

ARSHIA SOLTANI MOAKHAR

✉ arshia.soltani2@gmail.com  [Arshia Soltani](#)  [ckodser](#)  [webpage](#)

Education

Sharif University of Technology

September 2019 – Present

Bachelor of Science in Computer Engineering

Tehran, Iran

- Overall GPA: 18.42/20.00
- Major GPA: 18.96/20.00
- Last two years: 19.51/20.00
- Last two years: 19.89/20.00

Publications and Preprints

- **A. Soltani Moakhar***, E. Iofinova*, D. Alistarh, "SPADE: Sparsity-Guided Debugging for Deep Neural Networks," *Submitted to ICLR 2024 Conference, 2023, ([arXiv](#))*.
- **A. Soltani Moakhar**, M. Azizmalayeri, H. Mirzaei, M.T. Manzuri, M.H. Rohban, "Seeking Next Layer Neurons' Attention for Error-Backpropagation-Like Training in a Multi-Agent Network Framework," *to be Submitted to ICML 2024 Conference, 2023, ([arXiv](#))*.
- H. Mirzaei, M. Jafari, H.R. Dehbashi, A. Ansari, S. Ghobadi, M. Hadi, **A. Soltani Moakhar**, M. Soleymani Baghshah, M.H. Rohban, "RODEO: Robust Out-of-Distribution Detection Via Exposing Adaptive Outliers," *Submitted to ICLR 2024 Conference*.
- M. Azizmalayeri, **A. Soltani Moakhar**, A. Zarei, R. Zohrabi, M.T. Manzuri, M.H. Rohban, "Your Out-of-Distribution Detection Method is Not Robust!," *Advances in Neural Information Processing Systems 36, 2022, ([NeurIPS 2022](#))*.

Research Interests

- Machine Learning
- Robustness
- Interpretability
- Sparsity
- Out-of-Distribution Detection
- Game Theory

Research Experience

Internship in Interpretability and Sparsity in Deep Neural Networks

Feb 2023 - Sep 2023

Institute of Science and Technology Austria (IST Austria)

Supervised by: [Prof. Dan Alistarh](#)

• Sparsity-Guided Debugging for Deep Neural Networks ([arXiv](#))

In this study, we enhanced the performance of various interpretability methods by initially sparsifying the network on a selected sample, followed by the application of the interpretability method. This approach proposes a solution to the challenges presented by polysemantic neurons, which are activated by multiple distinct concepts, allowing for an in-depth investigation into the activation triggers of a neuron on a sample. This is particularly useful when the sample may pertain to a lesser known functionality of the neuron.

Research Assistant in Robust and Interpretable Machine Learning Lab

Aug 2021 - Present

Sharif University of Technology

Supervised by: [Prof. Mohammad Hossein Rohban](#)

• Robust Out-of-Distribution Detection Using GAN Architecture ([NeurIPS 2022](#))

Initially, we identified vulnerabilities in existing Robust Out-of-Distribution (OOD) detection methods to end-to-end adversarial attacks. Subsequently, we proposed a novel OOD detection method, inspired by Generative Adversarial Network (GAN) architecture and adversarial training, which achieved state-of-the-art results in robust OOD detection.

• Aligning Self-Interested Neurons in Deep Neural Networks([arXiv](#))

We explored a scenario where neurons are self-interested, aiming to increase the strength of their connection weights to subsequent layer neurons. We demonstrated that when all neurons rationally update their weights, the resulting behavior closely resembles Gradient Descent. This characteristic enables the training of deeper networks under a self-interested neurons scenario compared to previous approaches, thereby extending the applicability of this method to real-world problems.

Honors and Awards

2022 International Collegiate Programming Contest (**ICPC**) World Final participation

2020 **First Place** in Sharif CodeJam

2019 **Silver Medal** in International Olympiad of Informatics (IOI)

2019 **Bronze Medal** in International junior competitive programming competition, infO(1)CUP

2019 **First Team** in the 7th Ferdowsi Collegiate Programming Contest

2018 **First Place** in National Olympiad of Informatics

2017 **Silver Medal** in National Olympiad of Informatics

Selected Presentations

Leveraging Sparsity for Enhanced Interpretability in Deep Neural Networks **Sep 2023**
Institute of Science and Technology Austria (IST Austria)

Delivered a lecture on various interpretability methods for Neural Networks, evaluated the effectiveness of these methods, and discussed how sparsity could significantly improve their performance.

Teaching Experience

Teacher Assistant **Spring 2023**
Sharif University of Technology

Teaching Assistant for the course Machine Learning lectured by Prof. [Abolfazl Motahari](#).
Organized and taught Jupyter, NumPy, Matplotlib, and PyTorch.
Designed and graded four assignments of the course.

National Committee Member **Aug 2020 - Present**
Iranian National Olympiad in Informatics

Proposed and selected algorithmic problems for the Iranian National Olympiad in Informatics, specifically for summer camp exams and IOI team selection tests.
Proposing and selecting Combinatorial problems for the Iranian National Olympiad in Informatics.
Instructor and consultant in competitive programming for the Iranian gold medalists and Iran's IOI team members.
Graph theory lecturer in olympiad of Informatics national summer camp.

Scientific Committee Member **Feb 2021 - Sep 2021**
Rasta NGO

Designing online workshops in Game Theory for high school students.
Taught the optimal kidney exchange problem and fair profit distribution in the workshop.

Related Coursework

Sharif University		Online Courses	
Artificial Intelligence	20.0/20.0	Deep Learning Specialization	DeepLearning.AI
Machine Learning	20.0/20.0	Practical Reinforcement Learning	HSE university
Adv Information Retrieval(NLP)	19.9/20.0	Game Theory I, II	Stanford University
Introduction to Bio informatics	20.0/20.0	Introduction to Deep Learning 11-785	CMU University
Medical Image Processing	18.9/20.0		
Discrete Structures	20.0/20.0		
Design of Algorithms	20.0/20.0		
Game Theory	20.0/20.0		
Linear Algebra	18.8/20.0		
Eng Probability & Statistic	18.2/20.0		

Skills

Programming Python | R | Java | C | C++ | SQL | Bash | \LaTeX | Racket | CUDA

Frameworks PyTorch | Tensorflow, Keras | NumPy | Pandas | Scikit-Learn | Matplotlib | Jupyter | Django

Languages Persian: Native | English: TOEFL iBT Score 106