Sub-10B Model Accuracy in Mixed-Language Vulnerability Detection After Python and Java Fine-Tuning

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does the accuracy of sub-10B models fine-tuned on Python code compare to models fine-tuned on Java code when evaluated on mixed-language vulnerability detection benchmarks. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Case Study: Fine-tuning Small Language Models for Accurate and Private CWE Detection in Python Code. Research question: How does the accuracy of sub-10B models fine-tuned on Python code compare to models fine-tuned on Java code when evaluated on mixed-language vulnerability detection benchmarks?.

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 3.7/10.

3 Results

16 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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.

References

- http://arxiv.org/abs/2601.08691v1
- http://arxiv.org/abs/2504.16584v1
- http://arxiv.org/abs/2508.11281v3