

Yue Zhao

RESEARCH	<input type="checkbox"/> Unsupervised Machine Learning	<input type="checkbox"/> Machine Learning Systems
KEYWORDS	<input type="checkbox"/> Outlier & Anomaly Detection	<input type="checkbox"/> Automated Machine Learning
	<input type="checkbox"/> Fraud Detection	<input type="checkbox"/> Risk Modeling
	<input type="checkbox"/> Healthcare AI	<input type="checkbox"/> Security AI
	<input type="checkbox"/> Out-of-distribution (OOD) Detection	<input type="checkbox"/> Graph Neural Networks
	<input type="checkbox"/> Meta Learning	<input type="checkbox"/> Parallel Computing

CONTACT	✉ yzhao010@usc.edu	TBA
INFORMATION	🐙 github.com/yzhao062	TBA
	🌐 linkedin.com/in/yzhao062	Los Angeles, CA
	🏠 viterbi-web.usc.edu/~yzhao010/	United States, 90089
	🔍 Google Scholar	University of Southern California

BIOGRAPHY I got my Ph.D. in 4 years at Carnegie Mellon University (CMU). **My research accelerates and automates unsupervised ML:** (1) how to support large-scale learning with **ML systems** and (2) how to **automate** unsupervised ML model selection and hyperparameter optimization. I build interdisciplinary AI/ML applications in **finance, healthcare, and security**.

OPEN-SOURCE HIGHLIGHTS **Open-source Contribution:** I have led or contributed as a core developer to more than 10 ML open-source initiatives. Popular ones include PyOD (A Python Toolbox for Scalable Outlier Detection), ADBench (Anomaly Detection Benchmark), and TDC (An ML Data Hub for Drug Discovery)

🔍 YZHAO062

My works receive 🌟**16,000 GitHub Stars** and 15,000,000 downloads as of July 4, 2023.

FULL-TIME	University of Southern California	
PROFESSIONAL	<i>Department of Computer Science</i>	
EXPERIENCE	Assistant Professor (Tenure-Track)	Aug. 2023 - Present

PwC Canada	
<i>Consulting & Deals</i>	
Senior Consultant (Data Scientist)	Aug. 2017 - Jun. 2019
Consultant (Data Scientist)	Feb. 2017 - Jul. 2017
Research Associate (Intern)	May. 2016 - Jan. 2017

EDUCATION	Carnegie Mellon University	Pittsburgh, PA
	<i>Ph.D. in Information Systems</i>	Sep. 2019 - May. 2023
	<ul style="list-style-type: none">• Affiliation: CMU automated learning systems group (Catalyst) and Data Analytics Techniques Algorithms (DATA) Lab• Advisors and Mentors: CMU: Prof. Leman Akoglu, Prof. Zhihao Jia, and Prof. George Chen. I collaborate with Prof. Jure Leskovec at Stanford, and Prof. Philip S. Yu at UIC.• Thesis: Outlier Detection: Automation, Systems, and Applications	

University of Toronto	Toronto, ON
<i>Master of Science in Computer Science</i>	Sep. 2015 - Dec. 2016

University of Cincinnati	Cincinnati, OH
<i>Bachelor of Science in Computer Engineering</i>	Sep. 2010 - May. 2015
<i>Minor: Computer Science and Mathematics</i>	

Preprints & Under Submission

34. Minqi Jiang, Chaochuan Hou, Ao Zheng, Xiyang Hu, Songqiao Han, Hailiang Huang, Xiangnan He, Philip S. Yu, Yue Zhao
 Weakly Supervised Anomaly Detection: A Survey
Under submission
arXiv preprint arXiv:2302.04549
33. Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Yingxia Shao, Wentao Zhang, Bin Cui, Ming-Hsuan Yang
 Diffusion Models: A Comprehensive Survey of Methods and Applications
Under submission
arXiv preprint arXiv:2209.00796
32. Yue Zhao, Leman Akoglu
 Hyperparameter Optimization for Unsupervised Outlier Detection
Under submission
arXiv preprint arXiv:2208.11727
31. Kay Liu*, Yingtong Dou*, Yue Zhao*, et al.
 PyGOD: A Python Library for Graph Outlier Detection
arXiv preprint arXiv:2204.12095
 (*equal contribution)

