

curriculum vitae of
Yuan Tian

🏠 Goodwin Hall, Room 635 ✉ y[dot]tian[at]queensu[dot]com
☎ +1 613 5837672 📄 Google Scholar (2,413 citations, h-index: 25, i10-index: 29)

RESEARCH INTERESTS

- Develop innovative machine learning algorithms to assist software development and enhance project management efficiency and software reliability through automation.
- Apply AI and data analytics to extract actionable insights from software repositories, benefiting software developers and researchers.
- Create cutting-edge solutions for Foundation Model (FM)-driven software and software development.

EMPLOYMENT

2018.10 - present **Assistant Professor**
School of Computing, Queen's University, Kingston, ON, Canada

2017.06 - 2018.09 **Data Scientist (Research Fellow)**
Living Analytics Research Centre (LARC), Singapore

EDUCATION

2012.09 - 2017.05 **Ph.D. in Information Systems**
Singapore Management University, Singapore
Thesis: Mining software repositories for automatic software bug management from bug triaging to patch backporting
Supervisor: David Lo (ACM/IEEE Fellow)

2015.09 - 2016.04 **Visiting Ph.D.**
Carnegie Mellon University, United States
Supervisor: Claire Le Goues

2013.11 - 2013.12 **Visiting Ph.D.**
The National Institute for Research in Digital Science and Technology (Inria), France
Supervisor: Julia Lawall

2008.08 - 2012.05 **B.Sc. in Computer Science and Technology**
Zhejiang University, Hangzhou, China

FUNDING

Principal investigator [F10] **NSERC Research Tools and Instruments (RTI) Grant**, \$129,760 2024.04 - 2025.04
An experimental platform to comparatively evaluate large pre-trained models on SE tasks
Co-PI: Bram Adams

[F9] **NSERC Discovery Grant**, \$198,000 2019.04 - 2025.04
Reliable and explainable recommender systems for efficient software development
Note: awarded an extra supplement fund of \$12,500 for Early Career Researchers (ECR)

[F8] **Mitacs Business Strategy Grant**, \$15,000 2023.10 - 2024.04
Predicting the winning bidding price of construction projects by ensemble machine learning
Trainee: Kazi Amit Hasan
Industry partner: Louis W. Bray Construction Limited

[F7] **FAS Research Infrastructure Grant**, \$200,000 2019.12 - 2023.12
Data mining for intelligent and reliable software development

[F6] **Mitacs Accelerate Grant**, \$15,000 2022.09 - 2023.05
Learning personal traits, value, skill representations for improved matching of jobs, talent, and courses
Trainee: Mohammad Sadegh Sheikhaeil
Industry partner: FutureCite

	[F5] Mitacs Accelerate Explore Grant , \$6,000	2021.09 - 2021.12
	<i>Optimizing and simplifying knowledge navigation for job seekers, employers and educators via data-driven personalized recommendations</i>	
	Trainee: Jefferson Jin	
	Industry partner: FutureCite	
	[F4] Queen's University Research Initiation Grant (RIG) , \$60,000	2021.09 - 2021.12
	<i>Reliable and explainable recommender systems in software engineering context</i>	
	[F3] NVIDIA GPU Grant , \$3,000	2018.12 - 2018.12
	<i>Deep learning-based automated software engineering</i>	
Collaborator/co-applicant	[F2] NSERC CREATE Training Program , \$1,650,000	2021.09 - 2027.08
	<i>Preparing Canada's workforce for health data of tomorrow</i>	
	PI: Parvin Mousavi	
	[F1] NSERC/FRQNT NOVA Program , \$225,000	2024.04 - 2027.04
	<i>Towards reliable smart home ecosystems</i>	
	PI: Jinqiu Yang	
Under review	[UF11] NSERC Alliance Advantage-Mitacs Accelerate Program , \$120,000	
	<i>Pragmatic automated code transformation leveraging large language models</i>	
	Industry partner: Ross Video	

HONORS AND AWARDS

	[H12] Best Research Paper Award (1 from 32 submissions)	2024.04
	This award is given to the best (only one) full research paper accepted at the <i>AI Foundation Models and Software Engineering Conference (FORGE'24)</i> co-hosted in ICSE'24. This paper has been invited to be extended and published in EMSE, a top Software Engineering journal.	
	[H11] Distinguished Reviewer Award (13 from 127 reviewers)	2023.05
	This award recognizes outstanding reviewers who served in the main research paper track of the <i>20th IEEE/ACM Mining Software Repository Conference (MSR'23)</i> .	
	[H10] Best Early Research Achievements Paper Award (1 from 19 submissions)	2022.05
	This award is given to the best (only one) early research achievement paper accepted at the <i>30th IEEE/ACM International Conference on Program Comprehension (ICPC'22)</i> .	
	[H9] The Most Influential Paper (MIP) Award (1 out of 96 accepted papers)	2022.03
	This award is given to the Most Influential Paper (only one) at the <i>29th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'22)</i> . An influential paper should have made an impact on research and/or practice, by generating/influencing further research, by releasing tools or datasets everybody is using, or by influencing practitioners.	
	[H8] Best New Ideas and Emerging Results Paper Award (1 out of 28 submissions)	2021.09
	This award is given to the best (only one) new emerging idea paper accepted at the <i>37th International Conference on Software Maintenance and Evolution (ICSME'21)</i> .	
	[H7] Best Early Research Achievements Paper Award (1 out of 43 submissions)	2017.05
	This award is given to the best (only one) early research achievement paper accepted at the <i>24th IEEE International Conference on Software Analysis, Evolution, and Reengineering (SANER'17)</i> .	
	[H6] Representative of SMU at Royal Society Commonwealth Science Conference	2016.05
	Only one student was selected to represent SMU's School of Information Systems at this conference.	
	[H5] SMU Presidential Doctoral Fellowship Award (2 out of 50+ Ph.D students)	2015.05
	The SMU Presidential Doctoral Fellowship is awarded to exceptionally qualified students admitted to SMU's academic research PhD programs. Notably, only two students from SMU's School of Information Systems received this prestigious fellowship in 2015.	
	[H4] Best Research Paper Award Nomination (5 out of 163 submissions)	2013.09
	This nomination is given to the best (only five) full research papers accepted at the <i>29th IEEE International Conference on Software Maintenance (ICSM'13)</i> . This paper was invited to be extended and published in EMSE, a top Software Engineering journal.	

