He Cheng

he.cheng@usu.edu +1 (607) 304 - 0392 Personal Website He Cheng Serendipity618

Research Interests

I am a **machine learning researcher** specializing in **deep anomaly detection**, including its **interpretability** and **security**. My research focuses on developing explainable and robust anomaly detection models, addressing the black-box nature of deep learning systems while enhancing their security against backdoor attacks. I have experience designing counterfactual, global/local, and prototype-based explanation methods to improve model transparency. Additionally, I have worked on attack strategies tailored for anomaly detection models to expose vulnerabilities in anomaly detection systems. Skilled in Python, PyTorch, and data-driven security analysis, with a strong interest in advancing both the theoretical foundations and practical applications of anomaly detection in high-stakes environments.

Education

Utah State University *Aug 2020 - Dec 2024*
***Ph.D.** in Computer Science*

- Advisor: Dr. Shuhan Yuan
- Dissertation: Interpretable and Robust Deep Anomaly Detection

State University of New York at Binghamton *Aug 2018 - May 2020*
***M.S.** in Electrical and Computer Engineering*

- Advisor: Dr. Jian Li
- Thesis: Consensus-Based Quickest Change Detection in IoT Systems

China University of Petroleum (East China) *Sep 2012 - Jun 2016*
***B.E.** in Mechanical Engineering*

Publications

He Cheng, Depeng Xu, and Shuhan Yuan. "Backdoor Attack against Log Anomaly Detection Models." In *Companion Proceedings of the 2025 ACM Web Conference*, 2025. (Short Paper Track)

He Cheng, Depeng Xu, and Shuhan Yuan. "BadSAD: Clean-Label Backdoor Attacks against Deep Semi-Supervised Anomaly Detection." In *arXiv preprint*, arXiv:2412.13324, 2024.

He Cheng, Depeng Xu, Shuhan Yuan, and Xintao Wu. "Achieving Counterfactual Explanation for Sequence Anomaly Detection." In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2024.

He Cheng and Shuhan Yuan. "Backdoor Attack against One-Class Sequential Anomaly Detection Models." In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2024.

He Cheng, Depeng Xu, and Shuhan Yuan. "Explainable Sequential Anomaly Detection via Prototypes." In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2023.

He Cheng, Depeng Xu, and Shuhan Yuan. "Sequential Anomaly Detection with Local and Global Explanations." In *Proceedings of the IEEE International Conference on Big Data (Big Data)*, 2022.

Xiao Han, **He Cheng**, Depeng Xu, and Shuhan Yuan. "InterpretableSAD: Interpretable Anomaly Detection in Sequential Log Data." In *Proceedings of the IEEE International Conference on Big Data (Big Data)*, 2021.

Research Projects

Backdoor Attacks Against Deep Log Anomaly Detection Models *Sep 2023 - Present*

- Developed a backdoor attack framework for self-supervised log anomaly detection models (DeepLog, LogBERT), enabling abnormal logs to evade detection while maintaining benign performance. The attack involves data poisoning (injecting imperceptible triggers using only normal log entries) and model infection (modifying the training objective to embed the backdoor). (***WWW 2025 (Short Paper Track)***)

- Proposed BadSAD, a clean-label backdoor attack against Deep Semi-Supervised Anomaly Detection (DeepSAD) models, widely used for image anomaly detection. The attack involves trigger injection (embedding subtle triggers in normal images) and latent space poisoning (manipulating feature representations to misclassify triggered anomalies as normal). (***arXiv:2412.13324, 2024***)
- Developed a backdoor attack framework for one-class sequential anomaly detection models (Deep SVDD, OC4Seq), enabling anomalies to evade detection while maintaining benign performance. Designed an attack strategy with trigger generation and backdoor injection. (***PAKDD 2024***)

Developing Inherently Interpretable Deep Anomaly Detection Models *Jan 2022 - Aug 2023*

- Proposed a prototype-based framework for explainable sequential anomaly detection, addressing the lack of interpretability in deep learning models. The method derives multiple prototypes to represent different normal and abnormal sequence patterns, enabling instance-based explanations. It employs contrastive learning and k-means clustering to identify and explain anomalies. (***IJCNN 2023***)
- Proposed GLEAD, a globally and locally explainable anomaly detection method for sequential anomaly detection. GLEAD leverages multi-head self-attention to capture diverse anomaly patterns and derives prototypes for normal and abnormal sequences to enhance explainability. It provides local explanations by highlighting anomalous entries in sequences and global explanations by identifying common anomaly patterns across data. (***IEEE Big Data 2022***)

Explaining One-class Deep Anomaly Detection Models via Counterfactual Explanations *Aug 2020 - Dec 2021*

- Developed CFDet, a counterfactual explanation framework for one-class sequence anomaly detection (Deep SVDD, OC4Seq), identifying anomalous entries by determining the minimum changes needed to make a sequence normal. CFDet enhances interpretability in anomaly detection models, aiding security analysis and reducing false alarms. (***ECML PKDD 2024***)

Consensus-based Quickest Change Detection in IoT Systems *Aug 2019 - May 2020*

- Developed a distributed anomaly detection framework for IoT systems, enabling quickest change detection while minimizing false alarms and communication overhead. Proposed a consensus-based distributed S-CuSum algorithm, achieving performance comparable to centralized methods with reduced information exchange. Designed two optimized variants to further decrease communication costs while maintaining detection accuracy.

Research & Professional Experience

Research Assistant *Logan, UT*
Utah State University *Aug 2020 - Present*

- **End-to-End Machine Learning Systems:** Designed and implemented machine learning pipelines for log anomaly detection using **PyTorch** and **TensorFlow**. Developed **counterfactual explanations**, **global and local interpretability**, and **prototype-based insights** to enhance user trust and identify vulnerabilities in deep anomaly detection systems.

- **Linux Server Management:** Configured **SSH** access and managed distributed servers for **machine learning experiments**. Automated server maintenance using **Bash** scripts, improving uptime and reliability. Implemented **performance monitoring**, **log analysis**, and resource tracking to optimize GPU utilization and ensure efficient model training.

Teaching & Mentorship

Teaching Assistant *Logan, UT*
Department of Computer Science, Utah State University *Aug 2021 - Dec 2024*

- Assisted in teaching and grading coursework for the following courses:
  - **CS 5665** Introduction to Data Science
  - **CS 5080/CS 6665** Data Mining

Academic Service

Reviewer, International Joint Conference on Neural Networks *2025*

Reviewer, Pacific-Asia Conference on Knowledge Discovery and Data Mining *2025*

Reviewer, International Conference on Machine Learning and Applications *2024*

Reviewer, International Joint Conference on Neural Networks *2024*

Reviewer, International Joint Conference on Neural Networks *2023*

IEEE Transactions on Computational Social Systems *Journal Reviewer*

Concurrency and Computation: Practice and Experience *Journal Reviewer*

Intelligent Data Analysis *Journal Reviewer*

Awards

IEEE Big Data 2024 Volunteer Lead *2024*

SDM24 Doctoral Forum Travel Award ($1,000) *2024*

Utah State University Graduate Student Travel Award *2023*

Utah State University Graduate Student Travel Award *2022*

Technical Skills

Programming and Tools: Python, SQL, C/C++, Docker, Git, Linux, PyTorch, TensorFlow, NumPy, Pandas

Cloud and Databases: AWS (EC2, S3, Lambda, SageMaker), GCP, PostgreSQL, MongoDB, SQLite

Machine Learning and System Design: Data preprocessing, feature engineering, hyperparameter tuning, distributed training (multi-GPU), explainable AI, model deployment (Flask, FastAPI, Docker containers), model optimization