image verification and recognition tasks and neural style transfer to generate new arts.
- **NLP:** Applied LSTM and GRU models to synthesize Shakespeare's text, speech recognition, and music synthesis.
- **RL:** Implemented Q-Learning and Expected Sarsa both with  $\epsilon$ -greedy action selection on Cliff World.
- **RL:** Implemented semi-gradient TD with a Neural Network as the function approximator for an RL agent in 500-State Random Walk Environment.
- **Drug Discovery:** Employed Variational AutoEncoders (VAE) on SMILES representation to design chemical compounds.
- **Biomarker Discovery:** Developed an ensemble feature selection method for biomarker discovery and prediction of clinical Alzheimer's diagnosis using plasma-based proteomics data.

## RESEARCH INTERESTS

---

- **Transfer Learning:** Out-of-distribution generalization, Domain adaptation, Multi-task learning
- **Learning with limited data:** Semi-supervised learning, Meta-learning, Self-supervision
- **Other:** Federated learning (for personalization and under data/system heterogeneity), Sequential decision making under distribution shifts