- [H3] **Temasek Foundation Leadership Enrichment and Regional Networking Scholarship** 2011.12
Sponsored by Temasek Foundation, this Scholarship Programme provides a global learning experience for talented ASEAN undergraduates with good academic standing and leadership potential to participate and learn alongside a group of international students from varied cultures and Singaporean students.
- [H2] **Third Scholarship for Academic Performance at Zhejiang University** 2009-2010, 2010-2011
- [H1] **Excellent Student Award at Zhejiang University (for top 5% students)** 2008-2009

PRESENTATIONS AND TALKS

Invited talk

- [T18] Optimizing Software Project Management with AI-powered Tracking Tool 2023.10
Invited talk at the colloquium series in the School of Computer Science at the University of Windsor, Online
- [T17] Optimizing Software Project Management with AI-powered Tracking Tool 2023.06
Invited talk at the Software Engineering for Machine Learning Applications (SEMLA) international symposium, Montreal, Canada
- [T16] Information Retrieval Based Nearest Neighbour Classification for Fine-grained Bug Severity Prediction 2022.03
Invited MIP Award talk at the *29th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'22)*, Online
- [T15] Intelligent Software Project Management Leveraging Big Data: Promises and Perils 2022.03
Invited talk at the 3rd SEMLA Exchange Session with IVADO Industrial Member, Online
- [T14] Modelling Jobs and Skills: A Network Embedding-based Approach 2018.04
Invited talk at the Meeting on Jobs and Skills Intelligence with Workforce Singapore and SkillsFuture Singapore, Singapore
- [T13] Leveraging Community Intelligence to Cope with Inconsistent Labels 2018.04
Invited talk at Queen's University, Kingston, Canada
- [T12] Smart Nation Domain 2: Jobs and Skills Intelligence 2018.04
Invited talk at the RIE (Research, Innovation, and Enterprise) 2020 Mid-Term Review of the Service and Digital Economy Domain, Singapore
- [T11] Automated Bug Report Analytics 2016.04
Invited talk at Advanced Digital Sciences Center (ADSC), Singapore

Conference talk

- [T10] Analyzing Gamer Complaints in Reviews of Cross-Platform Video Games on Steam 2023.08
Conference talk at the *IEEE Conference on Games (CoG)*, Boston, United States
- [T9] Understanding the Time to First Response in GitHub Pull Requests 2023.05
Conference talk at the *20th IEEE/ACM International Conference on Mining Software Repositories (MSR'23)*, Melbourne, Australia
- [T8] Automated Self-Admitted Technical Debt Tracking at Commit-Level: A Language-independent Approach 2023.05
Conference talk at the *ACM/IEEE International Conference on Technical Debt (TechDebt'23)*, Melbourne, Australia
- [T7] A Study of Update Request Comments in Stack Overflow Answer Posts 2023.05
Conference talk at the *31st IEEE/ACM International Conference on Program Comprehension (ICPC'23)*, Melbourne, Australia
- [T6] Adoption of Third-party Libraries in Mobile Apps: A Case Study on Open-source Android Applications 2022.05
Conference talk at the *9th IEEE/ACM International Conference on Mobile Software Engineering and Systems (MobileSoft'22)*, Online
- [T5] A First Look at Duplicate and Near-duplicate SATD Comments 2022.05
Conference talk at the *30th IEEE/ACM International Conference on Program Comprehension (ICPC'22)*, Online
- [T4] Rule-based Specification Mining Leveraging Learning to Rank 2019.11
Conference talk at the *34th IEEE/ACM International Conference on Automated Software Engineering (ASE'19)*, San Diego, United States

- [T3] Recommending Who to Follow in the Software Engineering Twitter Space 2019.11
Conference talk at the 34th IEEE/ACM International Conference on Automated Software Engineering (ASE'19), San Diego, United States
- [T2] Drone: Predicting Priority of Reported bugs by Multi-factor Analysis 2013.09
Conference talk at the 29th International Conference on Software Maintenance (ICSM'13), Eindhoven, Netherlands
- [T1] Automatic Classification of Software Related Microblogs 2012.04
Conference talk at the 28th International Conference on Software Maintenance (ICSM'12), Trento, Italy

PUBLICATIONS

Graduate or undergraduate students under my direct supervision are marked with a †.

My research on designing novel automated software engineering approaches and mining insights from software repositories has been published in top (Core-A) software engineering journals, including IEEE Transactions on Software Engineering (TSE) (IF: 7.4), ACM Transactions on Software Engineering and Methodology (TOSEM) (IF: 4.4), Empirical Software Engineering Journal (EMSE) (IF: 4.1), Information and Software Technology (IST) (IF: 3.9), Journal of Systems and Software (JSS) (IF: 3.5); as well as in AI and Cybersecurity journals, including Computer & Security (IF: 5.6), IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI) (IF: 5.3), ACM Transactions on Privacy and Security (TOPS) (IF: 2.3), ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) (IF: 2.0).