Peer-reviewed Journal Papers

31. Yue Zhao*, Martin Q. Ma*, Xiaorong Zhang, Leman Akoglu
 The Need for Unsupervised Outlier Model Selection: A Review and Evaluation of Internal Evaluation Strategies
ACM SIGKDD Explorations Newsletter (SIGKDD Explor.), 2023
 (*equal contribution)
30. Kexin Huang*, Tianfan Fu*, Wenhao Gao*, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, Marinka Zitnik
 Artificial Intelligence Foundation for Therapeutic Science
Nature Chemical Biology (NCHEMB), 2022
 (*equal contribution)
29. Yue Zhao*, Zheng Li*, Xiyang Hu, Nicola Botta, Cezar Ionescu, George H. Chen
 ECOD: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions
IEEE Transactions on Knowledge and Data Engineering (TKDE), 2022.
 (*equal contribution)
28. Yue Zhao, Zain Nasrullah, Zheng Li
 PyOD: A Python Toolbox for Scalable Outlier Detection
Journal of Machine Learning Research (JMLR), 2019.

Peer-reviewed Conference & Workshop Papers (with proceedings)


27. Jaemin Yoo, Yue Zhao, Lingxiao Zhao, Leman Akoglu
 DSV: Capturing Self-supervision Alignment for Unsupervised Outlier Model Selection
European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2023
26. Peng Xu, Lin Zhang, Xuanzhou Liu, Jiaqi Sun, Yue Zhao, Haiqin Yang, Bei Yu
 Do Not Train It: A Linear Neural Architecture Search of Graph Neural Networks
International Conference on Machine Learning (ICML), 2023
25. Yue Zhao, Guoqing Zheng, Subhabrata Mukherjee, Robert McCann, Ahmed Awadallah
 ADMoE: Anomaly Detection with Mixture-of-Experts from Noisy Labels
Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI), 2023

24. Yue Zhao, George H. Chen, Zhihao Jia
TOD: GPU-accelerated Outlier Detection via Tensor Operations
International Conference on Very Large Data Bases (VLDB), 2023
23. Songqiao Han*, Xiyang Hu*, Hailiang Huang*, Minqi Jiang*, Yue Zhao*
ADBench: Anomaly Detection Benchmark
Advances in Neural Information Processing Systems (NeurIPS), 2022
(*equal contribution & the corresponding author)
22. Yue Zhao*, Kay Liu*, Yingdong Dou*, et al.
Benchmarking Node Outlier Detection on Graphs
Advances in Neural Information Processing Systems (NeurIPS), 2022
(*equal contribution)
21. Yue Zhao, Xiaorong Zhang, Leman Akoglu
ELECT: Toward Unsupervised Outlier Model Selection
IEEE International Conference on Data Mining (ICDM), 2022.
Regular paper. Acceptance rate 9.77% (85/870); overall acceptance 20% (174/870).
20. Zhiming Xu, Xiao Huang, Yue Zhao, Yushun Dong, Jundong Li
Contrastive Attributed Network Anomaly Detection with Data Augmentation
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2022
Acceptance rate 19%.
19. Yue Zhao, Ryan A. Rossi, Leman Akoglu
Automatic Unsupervised Outlier Model Selection
Advances in Neural Information Processing Systems (NeurIPS), 2021
Acceptance rate 26%.
18. Kwei-Herng Lai, Daochen Zha, Junjie Xu, Yue Zhao, Guanchu Wang, Xia Hu
Revisiting Time Series Outlier Detection: Definitions and Benchmarks
Advances in Neural Information Processing Systems (NeurIPS), 2021
17. Kexin Huang*, Tianfan Fu*, Wenhao Gao*, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, Marinka Zitnik
Therapeutics Data Commons: Machine Learning Datasets and Tasks for Drug Discovery and Development
Advances in Neural Information Processing Systems (NeurIPS), 2021
(*equal contribution)
16. Yue Zhao*, Xiyang Hu*, Cheng Cheng, Cong Wang, Changlin Wan, Wen Wang, Jianing Yang, Haoping Bai, Zheng Li, Cao Xiao, Yunlong Wang, Zhi Qiao, Jimeng Sun, Leman Akoglu
SUOD: Accelerating Large-scale Unsupervised Heterogeneous Outlier Detection
Conference on Machine and Learning Systems (MLSys), 2021.
Acceptance rate 23.5% (52/221). (*equal contribution)
15. Kwei-Herng Lai*, Daochen Zha*, Guanchu Wang, Junjie Xu, Yue Zhao, Devesh Kumar, Yile Chen, Purav Zumkhawaka, Minyang Wan, Diego Martinez and Xia Ben Hu
TODS: An Automated Time Series Outlier Detection System (Demo paper)
Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI), 2021.
(*equal contribution)
14. Meng-Chieh Lee, Yue Zhao, Aluna Wang, Pierre Jinghong Liang, Leman Akoglu, Vincent S. Tseng, Christos Faloutsos
AutoAudit: Mining Accounting and Time-Evolving Graphs
IEEE International Conference on Big Data (Big Data), 2020
13. Changlin Wan, Dongya Jia, Yue Zhao, Wennan Chang, Sha Cao, Xiao Wang, and Chi Zhang
A Data Denoising Approach to Optimize Functional Clustering of Single Cell RNA-sequencing Data
IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020

