

Hui Guo

🏠 [hguo1728.github.io](https://github.com/hguo1728) ✉ hguo288@uwo.ca 🌐 github.com/hguo1728

Education

- **Ph.D. in Computer Science, Western University** 09/2021–Present
Department of Computer Science
Advisors: Prof. Grace Y. Yi and Prof. Boyu Wang
Dissertation focus: **Learning with Imperfect Data: Statistical and Machine Learning Methods**
Expected completion: Summer 2026 | Graduate course average: 95.75%
- **B.Sc. in Mathematics and Applied Mathematics, Nankai University** 09/2016–06/2021
Thesis: A Generalization on Conditional Distortion Risk Measures
- **B.Econ. in Finance, Nankai University** 09/2016–06/2021
Thesis: Pairs Trading Strategy Based on Hurst Exponent
Undergraduate course average: 90.30%

Refereed Publications and Manuscripts

Note: NeurIPS and ICLR are leading peer-reviewed conferences in machine learning.

1. **Hui Guo**, Grace Y. Yi, Boyu Wang (2026). Addressing both Variable Selection and Misclassified Responses with Parametric and Semiparametric Methods. *Bernoulli* 32(2): 1303–1327. [\[paper\]](#) [\[code\]](#) [\[arXiv\]](#)
2. Yuan Bian, Yu Shi, **Hui Guo**, Grace Y. Yi, Wenqing He (2025). Physician Effects in Critical Care: A Causal Inference Approach Through Propensity Weighting with Parametric and Super Learning Methods. *Journal of Data Science* 23(1): 130–148. [\[paper\]](#)
3. Gezheng Xu*, **Hui Guo***, Li Yi, Charles Ling, Boyu Wang, Grace Y. Yi (2025). Revisiting Source-Free Domain Adaptation: A New Perspective via Uncertainty Control. *International Conference on Learning Representations (ICLR)*. [\[paper\]](#) [\[code\]](#)
**Equal contribution*
4. **Hui Guo**, Grace Y. Yi, Boyu Wang (2024). Learning from Noisy Labels via Conditional Distributionally Robust Optimization. *Neural Information Processing Systems (NeurIPS)*. [\[paper\]](#) [\[code\]](#) [\[arXiv\]](#)
5. **Hui Guo**, Boyu Wang, Grace Y. Yi (2023). Label Correction of Crowdsourced Noisy Annotations with an Instance-Dependent Noise Transition Model. *Neural Information Processing Systems (NeurIPS)*. [\[paper\]](#) [\[code\]](#)
6. **Hui Guo**, Boyu Wang, Grace Y. Yi (2026). Dual-Granularity Learning for Regression with Continuous Noisy Labels. *Submitted to Neural Information Processing Systems (NeurIPS)*.
7. **Hui Guo**, Grace Y. Yi (2026). Instance-Dependent Bayesian Modeling and Robust Training with Crowdsourced Noisy Data. *To be submitted shortly*.
8. **Hui Guo**, Yi Wan, Boyu Wang, Grace Y. Yi (2026). Reliability-Aware Longitudinal Clinical Prediction from Multimodal Electronic Health Records. *Manuscript in preparation*.

Awards and Distinctions

- **Robert and Ruth Lumsden Graduate Fellowship in Science, Western University** 2024–2025
Awarded annually to one PhD student in the Faculty of Science in recognition of exceptional academic achievement, research excellence, and leadership.

- **The Fourth Research Challenge Contest of Nankai University** May 2019
Received Honorable Mention.
- **Students' Innovation and Entrepreneurship Training Program** 2018–2019
Excellence Award (team project).
- **Nankai University Academic Excellence Scholarship** 2017–2018
Awarded annually to the top 6% of students in recognition of outstanding academic performance.
- **National Mathematical Modeling Contest for University Students** September 2017
Second Prize, Tianjin Division (team award).
- **Nankai University Talent Scholarship** 2016–2017
Awarded annually to the top 5% of students in recognition of excellence in academics, research, leadership, and extracurricular activities.

Selected Talks and Conference Presentations

- Learning with Noisy Labels: Variable Selection and Misclassification Probability Modeling
Lightning talk for Dr. Raymond Carroll's visit, Western University, June 2025.
- Revisiting Source-Free Domain Adaptation: A New Perspective via Uncertainty Control
International Conference on Learning Representations (ICLR), Singapore, April 2025.
- Learning from Noisy Labels via Conditional Distributionally Robust Optimization
Neural Information Processing Systems (NeurIPS), Vancouver, Canada, December 2024.
- Label Correction of Crowdsourced Noisy Annotations with an Instance-Dependent Noise Transition Model
Neural Information Processing Systems (NeurIPS), New Orleans, USA, December 2023.
- Variable Selection for Logistic Regression Models with Misclassified Response
Annual Meeting of the Statistical Society of Canada, Online, May 2022.
- Personalized Physician Recommendation for Critical Care Using the TreeSHAP Method (case study)
Annual Meeting of the Statistical Society of Canada, Online, May 2022.

Research Presentations

Presented at the Data Science Research Group, Western University.

- Large Language Models: From Transformers to Multimodal AI. April 2026.
- An Introduction to Modern Large Language Models: Architecture, Frontiers, and Applications. March 2026.
- Introduction, Basic Tail and Concentration Bounds. October 2025.
- Posterior Convergence of Bayesian Models: Theory, Application, and Implementation. April 2025.
- Introduction to Large Language Models (LLMs). November 2024.
- An Introduction to Deep Generative Models. April 2024.
- Generalization Problem in Adversarial Training. October 2023.
- VC Dimension and Rademacher Complexity. February 2023.
- Review of Machine Learning with Label Noise. September 2022.
- Ising Model Selection Using Penalized Likelihood Function. March 2022.
- Review of Variable Selection with Measurement Error. February 2022.

Workshops & Events

- Remarkable, Vector Institute
Toronto, ON, February 2026
- Southern Ontario Learning Theory (SOLT) Workshop, Vector Institute
University of Waterloo, Waterloo, ON, November 2025

Academic Service

- **Conference reviewer:** ICML 2026; ICLR 2024–2026 (3 years); AISTATS 2024–2026 (3 years); NeurIPS 2024–2026 (3 years).
- **Journal reviewer:** *Foundations and Trends in Signal Processing*; *Neural Networks*.

Selected Teaching Experience

Teaching Assistant, Western University	London, ON, Canada
– CS2209: Applied Logic for Computer Science	2021–2026 (7 terms)
An undergraduate core course for Computer Science students, with around 300 students, covering propositional and predicate logic, resolution, and logic programming in Prolog.	
– CS2035: Data Analysis and Visualization	2022–2024 (3 terms)
An undergraduate elective for Computer Science students, with around 80 students, covering statistical data analysis and visualization techniques using MATLAB.	

Technical Skills

- **Programming:** Python, R, MATLAB, C++, \LaTeX , Prolog.

Research Interests

My research interests broadly span **statistics, machine learning, and artificial intelligence**. I am interested in developing principled methodologies for learning, inference, and decision-making from **complex, imperfect, and uncertain data**. My previous work has focused on imperfect supervision, noisy labels, crowdsourced annotations, robust learning, and uncertainty quantification. Going forward, I am excited to explore broader directions in both statistics and machine learning, including probabilistic and latent variable modeling, missing data, causal inference, foundation models, trustworthy AI, and other emerging areas of data-driven discovery. On the application side, I am particularly interested in healthcare and biomedical research, including multimodal electronic health records, medical imaging, and large language models, while remaining open to a broad range of scientific and real-world applications.