Journal articles

- [U]21 Jerin Yasmin[†], Jiale Wang[†], **Yuan Tian**, Bram Adams (2024). An Empirical Study of Developers' Challenges in Implementing Workflows as Code: A Case Study on Apache Airflow. Major Revision Requested by *Journal of Systems and Software (JSS)*
- [U]20 Mohammad Sadegh Sheikhaeil[†], **Yuan Tian**, Shaowei Wang, Bowen Xu (2024). An Empirical Study on the Effectiveness of Large Language Models for SATD Identification and Classification. Empirical. Major Revision Requested by *Empirical Software Engineering Journal (EMSE)*
- [U]19 Huiyi Hao[†], Kazi Amit Hasan[†], Hong Qin[†], Marcos Macedo[†], **Yuan Tian**, Steven H.H. Ding, Ahmed E. Hassan (2024). An Empirical Study on Developers Shared Conversations with ChatGPT in GitHub Pull Requests and Issues. Under Second Review by *Empirical Software Engineering Journal (EMSE)*
- [U]18 Azmain Kabir, Shaowei Wang, **Yuan Tian**, Tse-Hsun Chen, Muhammad Asaduzzaman, Wenbing Zhang (2024). ZS4C: Zero-Shot Synthesis of Compilable Code for Incomplete Code Snippets using ChatGPT. Under Second Review by *ACM Transactions on Software Engineering and Methodology (TOSEM)*
- [U]17 Leo Song, Steven H.H Ding, **Yuan Tian**, Li Tao Li, Weihan Ou, Philippe Charland, Andrew Walenstein (2024). Reinforcement Subsequence Learning for Obfuscated Clone Search in Scripting Languages. Under Second Review by *ACM Transactions on Software Engineering and Methodology (TOSEM)*
- [J]16 Rongcun Wang, Senlei Xu, **Yuan Tian**, Xiaobing Sun (2024). SCL-CVD: Supervised Contrastive Learning for Code Vulnerability Detection via GraphCodeBERT. *Computer & Security Journal (C&S)*, DOI: 10.1016/j.cose.2024.103994
- [J]15 Shahla Shaan Ahmed, Shaowei Wang, **Yuan Tian**, Tse-Hsun Chen, Haoxiang Zhang (2024). Studying and recommending information highlighting in Stack Overflow answers. *Information and Software Technology (IST)*, DOI:10.1016/j.infsof.2024.107478
- [J]14 Rongcun Wang, Xingyu Ji, Senlei Xu, **Yuan Tian**, Shujuan Jiang, Rubing Huang (2024). An Empirical Assessment of Different Word Embedding and Deep Learning Models for Bug Assignment. *Journal of Systems and Software (JSS)*, DOI: 10.1016/j.jss.2024.111961
- [J]13 Rongcun Wang, Senlei Xu, Xingyu Ji, **Yuan Tian**, Lina Gong, Ke Wang (2024). An Extensive Study of The Effects of Different Models on Code Vulnerability Detection in Python Code. *Automated Software Engineering Journal (AUSE)*, DOI: 10.1007/s10515-024-00413-4
- [J]12 Mohamed Amine Batoun, Ka Lai Yung[†], **Yuan Tian**, Mohammed Sayagh (2023). An Empirical Study on GitHub Pull Requests' Reactions. *ACM Transactions on Software Engineering Methodology (TOSEM)*, DOI:10.1145/3597208
- [J]11 Litao Li, Steven H.H Ding, **Yuan Tian**, Fung, Benjamin C. M. Fung, Philippe Charland, Weihan Ou, Leo Song, Chen, Congwei (2023). VulANalyzeR: Explainable Binary Vulnerability Detection with Multi-Task Learning and Attentional Graph Convolution. *ACM Transactions on Privacy and Security (TOPS)*, DOI:10.1145/3585386

- [J10] Weihan Ou, Steven H.H Ding, **Yuan Tian**, Leo Song (2022). SCS-Gan: Learning Functionality-Agnostic Stylometric Representations for Source Code Authorship Verification. *IEEE Transactions on Software Engineering (TSE)*, DOI:10.1109/TSE.2022.3177228
- [J9] Mohammad Sadegh Sheikhaei[†], **Yuan Tian**, Shaowei Wang (2022). A Study of Update Request Comments in Stack Overflow Answer Posts. *Journal of Systems and Software (JSS)*, DOI:10.1016/j.jss.2022.111590
- [J8] Mohammad Sadegh Sheikhaei[†], Hasan Zafari, **Yuan Tian** (2021). Joined Type Length Encoding for Nested Named Entity Recognition. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, DOI:10.1145/3487057
- [J7] Maram Assi, Safwat Hassan, **Yuan Tian**, Ying Zou (2021). Joined Type Length Encoding for Nested Named Entity Recognition. *Empirical Software Engineering Journal (EMSE)*, DOI:10.1007/s10664-021-09988-y
- [J6] Thong Hoang, Julia Lawall, **Yuan Tian**, Richard J Oentaryo, David Lo (2019). Patchnet: Hierarchical deep learning-based stable patch identification for the Linux kernel. *IEEE Transactions on Software Engineering (TSE)*, DOI:10.1109/TSE.2019.2952614
- [J5] Zherui Cao[†], **Yuan Tian**, Tien-Duy B Le, David Lo (2018). Rule-based Specification Mining Leveraging Learning to Rank. *Automated Software Engineering Journal (AUSE)*, DOI:10.1007/s10515-018-0231-z
- [J4] Abhishek Sharma[†], **Yuan Tian**, Agus Sulistya, Dinusha Wijedasa, David Lo (2018). Recommending Who to Follow in the Software Engineering Twitter Space. *ACM Transactions on Software Engineering Methodology (TOSEM)*, DOI:10.1145/3266426
- [J3] Ferdian Thung, Richard J Oentaryo, David Lo, **Yuan Tian** (2017). WebAPIRec: Recommending Web APIs to Software Projects via Personalized Ranking. *IEEE Transactions on Emerging Topics in Computational Intelligence (TETCI)*, DOI:10.1109/TETCI.2017.2699222
- [J2] **Yuan Tian**, Nasir Ali, David Lo, Ahmed E. Hassan (2015). On the Unreliability of Bug Severity Data. *Empirical Software Engineering Journal (EMSE)*, DOI:10.1007/s10664-015-9409-1
- [J1] **Yuan Tian**, David Lo, Xin Xia, Chengnian Sun (2015). Automated Prediction of Bug Report Priority using Multi-factor Analysis. *Empirical Software Engineering Journal (EMSE)*, DOI:10.1007/s10664-014-9331-y

My research on data-driven software engineering has been published and presented at top software engineering conferences, including the International Conference on Software Engineering (ICSE), the International Conference on Automated Software Engineering (ASE), and top conferences in sub-domains of software engineering, such as the International Conference on Software Maintenance and Evolution (ICSME, previously ICSM), the International Conference on Mining Software Repositories (MSR), the International Conference on Software Analysis, Evolution and Reengineering (SANER, previously CSMR and WCRE), the International Conference on Program Comprehension (ICPC), and the International Conference on Technical Debt (TechDebt)

