Assignment for WASP Software Engineering Course

Anindya Sundar Das

August 2023

1 Q1:Introduction

Anomaly exploration holds great significance in domains such as data mining and machine learning with extensive practical uses in areas like healthcare, finance, and intrusion detection. The rapid advancement of technology, encompassing trends like rise of edge devices, widespread internet usage, and social media, has led to the generation of complex, high-dimensional and multi-faceted data comprising images, text, audio, video, multimodal and diverse formats. Nevertheless, the growing complexity of this data's dimensions presents challenges for detecting anomalies, as the proximity between data objects decreases in high-dimensional space. Additionally, many of such anomaly detection algorithms tasks involves human beings; hence it's legitimate to ask whether we can trust these algorithms for labelling a human being as anomalous as the person could be likely be a victim of the deliberate or accidental victim of the algorithm misuse. In addition to ethical concerns, the vulnerability of deep learning models to manipulation and intrusion poses significant challenges. Adversarial attacks can also exploit weaknesses in data, models, or outputs, compromising the integrity of anomaly detection systems [1]. Furthermore, the scarcity of anomalous instances in training data presents a critical hurdle. As anomalies are infrequent occurrences, the resulting dearth of samples hinders the performance of deep learning models. This limitation becomes particularly pronounced in real-world scenarios with constrained data availability.

We aim to overcome these challenges by integrating core ethical values such as explainability, fairness, and robustness into anomaly detection algorithms. In this project, our focus is address anomaly detection in complex data formats like text, images, and graphs. The project encompasses three primary three main objectives. Initially, it aims to investigate attacking strategies and defenses against different attacks to design robust anomaly detection (AD) [2] models. The second goal revolves around making the models explainable by identifying the features that characterize anomalous instances, thereby localizing the factors contributing to their classification as anomalies. The third objective focuses on addressing the challenges of anomaly detection in limited and incomplete data with high levels of missing values. This includes optimizing training strategies for scenarios where data is scarce, unlabeled, imbalanced, or constrained. By tackling these challenges, the project aims to enhance the trustworthiness, performance, and practicality of deep learning-based anomaly detection methods.

2 Q2

2.1 Requirement Engineering for AI/ML

Requirement Engineering for Machine Learning (RE4ML) refers to the systematic process of gathering, analyzing, documenting, and validating the functional and non-functional needs and constraints of machine learning projects. It is an essential step in the development life-cycle of machine learning systems, akin to traditional software requirement engineering, but tailored to the specific characteristics and challenges of machine learning.

The systematic process of RE4ML can greatly contribute to the success, effectiveness, and ethical alignment of our research in the following ways:

i)**Problem Specification:** It involves identifying and defining the specific types of anomalies that we aim to detect, the data formats we will be working with (text, images, graphs), and the complexity of the targeted anomalies. ii)**Data Specification:** It involves detailing the sources of data, the amount of data needed for training and testing, data quality standards, and potential data augmentation techniques to address scarcity. iii)**Functional Requirements:** It involves defining the functional expectation of the anomaly detection model in terms of minimum accuracy rate, False Positive Rate (FPR), Area Under the Receiver Operating Characteristic curve (AUROC) value the anomaly detection models need to achieve. iv)**Non-Functional Requirements:** Non-functional requirements are essential aspects of RE4ML that define the qualities and attributes that a machine learning system or model should exhibit beyond its core functionality. a) ***Robustness*** ensures that the models can accurately identify anomalies even when faced with noisy or uncertain data. We need to specify a robustness requirement that outlines the acceptable level of noise, data variations, or outliers that the models should handle. b) ***Reliability*** in anomaly detection is crucial for maintaining consistent performance over different datasets and time periods. c) ***Explainability*** requirement mandates the AD models to generate clear explanations for their anomaly predictions. d) ***Maintainability*** guidelines include version control, documentation practices, and modular code structure to ensure that the models can be updated, enhanced, and retrained without compromising their performance. e) ***Safety*** requirements outline measures to prevent harmful consequences resulting from misclassified anomalies. This could involve human oversight, or validation steps before taking any critical actions based on model predictions.