12. Zheng Li, Yue Zhao, Nicola Botta, Cezar Ionescu, Xiyang Hu
COPOD: Copula-Based Outlier Detection
IEEE International Conference on Data Mining (ICDM), 2020.
11. Zheng Li, Yue Zhao, Jialin Fu
SYNC: A Copula based Framework for Generating Synthetic Data from Aggregated Sources
IEEE International Conference on Data Mining Workshops (ICDMW), 2020.
10. Yiqun Mei, Yue Zhao, Wei Liang
DSR: An Accurate Single Image Super Resolution Approach for Various Degradations
IEEE International Conference on Multimedia and Expo (ICME), 2020, London, UK.
9. Yue Zhao, Xuejian Wang*, Cheng Cheng*, Xueying Ding*
Combining Machine Learning Models and Scores using combo Library (Demo paper)
Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI), 2020.
(*equal contribution)
8. Zain Nasrullah, Yue Zhao
Music Artist Classification with Convolutional Recurrent Neural Networks
IEEE International Joint Conference on Neural Networks (IJCNN), 2019, Hungary.
7. Yue Zhao, Zain Nasrullah, Maciej K. Hryniewicki, Zheng Li
LSCP: Locally Selective Combination in Parallel Outlier Ensembles
SIAM International Conference on Data Mining (SDM), 2019, Calgary, Canada.
Acceptance rate 22.7% (90/397).
6. Yue Zhao, Maciej K. Hryniewicki
XGBOD: Improving Supervised Outlier Detection with Unsupervised Representation Learning
IEEE International Joint Conference on Neural Networks (IJCNN), 2018, Rio, Brazil.
5. Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, Xiaoyu Zhu
Employee Turnover Prediction with Machine Learning: A Reliable Approach
Intelligent System Conference (Intellisys), 2018, London, UK.
Acceptance rate 34% (194/568).
4. Yue Zhao*, Zhongtian Qiu*, Yiqing Yang*, Weiwei Li*, Mingming Fan
An Empirical Study of Touch-based Authentication Methods on Smartwatches
ACM International Symposium on Wearable Computers (ISWC), 2017, Maui, USA.
Acceptance rate 25.6% (23/90). (*equal contribution)

Peer-reviewed Workshop Papers (without proceedings)