Conference papers

- [C39] Junjie Li, Aseem Sangalay, Cheng Cheng, **Yuan Tian**, Jinqiu Yang (2024). Fine Tuning Large Language Model for Secure Code Generation. In *Proceedings of the 1st International Conference on AI Foundation Models and Software Engineering (Forge'24) co-hosted by the 46th International Conference on Software Engineering (ICSE'24)*
- [C38] Macedo Marcos[†], **Yuan Tian**, Filipe Cogo, Bram Adams (2024). Exploring the Impact of the Output Format on the Evaluation of Large Language Models for Code Translation. In *Proceedings of the 1st International Conference on AI Foundation Models and Software Engineering (Forge'24) co-hosted by the 46th International Conference on Software Engineering (ICSE'24) (Best Paper Award)*
- [C37] Wenxin Jiang, Jerin Yasmin[†], Jason Jones, Nicholas Synovic, Jiashen Kuo, Nathaniel Bielanski, **Yuan Tian**, Thiruvathukal, George K Thiruvathukal, James C Davis (2024). PeaTMOSS: A Dataset and Initial Analysis of Pre-Trained Models in Open-Source Software. In *Proceedings of the 21st IEEE/ACM International Conference on Mining Software Repositories (MSR'24)*
- [C36] Hanwen Hu[†], **Yuan Tian**, Safwat Hassan, Dayi Lin (2023). Analyzing Gamer Complaints in Reviews of Cross-Platform Video Games on Steam. In *Proceedings of the IEEE Conference on Games (CoG'23)*
- [C35] Kazi Amit Hasan[†], Marcos Macedo[†], **Yuan Tian**, Bram Adams, Steven H.H. Ding (2023). Understanding the Time to First Response in GitHub Pull Requests. In *Proceedings of the 20th IEEE/ACM International Conference on Mining Software Repositories (MSR'23)*
- [C34] Mohammad Sadegh Sheikhaei[†], **Yuan Tian** (2023). Automated Self-Admitted Technical Debt Tracking at Commit-Level: A Language-independent Approach. In *Proceedings of the 6th IEEE/ACM International Conference on Technical Debt (TechDebt'23)*

- [C33] Nafiz Sadman, Kazi Amit Hasan[†], Elyas Rashno, Furkan Alaca, **Yuan Tian**, and Farhana Zulkernine (2023). Vulnerability of Open-source Face Recognition Systems towards Blackbox Attacks: A Case Study with InsightFace. In *Proceedings of the IEEE Symposium on Computational Intelligence in Cyber Security (SSCI'23)*
- [C32] Shahla Shaan Ahmed, Shaowei Wang, Haoxiang Zhang, Tse-Hsun Chen, **Yuan Tian** (2022). A First Look at Information Highlighting in Stack Overflow Answers. In *Proceedings of the 38th IEEE International Conference on Software Maintenance and Evolution (ICSME'22)*
- [C31] Jerin Yasmin[†], Mohammad Sadegh Sheikhaei[†], **Yuan Tian** (2022). A First Look at Duplicate and Near-duplicate Self-admitted Technical Debt Comments. In *Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension (ICPC'22)*, **(Best ERA Paper Award)**
- [C30] Ratnadira Widyasari, Gede Artha Azriadi Prana, Stefanus A. Haryono, **Yuan Tian**, Hafil Noer Zachary, and David Lo (2022). XAI4FL: Enhancing Spectrum-based Fault Localization with Explainable Artificial Intelligence. In *Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension (ICPC'22)*
- [C29] Aidan Polese[†], Safwat Hassan, **Yuan Tian** (2022). Adoption of Third-party Libraries in Mobile Apps: A Case Study on Open-source Android Applications. In *Proceedings of the 9th IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft'22)*
- [C28] Dennis Huynh[†], Garrett Audet[†], Nikolay Alabi[†], **Yuan Tian** (2021). Stock Price Prediction Leveraging Reddit: The Role of Trust Filter and Sliding Window. In *Proceedings of the IEEE International Conference on Big Data (BigData'21)*
- [C27] Shaowei Wang, Daniel M Germán, Tse-Hsun Chen, **Yuan Tian**, Ahmed E. Hassan (2021). Is Reputation on Stack Overflow a Good Indicator for Users' Expertise? No!. In *Proceedings of the 37th International Conference on Software Maintenance and Evolution (ICSME'21)* **(Best NIER Paper Award)**
- [C26] Jerin Yasmin[†], **Yuan Tian**, Jinqiu Yang (2020). A First Look at the Deprecation of RESTful APIs: An Empirical Study. In *Proceedings of the 36th International Conference on Software Maintenance and Evolution (ICSME'20)*
- [C25] Hoang, Thong, Julia Lawall, Richard J. Oentaryo, **Yuan Tian**, David Lo. (2019). PatchNet: A Tool for Deep Patch Classification. In *Companion Proceedings of the 41st IEEE/ACM International Conference on Software Engineering (ICSE-Companion'19)*
- [C24] Chiang, Meng-Fen, Ee-Peng Lim, Wang-Chien Lee, **Yuan Tian**, Chih-Chieh Hung (2018). Are you on the Right Track? Learning Career Tracks for Job Movement Analysis. In *Proceedings of the 2nd International Data Science for Human Capital Management (DSHCM) workshop co-hosted in the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'18)*
- [C23] **Yuan Tian**, Ferdian Thung, Abhishek Sharma, David Lo (2017). APIBot: Question Answering Bot for API Documentation. In *Proceedings of the 32nd IEEE/ACM international conference on automated software engineering (ASE'17)*
- [C22] **Yuan Tian**, Pavneet Singh Kochhar, David Lo (2017). An Exploratory Study of Functionality and Learning Resources of Web APIs on ProgrammableWeb. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE'17)*
- [C21] Abhishek Sharma[†], **Yuan Tian**, Agus Sulistya, David Lo, Aiko Fallas Yamashita (2017). Harnessing Twitter to Support Serendipitous Learning of Developers. In *Proceedings of the 24th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'17)* **(Best ERA Paper Award)**
- [C20] **Yuan Tian**, Dinusha Wijedasa[†], David Lo, Claire LE Goues (2016). Learning to Rank for Bug Report Assignee Recommendation. In *Proceedings of the 24th IEEE International Conference on Program Comprehension (ICPC'16)*
- [C19] Xiao Xuan, David Lo, Xin Xia, **Yuan Tian** (2015). Evaluating Defect Prediction Approaches using a Massive Set of Metrics: an empirical study. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing (SAC'15)*
- [C18] **Yuan Tian**, David Lo (2015). A Comparative Study on the Effectiveness of Part-of-Speech Tagging Techniques on Bug Reports. In *Proceedings of the 22nd IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'15)*
- [C17] Abhishek Sharma[†], **Yuan Tian**, David Lo (2015). NIRMAL: Automatic Identification of Software Relevant Tweets Leveraging Language Model. In *Proceedings of the 22nd IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'15)*
- [C16] Abhishek Sharma[†], **Yuan Tian**, David Lo (2015). What's Hot in Software Engineering Twitter Space? In *Proceedings of the 31st International Conference on Software Maintenance and Evolution (ICSME'15)*