2.2 AI/ML Software Testing

One of the core concept of AI/ML software testing which ensures reliability, correctness, and effectiveness of machine learning models and systems. Some of the key aspects in Software Testing: i) Testing machine learning models presents unique challenges due to the data-centric nature, the non-deterministic behavior of ML systems, and the influence of training data on their performance. Our research targets these aspects by focusing on building AD systems that handles complex data, and performs relatively well in limited data-setup. ii) ***Validation and Verification:*** Validation ensures that the right system is built, while verification ensures that the system is built correctly. Validating anomaly detection models involves ensuring that they are designed appropriately to detect targeted anomalies and provide meaningful insights. While Validation involves checking that the models' design, algorithms, and implementations accurately reflect the intended objectives. iii) ***Robustness Testing:***Testing with noisy, varied, and unexpected data scenarios helps validate the models' ability to accurately detect anomalies even in challenging conditions. iV) ***Interpretability Testing:*** We also emphasize on model explainability, where the generated explanations need to accurately reflect the reasons behind anomaly classifications.

3 Q3.

3.1 AutoML and Anomaly Detection:

Automated Machine Learning (AutoML) [3] can help streamline the development of anomaly detection models and enhance their performance while addressing the challenges associated with complex data in the following way:

i) AutoML can automate the pre-processing of complex data such as text tokenization, image augmentation, and graph feature extraction. ii) It can be used to search for optimal hyperparameters for anomaly detection models. iii) AutoML can generate synthetic data samples to augment the training dataset, addressing the issue of limited data and improving the model's generalization capability. iv) It can be useful in the validation and verification of models by automating the process of checking that models are designed correctly, aligning with the research topic's focus on trustworthiness. v) AutoML can facilitate the ongoing monitoring of model performance, automatically triggering retraining or updates when the model's accuracy or behavior deviates from expectations.

3.2 Boundary Value Testing for ML:

In the context of ML software, Boundary Value Testing (BVT) [4] involves testing the behavior of the ML model at the edges of its input space. This is important because ML models can exhibit different behaviors near the boundaries, and these behaviors might have significant implications for the model's performance and reliability. In the context of anomaly detection, boundary value testing becomes essential for several reasons. Firstly, BVT can help assess the **robustness** of these models by evaluating their behavior at the edges of the input domain. This ensures that the models are capable of detecting anomalies even in challenging scenarios. Secondly, BVT can reveal how **sensitive** the anomaly detection models are to small changes in input data. This is especially important in scenarios where minor fluctuations or variations might lead to different anomaly predictions. Furthermore, evaluating anomaly detection model performance at the edges of the input space helps identify potential issues such as underfitting or overfitting. It ensures that the models are capable of **generalizing** to a wide range of input scenarios.

4 Q4.

4.1 Security & Privacy:

Security and privacy are two important concepts in modern digital era, encompassing safeguards and rights that individuals and organizations rely upon for protection and control over their information and assets. Security involves measures taken to safeguard systems, data, and networks from unauthorized access, attacks, and breaches. It includes practices such as encryption, authentication, and access controls to ensure the confidentiality, integrity, and availability of sensitive information.

On the other hand, privacy pertains to the control individuals have over their personal information. It involves the right to determine what data is collected, how it's used, and who has access to it. Privacy protection ensures that personal details remain confidential and are not exploited without consent. In the digital realm, concerns about privacy are heightened due to the vast amount of data collected and shared online. Privacy measures include data anonymization, consent mechanisms, and adherence to regulations such as GDPR.

Federated Learning (FL) can address some of the interesting research challenges AD systems face related to privacy. For instance, the detection of anomalies in IoT systems is crucial for tasks such as identifying falsified data injections and diagnosing transmission line faults in smart grids. However, Current AD methods for IOT devices still encounter challenges related to efficiency, robustness, and security. In [5], the authors proposed a decentralized and asynchronous FL framework utilizing blockchain technology to mitigate poisoning attacks against IoT anomaly detection models and introduced model with accurate, efficient, secure, and privacy-preserving anomaly detection capabilities.

Conversely,, AD systems could also be useful in detection of backdoored/ compromised models as deliberated in [6], It can also be used to detect backdoor triggers , by thus ensuring security of Machine Learning models as recently addressed in the work [7].