3. Yue Zhao, Xueying Ding, Jianing Yang, and Haoping Bai.
SUOD: Toward Scalable Unsupervised Outlier Detection
Workshops at the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
Extended version published in *MLSys 2021*.
2. Colin Wan, Zheng Li, Alicia Guo, Yue Zhao
SynC: A Unified Framework for Generating Synthetic Population with Gaussian Copula
Workshops at the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.
Extended version published in *ICDMW 2020*.
1. Yue Zhao, Maciej K. Hryniewicki
DCSO: Dynamic Combination of Detector Scores for Outlier Ensembles
ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD Workshop on Outlier Detection De-constructed), 2018, London, UK.
Extended version published in *SDM 2019*, renamed to LSCP.

 AWARDS, GRANTS, AND FUNDING	Meta 2022 AI4AI Research Award (co-PI)	\$50,000	Oct. 2022
	The Norton Labs Graduate Fellowship	\$20,000	Mar. 2022
	CMU Presidential Fellowship	\$80,000	2019
	Mitacs-Accelerate Research and Development Funding	\$30,000	2016-2017
	University Global Award and Scholarship	\$32,000	2010-2015
	Mantei/Mae Award & Scholar	\$40,000	2012-2015
	Engineer of the Month (University of Cincinnati)		Jun. 2014
INTERNSHIP EXPERIENCE	NortonLifeLock Research Group		
	Machine Learning Research Intern		2022
	• Supervised by Dr. Acar Tamersoy and Dr. Kevin Roundy.		
	Microsoft Research		
	Machine Learning Research Intern		2022
	• Designed weakly supervised anomaly detection algorithms		
	• Supervised by Dr. Guoqing Zheng and Dr. Subhabrata (Subho) Mukherjee.		
	Stanford University, Computer Science Department		
	Visiting Student Researcher		2021
	• Designed new GNN systems.		
	• Supervised by Prof. Jure Leskovec.		
	IQVIA, Analytics Center of Excellence		
	Machine Learning Research Intern		2020
	• Designed new machine learning models in healthcare.		
	• Supervised by Dr. Cao (Danica) Xiao (IQVIA) and Prof. Jimeng Sun (UIUC).		
	Siemens PLM Software USA		
	Software Engineer (Intern & Contract)	Mar. 2012 - Dec. 2014	
	• Managed a Java project to transition the LabManager system to vCloud Director.		
	• Refactored outdated automation code and added new modules and JUnit test cases.		
	• Led a C++ Code Coverage project on Teamcenter platform to strengthen its stability.		
TEACHING EXPERIENCE	Carnegie Mellon University	Pittsburgh, PA	
	Teaching Assistant	Fall 2022	
	<i>Managing Digital Business</i> (Prof. David Riel)		
	Teaching Assistant & co-Instructor (lectures on AutoML and MLSys)	Spring 2022	
	Teaching Assistant & co-Instructor (lectures on AutoML and MLSys)	Fall 2021	
	Teaching Assistant & co-Instructor (lectures on AutoML)	Spring 2021	
	Teaching Assistant & co-Instructor (lectures on AutoML)	Fall 2020	
	<i>Intro to Artificial Intelligence</i> (Prof. David Steier)		
	Teaching Assistant	Spring 2022	
	<i>Digital Transformation</i> (Prof. David Riel)		
	Teaching Assistant (helping on course topics)	Fall 2021	
	<i>Statistics for IT Managers</i> (Prof. Daniel Nagin)		
	University of Toronto	Toronto, ON	
	Teaching Assistant & Lab Session Instructor	Fall 2015	
	<i>Embedded Systems</i> (Prof. Philip Anderson)		
	University of Cincinnati	Cincinnati, OH	
	Teaching Assistant & Lab Session Instructor	Fall 2014	
	<i>Intro to Programming</i> (Prof. George Purdy)		

Conference Organizing Committee

- Workflow Co-Chair for KDD 2023

External Reviewer for Funding Proposals

- Dutch Research Council (NWO)