- [C15] **Yuan Tian**, Meiyappan Nagappan, David Lo, Ahmed E. Hassan (2015). What are the Characteristics of High-Rated Apps? A Case Study on Free Android Applications. In *Proceedings of the 31st International Conference on Software Maintenance and Evolution (ICSME'15)*
- [C14] Pavneet Singh Kochhar, **Yuan Tian**, David Lo (2014). Potential Biases in Bug Localization: Do They Matter? In *Proceedings of the 29th ACM/IEEE international conference on Automated Software Engineering (ASE'14)*
- [C13] **Yuan Tian**, David Lo, Julia Lawall (2014). SEWordSim: Software-specific Word Similarity Database. In *Companion Proceedings of the 36th ACM/IEEE International Conference on Software Engineering (ICSE-Companion'14)*
- [C12] Xinyu Wang, David Lo, Xin Xia, Xingen Wang, Pavneet Singh Kochhar, **Yuan Tian**, Xiaohu Yang, Shanping Li, Jianling Sun, Bo Zhou (2014). BOAT: an Experimental Platform for Researchers to Comparatively and Reproducibly Evaluate Bug Localization Techniques. In *Companion Proceedings of the 36th ACM/IEEE International Conference on Software Engineering (ICSE-Companion'14)*
- [C11] **Yuan Tian**, David Lo, Julia Lawall (2014). Automated Construction of a Software-specific Word Similarity Database. In *Proceedings of the Software Evolution Week-IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering (CSMR-WCRE'14)*
- [C10] **Yuan Tian**, David Lo (2014). An Exploratory Study on Software Microblogger Behaviors. In *Proceedings of the IEEE 4th Workshop on Mining Unstructured Data (MUD'14) co-hosted in the 29th IEEE International Conference on Software Maintenance and Evolution (ICSME'14)*
- [C9] **Yuan Tian**, Pavneet Singh Kochhar, Ee-Peng Lim, Feida Zhu, David Lo (2014). Predicting Best Answerers for New Questions: An Approach Leveraging Topic Modeling and Collaborative Voting. In *Proceedings of the 5th International Conference on Social Informatics (SoInfo'13)*
- [C8] **Yuan Tian**, David Lo, Chengnian Sun (2013). Drone: Predicting Priority of Reported Bugs by Multi-factor Analysis. In *Proceedings of the 28th International Conference on Software Maintenance (ICSM'13) (Best Research Paper Nomination)*
- [C7] Didi SURIAN, **Yuan Tian**, David LO, Hong CHENG, Ee Peng LIM (2013). Predicting Project Outcome Leveraging Socio-technical Network Patterns. In *Proceedings of the 17th European Conference on Software Maintenance and Reengineering (CSMR'13)*
- [C6] Palakorn Achananuparp, Ibrahim Nelman Lubis, **Yuan Tian**, David Lo, Ee-Peng Lim (2012). Observatory of Trends in Software Related Microblogs. In *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering (ASE'12)*
- [C5] **Yuan Tian**, Julia Lawall, David Lo (2012). Identifying Linux Bug Fixing Patches. In *Proceedings of the 34th IEEE/ACM International Conference on Software Engineering (ICSE'12)*
- [C4] **Yuan Tian**, Chengnian Sun, David Lo (2012). Improved Duplicate Bug Report Identification. In *Proceedings of the 16th European Conference on Software Maintenance and Reengineering (CSMR'12)*
- [C3] Philips Kokoh Prasetyo, David Lo, Achananuparp Palakorn, **Yuan Tian**, Ee Peng Lim (2012). Automatic Classification of Software Related Microblogs. In *Proceedings of the 28th IEEE International Conference on Software Maintenance (ICSM'12)*
- [C2] **Yuan Tian**, Chengnian Sun, David Lo (2012). Information Retrieval Based Nearest Neighbor Classification for Fine-Grained Bug Severity Prediction. In *Proceedings of the 19th Working Conference on Reverse Engineering (WCRE'12)*
- [C1] **Yuan Tian**, Palakorn Achananuparp, Ibrahim Nelman Lubis, David Lo, Ee-Peng Lim (2012). What Does Software Engineering Community Microblog About? In *Proceedings of the 9th Working Conference on Mining Software Repositories (MSR'12)*
- Book Chapter:** **Yuan Tian**, and David Lo. Leveraging Web 2.0 for Software Evolution. In *Evolving Software Systems*, pp. 163-197. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.
- US Patent:** Zhizhou Song, Steven H.H. Ding, **Yuan Tian**, Litao Li, Weihan Ou, BlackBerry Ltd, 2024. Method and apparatus for clone search. U.S. Patent Application 18/482,511.

Others

TEACHING

Developed new courses (courses that did not exist before) are marked with a †.

A course code starting with “8” indicates a graduate-level course, while any other starting number denotes an undergraduate-level course.