4.2 Quality Assurance:

Quality Assurance (QA) is essentially a method for guaranteeing the quality of software or product. It encompasses a range of actions that ensure that project processes, methods, and criteria are appropriate and properly executed. Quality Assurance (SQA) within software engineering centers on enhancing the software development process to preemptively address problems before they escalate into significant concerns. Software quality assurance centers on ensuring software's usability, reusability, portability, correctness, maintainability, and error control. It strives to optimize these dimensions through testing, process improvements, and adherence to standards, resulting in reliable, user-friendly, adaptable software with minimal errors.

To execute SQA with efficacy, it's vital to adopt a systematic strategy: i) Precisely outline the quality benchmarks the software must adhere to. This encompasses specifying prerequisites, acceptance measures, and performance indicators. ii) Methodically assess software elements like requirements, design documents, and code. iii) Execute various testing methods like unit, integration, system, and acceptance tests. iv) Constantly oversee and assess the software's quality during development. v) Enhance the SQA process by continually analyzing the outcomes of monitoring and assessment endeavors.

In addition to SQA, industrial quality assurance assumes a crucial role in production by enhancing reducing waste, product reliability, and boosting overall efficiency. Manual inspection methods often involve human beings for defect identification and are inefficient, error-prone, and laborious. Hence traditional approaches are being replaced by algorithmic techniques such as anomaly detection methodologies. In one of the recent work [8], researchers used anomaly detection for visual quality assurance. It involves training various models in an unsupervised fashion on a large dataset of intact vehicle identification number labels. These models learn defect-free patterns and identify deviations as potential defects.

5 Q5.

Future software engineering trends involve AI-driven automation, ethical AI integration, and advanced testing techniques. The research topic aligns with these trends by focusing on AI-powered anomaly detection. As software complexity increases, AI will aid in identifying anomalies in diverse data formats. The challenge lies in ensuring transparent and trustworthy AI models. AI's potential for precise anomaly detection will reshape software quality assurance and security protocols, optimizing performance across domains

As a researcher specializing in AI-powered anomaly detection there holds a promising career path within the software industry, including roles as AI/ML engineers, data scientists, or software developers with a focus on quality assurance and security. The expertise aligns with industry demands for AI-driven automation, enhancing software testing, and ensuring ethical AI integration. As companies increasingly adopt AI technologies, these skills will be sought after to develop robust, transparent, and responsible AI systems.

ML/AI Engineering will evolve significantly in the next 5-10 years. It will shift from focusing solely on model performance to emphasizing ethical AI, interpretability, and domain-specific expertise. As AI technologies become more integrated into daily life, engineers will prioritize responsible AI design, transparency, and fairness. Edge computing will rise, demanding efficient models for resource-constrained devices. Collaboration between humans and AI will increase, requiring engineers to design systems that enhance human capabilities. Overall, ML/AI Engineering's importance will center on ethical, practical, and human-centered AI solutions.

References

[1] Anirban Chakraborty, Manaar Alam, Vishal Dey, Anupam Chattopadhyay, and Debdeep Mukhopadhyay. Adversarial attacks and defences: A survey. *arXiv preprint arXiv:1810.00069*, 2018.

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.

[3] Xin He, Kaiyong Zhao, and Xiaowen Chu. Automl: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212:106622, 2021.

[4] Hong Zhu and Ian Bayley. Discovering boundary values of feature-based machine learning classifiers through exploratory datamorphic testing. *Journal of Systems and Software*, 187:111231, 2022.

[5] Lei Cui, Youyang Qu, Gang Xie, Deze Zeng, Ruidong Li, Shigen Shen, and Shui Yu. Security and privacy-enhanced federated learning for anomaly detection in iot infrastructures. *IEEE Transactions on Industrial Informatics*, 18(5):3492–3500, 2021.

[6] Alexander Unnervik and Sébastien Marcel. An anomaly detection approach for backdoored neural networks: face recognition as a case study. In *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2022.

[7] Hao Fu, Akshaj Kumar Veldanda, Prashanth Krishnamurthy, Siddharth Garg, and Farshad Khorrami. Detecting backdoors in neural networks using novel feature-based anomaly detection. *arXiv preprint arXiv:2011.02526*, 2020.

[8] Justus Zipfel, Felix Verworner, Marco Fischer, Uwe Wieland, Mathias Kraus, and Patrick Zschech. Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. *Computers & Industrial Engineering*, 177:109045, 2023.