Program Committee and/or Reviewer for Conferences and Workshops

- KDD 2020, 2021, 2022, 2023
- IJCAI 2022, 2023
- NeurIPS 2021, 2022, 2023
- AAAI 2021, 2022, 2023
- AAAI Demonstrations 2021, 2022
- MICCAI 2020, 2021, 2022
- ICDM 2020
- KDD Workshop on Outlier Detection and Description (ODD), 2021
- KDD Workshop on Anomaly and Novelty Detection (ANDEA), 2021, 2022
- IJCAI Workshop on Artificial Intelligence for Anomalies and Novelty (AI4AN), 2020, 2021
- INFORMS Workshop on Data Science 2021

Journal Reviewer

- Journal of Machine Learning Research (JMLR)
- Machine Learning
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Internet of Things Journal (IoT-J)
- IEEE Intelligent Systems
- IEEE Journal on Selected Areas in Communications (J-SAC)
- Data Mining and Knowledge Discovery (DMAI)
- ACM Transactions on Management Information Systems (TMIS)
- Knowledge and Information Systems (KAIS)
- INFORMS Journal on Computing (IJOC)
- Big Data
- Artificial Intelligence Review (AIRE)
- Neurocomputing
- IEEE Transactions on Systems, Man, and Cybernetics: Systems
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
- IEEE Network Magazine
- IEEE Computational Intelligence Magazine (CIM)
- BioData Mining
- European Journal of Management and Business Economics (EJMBE)
- The Journal of Open Source Software (JOSS)

TALKS AND	CMU Catalyst	<i>How (Not) to Fail Your Academic Job Search</i>	May. 2023
LECTURES	KAUST	<i>Automated and Scalable ML Algorithms and Systems</i>	Apr. 2023
	Emory University	<i>Automated and Scalable ML Algorithms and Systems</i>	Apr. 2023
	USC	<i>Automated and Scalable ML Algorithms and Systems</i>	Mar. 2023
	UC Davis	<i>Automated and Scalable ML Algorithms and Systems</i>	Mar. 2023
	Stony Brook University	<i>Automated and Scalable ML Algorithms and Systems</i>	Feb. 2023
	University of Chicago	<i>Automated and Scalable ML Algorithms and Systems</i>	Feb. 2023
	UC Merced	<i>Automated and Scalable ML Algorithms and Systems</i>	Feb. 2023
	CMU PDL Meeting	<i>Automated and Scalable ML Algorithms and Systems</i>	Jan. 2023
	CMU Data Science Seminar	Guest Lecture <i>Automated Anomaly Detection</i>	Nov. 2022
	LoG Seminar	<i>Large-scale Graph Anomaly Detection</i>	Oct. 2022
	Intuit	<i>Anomaly Detection for Financial Risk Modeling</i>	Aug. 2022
	Rice University	<i>Large-scale Anomaly Detection with Automation</i>	Sep. 2022
	Microsoft Research	<i>Weakly-supervised Anomaly Detection</i>	Sep. 2022
	Wells Fargo	<i>Anomaly Detection for Financial Risk Modeling</i>	Aug. 2022
	Columbia University	Guest Lecture <i>Anomaly Detection</i>	Jul. 2022
	Morgan Stanley	<i>Automated Outlier Detection</i>	Jun. 2022
	Microsoft Research	<i>Automated Outlier Detection</i>	Jun. 2022
	Morgan Stanley	<i>Large-scale Anomaly Detection Systems</i>	Mar. 2022
	Rutgers Business School	<i>Outlier Model Selection</i>	Mar. 2022
	Tesla	<i>Large-scale Anomaly Detection Systems</i>	Feb. 2022
	Catalyst, CMU	<i>Systems for Data Mining Algorithms</i>	Dec. 2021
	E&Y Canada	<i>ML applications in Data Analytics</i>	Oct. 2021
	University of Nottingham	<i>General Machine Learning Applications</i>	Jan. 2021