Courses taught at the School of Computing, Queen's University:

- CISC235 – Data Structures, Winter 2019, Fall 2019, Winter 2021 - 2024
- CISC351/372 CMPE351[†] – Advanced Data Analytics, Winter 2019 - 2024

- CISC843/880[†] – Mining Software Repositories, Winter 2019 - 2021
- CISC839 (for DEBI graduate program)[†] – Topics in Information Systems, Spring 2022 - 2024

Courses taught at the Smith School of Business, Queen's University:

- MDPM822 (for MDPM graduate program)[†] – Emerging Computer Technologies, Fall 2022 - 2023

HQP TRAINING

In my research group, we cultivate an environment where all members are treated equally, and their differences are respected and celebrated. I believe that equity, diversity, and inclusion are essential for effective communication and are fundamental to scientific innovation. Our lab currently has six graduate students from five different countries, three of whom are female. Please note that due to COVID-19, my student hiring process has been delayed, particularly given the challenges of hiring international graduate students.

Current PhDs

Kazi Amit Hasen (co-supervised with Prof. Steven H.H. Ding) 2023.06 - PRESENT

Thesis Topic: *Advanced Pull-based Software Development Leveraging Deep Learning and Data Mining.*

Author of publication items [C33], [C35], [U]19]

Highlight: Amit joined our lab in September 2022 as a research-based master's student and was promoted to the Ph.D track in June 2023. Amit participated in the Mitacs Business Strategy internship in collaboration with LWB.

Jerin Yasmin

2021.09 - PRESENT

Thesis Topic: *Understanding and Improving Programmable Data and Machine Learning Pipelines in Software Systems.*

Author of publication items [C26], [C31], [C37], [U]21]

Highlight: Jerin was awarded R.S McLaughlin Fellowship for 2021-2022 and the Ontario Graduate Scholarship (OGS) for 2022-2023. In collaboration with a software engineering team from Purdue University, Jerin published the largest dataset on the usage of pre-trained models. Jerin also won the Best ERA Paper Award at ICPC'22.

Mohammad Sadegh Sheikhaei

2020.09 - 2024.12 (EXPECTED)

Thesis Topic: *Automated Self-Admitted Technical Debt Tracking, Classification, and Repayment for Sustainable Software Development.*

Author of publication items [C31], [C34], [J8], [J9], [U]20]

Highlight: Sadegh participated in the Mitacs Accelerate internship in collaboration with FutureCite, where he contributed to the development of FutureCite's latest business product.

Current MScs

Cameron Parker

2024.01 - PRESENT

Thesis Topic: *Pragmatic Automated Library Transformation.*

Highlight: Cameron is currently a senior developer at Yelp. His master's degree is supported by Yelp.

Marcos Macedo Contrera

2023.09 - 2024.09 (WILL BE PROMOTED TO PHD IN FALL'24)

Thesis Topic: *Foundation Model (FM)-driven Automated Code Transformation.*

Author of publication items [C35], [C38], [U]19]

Highlight: Marcos won the Vector Scholarship in Artificial Intelligence for the year 2023 - 2024 and the Ontario Graduate Scholarship (OGS) for the year 2024 - 2025. His paper won the Best Research Paper Award at FORGE'24.

Huizi Hao (co-supervised with Prof. Ahmed E. Hassan)

2023.09 - PRESENT

Thesis Topic: *Understanding and Improving the Utilization of LLMs in Software Development Lifecycle.*

Author of publication item [U]19]

Highlight: Huizi won second place in the research poster competition at CSER'24.

I believe supervising undergraduate students on research projects is crucial. It provides them with invaluable hands-on experience, enhances their understanding of theoretical concepts, and fosters their critical thinking and problem-solving skills. Engaging in research projects also prepares students for future academic or industry careers. I primarily work with students on research projects during the summer, hiring them as research assistants. Currently, I am supervising the following 6 students:

Current UGs

- Summer research assistants: Anthony Guglya (since 2023.04), Vu Thanh Loc Mai (since 2024.05), Renee Kim (since 2024.05), Natasha Carneiro (since 2024.06)
- Undergraduate thesis (CISC500) students: Mingjun Gao (since 2024.06), Conor Gallagher (since 2024.04), visiting student from Queen's University Belfast, UK

Graduated MScs

Hong Qin 2022.09 - 2024.05

Msc Report Title: *A Comprehensive Study of Developers' Conversations with ChatGPT Shared in GitHub*.
Highlight: Hong will join Duke University in July 2024 as a Programmer Analyst/Project Manager. She was a qualifying student transferred from non-CS domain during the year 2022 - 2023.

DongJun (Jefferson) Jin 2021.09 - 2022.09

Msc Report Title: *Mining and Analyzing the Cell-based Development and Evolution of Jupyter Notebooks*.
Highlight: DongJun is now a software engineer at Veeva System, a leader in cloud-based software for the global life sciences industry. He participated in the Mitacs Accelerate Explorer internship.

Aidan Polese 2020.05 - 2022.08

Msc Report Title: *Adoption of Third-party Libraries in Mobile Apps: A Case Study on Open-source Android Applications*.
Highlight: Aidan is now a software engineer at Tango, a mobile application company in Toronto.

Julia Liu 2020.09 - 2022.03

Msc Report Title: *Characterizing Discussions on SE bots: A Case Study of OSS projects on GitHub*.
Highlight: Julia is now a software engineer at Alibaba Group, China.

Hanwen Hu 2020.09 - 2021.10

Msc Report Title: *An Empirical Study of Issues in Cross-Platform Video Games on Steam*.
Highlight: Hanwen is now a software engineer at a software company in Toronto.

Jerin Yasmin 2019.09 - 2021.07

Msc Thesis Title: *RESTful API Deprecation: An Empirical Study on Deprecation Practices and a Method for Detecting Deprecated API Usages in Client-side Web Applications*.
Highlight: Jerin is now a PhD in our research lab.

Ka Lai Yung 2019.09 - 2021.05

Msc Report Title: *Beyond Emojis: an Exploratory Study on Pull Request Reactions on GitHub*.
Highlight: Kalai is now a software engineer at a Financial Company in Toronto.

