

DeepSeek R1 and Codestral Code Repair Success Under Aligned Vulnerability Taxonomies

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: What is the impact of vulnerability taxonomy alignment on the code repair success rates of DeepSeek R1 versus Codestral on the Big-Vul dataset. Many ML-based approaches have been proposed to automatically detect, localize, and repair software vulnerabilities. While ML-based methods are more effective than program analysis-based vulnerability analysis tools, few have been integrated into modern IDEs, hindering practical. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: AIBugHunter: A Practical Tool for Predicting, Classifying and Repairing Software Vulnerabilities. Research question: What is the impact of vulnerability taxonomy alignment on the code repair success rates of DeepSeek R1 versus Codestral on the Big-Vul dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.7/10.

3 Results

9 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 2.7/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The AIBugHunter tool uses a NVIDIA GTX 3090 GPU with 24GB of memory for its experiments.	×	0.07
The AIBugHunter approach achieves a multiclass accuracy of 65% for CWE-ID classification.	×	0.05
The CodeBERT model achieves a multiclass accuracy of 59% for CWE-ID classification.	×	0.01
The BERT-base model achieves a multiclass accuracy of 50% for CWE-ID classification.	×	0.01
The BoW+RF model achieves a multiclass accuracy of 27% for CWE-ID classification.	×	0.01
The Naive Bayes model achieves a multiclass accuracy of 27% for CWE-ID classification.	×	0.01
The AIBugHunter approach achieves a multiclass accuracy of 74% for CWE-Type classification.	×	0.05
The BERT-base model achieves a multiclass accuracy of 71% for CWE-Type classification.	×	0.01
The CodeBERT model achieves a multiclass accuracy of 72% for CWE-Type classification.	×	0.01
The BoW+RF model achieves a multiclass accuracy of 52% for CWE-Type classification.	×	0.01
The Naive Bayes model achieves a multiclass accuracy of 51% for CWE-Type classification.	×	0.01

References

- <http://arxiv.org/abs/2305.16615v1>
- <http://arxiv.org/abs/2506.11697v1>
- <http://arxiv.org/abs/2601.08691v1>