Engineering Village™

1. **Explainable AI for Android Malware Detection: Towards Understanding Why the Models Perform So Well?**
Liu, Yue; Tantithamthavorn, Chakkrit; Li, Li; Liu, Yepang **Source:** *arXiv*, September 2, 2022; **E-ISSN:** 23318422; **DOI:** 10.48550/arXiv.2209.00812; **Publisher:** arXiv

Author affiliation:
Faculty of Information Technology, Monash University, Melbourne, Australia
Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

Abstract:
Machine learning (ML)-based Android malware detection has been one of the most popular research topics in the mobile security community. An increasing number of research studies have demonstrated that machine learning is an effective and promising approach for malware detection, and some works have even claimed that their proposed models could achieve 99% detection accuracy, leaving little room for further improvement. However, numerous prior studies have suggested that unrealistic experimental designs bring substantial biases, resulting in over-optimistic performance in malware detection. Unlike previous research that examined the detection performance of ML classifiers to locate the causes, this study employs Explainable AI (XAI) approaches to explore what ML-based models learned during the training process, inspecting and interpreting why ML-based malware classifiers perform so well under unrealistic experimental settings. We discover that temporal sample inconsistency in the training dataset brings over-optimistic classification performance (up to 99% F1 score and accuracy). Importantly, our results indicate that ML models classify malware based on temporal differences between malware and benign, rather than the actual malicious behaviors. Our evaluation also confirms the fact that unrealistic experimental designs lead to not only unrealistic detection performance but also poor reliability, posing a significant obstacle to real-world applications. These findings suggest that XAI approaches should be used to help practitioners/researchers better understand how do AI/ML models (i.e., malware detection) work-not just focusing on accuracy improvement.

Main Heading: Android malware **Controlled terms:** Android (operating system) - Chemical detection - Classification (of information) - Machine learning - Mobile security - Statistics

Uncontrolled terms: Android malware - Android malware detection - Android securities - Detection performance - Explainable AI - Machine learning models - Machine-learning - Malware detection - Malwares - Optimistics

Classification Code: 716.1 Information Theory and Signal Processing - 723 Computer Software, Data Handling and Applications - 723.2 Data Processing and Image Processing - 723.4 Artificial Intelligence - 801 Chemistry - 903.1 Information Sources and Analysis - 922.2 Mathematical Statistics

Database: Compendex