

How does the inference efficiency (throughput, latency) of pruned BERT models compare to quantization-optimize

Assignee Research

May 29, 2026

Abstract

In this paper, we present a comparative analysis of benign and malicious Android applications, based on static features. In particular, we focus our attention on the permissions requested by an application. We consider both binary classification of malware versus benign, as well as the multiclass problem, where we classify malware samples into their respective families. Our experiments are based on substantial malware datasets and we employ a wide variety of machine learning techniques, including decision trees and random forests, support vector machines, logistic model trees, AdaBoost, and ar

1 Introduction

This paper examines: A Comparative Analysis of Android Malware. Research question: How does the inference efficiency (throughput, latency) of pruned BERT models compare to quantization-optimized random forest classifiers for malware detection on resource-constrained Android devices?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

16 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 4.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The Android Malware Genome Project dataset consists mainly of apk files obtained from various malware forums and Android	×	0.06
The dataset includes labels specifying the family to which each sample belongs.	×	0.06
The benign dataset was crawled from the Play-Drone project.	×	0.03
Androguard was used to filter broken and potentially malicious apk files from the benign dataset.	×	0.04
The number of malware samples obtained is 989 for detection and 1260 for classification.	×	0.07
The number of benign samples obtained is 2657.	×	0.06
There are 230 distinct permissions found in the superset of malware and benign samples.	×	0.07
For binary classification, a feature vector of length 230 is generated based on permissions.	×	0.11
For multiclass classification, a feature vector of length 118 is generated based on permissions.	×	0.10
Many of the features under consideration (i.e., permissions) provide little or no discriminating information.	×	0.04
Removing non-discriminating features can improve classification performance.	×	0.08
The accuracy of the binary classification model is 100% for some experiments.	×	0.07
The accuracy of the multiclass classification model is 98% for some experiments.	×	0.06

References

- <http://arxiv.org/abs/1904.00735v1>

- <http://arxiv.org/abs/1811.01027v2>
- <http://arxiv.org/abs/2104.09994v3>