I supervised the following 36 students for their advanced undergraduate projects (CISC499), research projects during summer terms, or special competitions.

Supervised UGs

- Amber Wang (2023.04 - 2024.04). Amber will join the research-based master's program at the University of Waterloo.
- Rabab Azeem, Bartek Kowalski, Kiarash Mirkamandari, Sanindie Silva. (2023.10 - 2024.01). I supervised the team to work on a research project addressing data efficiency challenges associated with fine-tuning pre-trained large language models (LLMs). Through this project, the team achieved 2nd place in 2024's AI ProjectX competition.

- Dennis Huynh, Garrett Audet, Nikolay Alabi (2021.05 - 2021.09). I supervised the team during the summer of 2021 on a research project that predicts stock prices leveraging Reddit. The project resulted in a conference paper at BigData'21.
- Dongjun Jin (2020.08 - 2020.12) via Mitacs Research Training Award. Dongjun completed his master's degree under my supervision.
- Alice Petrov (2020.01 - 2021.01) via the Undergraduate Student Research Awards (USRA) program. Alice is now a master's student in the Mathematics and Foundations of Computer Science program at Oxford University.
- Xiaotian Liu (2019.04 - 2019.12). Xiaotian achieved a research-based master's in AI from Queen's University and is now a PhD student at the University of Toronto.
- CISC 499 students: Xueyi Jia, Erin Atacan, Gaoyuan Bao, Peter Gelgor, Christopher Iwamoto (2024.01 - 2024.04), Runze Liu, Ruifeng Qian, Ruiyang Su, Yuen Zhou (2023.01 - 2023.04), Emily Zhao, Xiaohan Xie, Shuyang Wang, Randy Shao, Mudra Patel, Tom Lu Lu (2022.01 - 2022.04), Xinmeng Wang, Xinrui Li, Jefferson Jin, Jessica Dassanayake (2021.01 - 2021.04), Noah Cabral, Chin Chen, Shikai Liu, Yuanhao Lou (2020.01 - 2020.04), Lian Strathdee, Roman Sokolovski (2019.01 - 2019.04).

Supervisory committee

I am currently serving or have previously served on the Ph.D. supervisory committee for the following 23 students at the School of Computing, Queen's University:

Nazmul Islam (2024.05 - present), Ephraim Govere (2023.09 - present), Emilie Coote (2023.09 - present), Elyas Rashno (2022.09 - present), Christopher Molloy (2022.05 - present), Fangjian Lei (2022.05 - present), Ayesha Babar (2022.09 - present), M Tanvir Kaykobad (2021.09 - present), Mahzabeen Emu (2021.09 - present), Yingzhe Lyu (2021.05 - present), Zhizhou Song (2021.05 - present), Yuhao Chen (2021.09 - present), Nisha Simon (2020.09 - present), Litao Li (2020.09 - present), Aaditya Bhatia (2020.09 - present), Weihan Ou (2020.09 - present), Maram Assi (2019.09 - present), Isra Arjumand (2019.09 - present), Kundi Yao (2018.10 - 2024.01), Anika Anwar (2018.10 - 2023.08), Seyed Farnood Faghihi Ghazvini (2018.10 - 2022.05), Jiayuan Zhou (2019.05 - 2020.09), Haoxiang Zhang (2018.10 - 2020.09).

PROFESSIONAL ACTIVITIES

Conference organization

I have participated in organizing the following 14 events at conferences and workshops:

- Data/Tool Showcase Track Co-chair, the 22nd IEEE/ACM International Conference on Mining Software Repositories (MSR'25).
- Panelist Co-Chair, the Consortium for Software Engineering Research 2024 Spring Meeting (CSER'24).
- Co-Organizer, the 5th Workshop on Human-Centric Software Engineering & Cyber Security (HCSE&CS-2024) co-hosted with the 39th IEEE/ACM International Conference on Automated Software Engineering (ASE'24).
- Vision Breakout Session Co-organizer, the FM+SE Vision 2030 Meeting. The Vision Meeting was held from Nov 16-19, 2023 in Mexico City. The Meeting includes free-form discussions by world-leading industrial and academic experts from many top institutions.
- Session Chair, the 20th IEEE/ACM International Conference on Mining Software Repositories (MSR'23).
- Publicity and Social Media Co-Chair, the 30th IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER'23).
- Diversity, Inclusion and Outreach Co-Chair, the 19th IEEE/ACM International Conference on Mining Software Repositories (MSR'22).
- Session Chair of Newcomer Orientation I, the 19th IEEE/ACM International Conference on Mining Software Repositories (MSR'22)

- Tutorials/Training Track Co-Chair, the 18th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE'22)
- Workshops Co-Chair, the 36th IEEE/ACM International Conference on Automated Software Engineering (ASE'21)
- Session Chair, the 36th IEEE/ACM International Conference on Automated Software Engineering (ASE'21)
- Publicity and Social Media Co-Chair, the 37th International Conference on Software Maintenance and Evolution (ICSME'21)
- Publicity and Social Media Chair, the 35th IEEE/ACM International Conference on Automated Software Engineering (ASE'20)
- General Co-Chair, the Consortium for Software Engineering Research 2019 Spring Meeting (CSER'19)

Since 2023, I have served as a member of the IEEE Computer Society Conference Activities Committee (CAC) within the Technical Conference Activities Board. In this role, I have reviewed applications from 22 conferences.

Research association

(Guest) Editor for journals **I have served, or am currently serving for the following journals:**

- Editor, Neurocomputing (Software Section), 2024
- Guest Editor, Human-Centric Intelligent Systems (HCIN), 2024
- Guest Editor, Automated Software Engineering Journal (AUSE), 2021

Funding evaluation

I have served as a reviewer for the following grants:

- MITACS Accelerate Grant, 2019 - 2024
- NSERC Discovery Grant, 2022, 2023
- NSERC Discovery Grant Peer Review for York's University, 2023
- John R. Evans Leaders Fund (JELF), Canada Foundation for Innovation (CFI), 2021

Journal review

I have served as a reviewer for the following journals:

- IEEE Transactions on Software Engineering (TSE), 2019 - present.
- ACM Transactions on Software Engineering and Methodology (TOSEM), 2022 - present.
- Empirical Software Engineering Journal (EMSE), 2017 - present
- Automated Software Engineering (AUSE), 2019 - present
- Journal of Systems and Software (JSS), 2019 - 2023
- IEEE Transactions on Reliability, 2021 - 2022
- Information and Software Technology Journal (IST), 2020 - 2021
- ACM Transactions on Information Systems (TOIS), 2019

Conference committee

I have served, or am currently serving, as the program committee member or award selection committee member for the following 53 tracks of conferences/workshops:

- For the IEEE/ACM International Conference on Software Engineering (ICSE), i.e., the best SE conference: ICSE'25 Research Track, ICSE'23 ACM Student Research Competition Track, ICSE'21 Technical Track, ICSE'20 ACM Student Research Competition Track.
- For the ACM International Conference on the Foundations of Software Engineering (FSE), a top-tier SE conference: FSE'25 Research Track, FSE'24 Research Track, FSE'23 Test of Time Award Committee, FSE'23 Student Research Competition Track, FSE'23 Ideas, Visions and Reflections Track.
- For the IEEE/ACM International Conference on Automated Software Engineering (ASE), a top-tier SE conference and the best conference in Automated Software Engineering domain: ASE'24 Research Papers Track, ASE'22 Research Papers Track, ASE'22 Student Research Competition Track, ASE'21 Research Papers Track, ASE'21 Student Research Competition Track, ASE'20 Research Papers Track, ASE'20 Late Breaking Results Track, ASE'19 Research Papers Track.
- For the IEEE/ACM 20th International Conference on Mining Software Repositories (MSR), the best in data mining for SE domain: MSR'24 Technical Track, MSR'24 Mining Challenge Track, MSR'24 Registered Reports Track, MSR'24 Industry Track. MSR'23 Technical Papers Track, MSR'21 Registered Reports Track, MSR'21 Data Showcase Track, MSR'20 Data Showcase Track.
- For the IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER), the premier event on the theory and practice of recovering information from existing software and systems: SANER'24 Industrial Track, SANER'24 ERA Track, SANER'24 Research Papers Track, SANER'23 Industry Track, SANER'19 Research Track, SANER'18 ERA Track.
- For the IEEE International Conference on Software Maintenance and Evolution (ICSME), the best conference in software maintenance and evolution: ICSME'24 Registered Reports Track.
- For Internetware, a conference for researchers and practitioners to discuss the trending software technologies in the Internet era: Internetware'24 Research Track, Internetware'20 Intelligent Software Engineering Track.
- For AIware, a conference promotes cross-disciplinary discussions, identifies emerging research challenges, and establishes a new research agenda for the community in the Foundation Model era: AIware'24 Main Track.
- For FORGE, which brings researchers, practitioners, and educators from the AI and Software Engineering community to solve the new challenges in the era of foundation models: FORGE'24 Research Track.
- For the Asia-Pacific Software Engineering Conference (APSEC), a leading regional SE conference: APSEC'24 Technical Track, APSEC'24 SEIP Track, APSEC'18 Technical Track, APSEC'19 SEIP Track.
- For the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA), a top-tier SE conference and the best in the testing domain: ISSTA'23 Tool Demonstrations Track and ISSTA'21 Tool Demonstrations Track.
- For the IEEE International Conference on Software Quality, Reliability, and Security conference (QRS), a well-known conference discussing the best and most efficient techniques for the development of reliable, secure, and trustworthy systems: QRS'23, QRS'20, QRS'19.
- For the IEEE International Symposium on Software Reliability Engineering (ISSRE), the leading conference of software reliability engineering: ISSRE'21.
- Others: Canadian AI'24, CoG'23, WAMA'19, NASAC'19, PROMISE'19, CollabTech'19, EASE'19 Short Papers Track.

Thesis examination

I have served as internal/external examiner, chair, or head representative for the thesis examinations of the following 33 graduate students:

- MSc. thesis examination at Queen's University: Shenyu Zheng (2024.05), Akshat Malik (2024.04), Kasra Khalafi (2024.01), Sana Arastehfar (2023.09), Sachin Velloor Sivasubramanian (2023.09), Ryan Kerr (2023.08), Divya Madhav Kamathdivya (2023.08), Henry van Herk (2023.06), Vignesh Rao (2023.04), Connor Little (2023.01), Martin Woo (2022.07), Victoria Armstrong (2022.04), Brennan Cruse (2022.04), Aawista Chaudhr (2021.10), Yuhao Chen (2021.05), Scarlett Taviss (2021.04), David Nam (2020.10), Akib Hirdoy (2020.04), Md Hasan Ibrahim (2019.09), Mojtaba Bagherzadeh (2019.09), Md Ahasanuzzaman (2019.07).
- Ph.D. thesis examination at Queen's University: Binghui He (2024.06), Kundi Yao (2023.12), Zhendong Sha (2023.12), Jia-Huei Lin (2023.06), Osama Ehsan (2023.05), Armstrong Foundjem (2022.05), Seyed Farnood Faghihi Ghazvini (2022.05), Karim Lounis (2020.12), Yu Zhao (2019.06), Dayi Lin (2019.01), Thomas Vaughan (2018.12).
- Ph.D. thesis examination for Mulong Xie (2023.05) from Australian National University, Australia.

University service

I have served, or am currently serving, on the following committees:

- Queen's Southeast Asia Advisory Council. Queen's University, 2024 - present.
- Graduate Committee. School of Computing, Queen's University, 2019 - present
- PhD Committee. School of Computing, Queen's University, 2019 - present
- Graduate Admission Committee. School of Computing, Queen's University, 2019 - 2022, 2023 - present
- Appointment Committee. School of Computing, Queen's University, 2020 - 2022, 2023 - present
- Graduate Scholarship Selection Committee. Queen's School of Graduate Studies, 2021, 2023
- Queen's NSERC Physical Sciences Canada Graduate Scholarships Selection Committee. Queen's University, 2021
- Distinguished Seminar Series Coordinator. School of Computing, Queen's University, 2019.