

# DeepSeek R1 Accuracy Variations in Vulnerability Classification Across Programming Languages

Assignee Research

June 4, 2026

## Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: How does the accuracy of Deepseek R1 in vulnerability classification vary across different programming languages when evaluated on a standardized dataset like Big-Vul. 11 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. 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: How does the accuracy of Deepseek R1 in vulnerability classification vary across different programming languages when evaluated on a standardized dataset like Big-Vul?.

## 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 3.5/10.

## 3 Results

16 papers retrieved. 11 claims extracted; 0 independently verified. Quality review score: 3.5/10.

## 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.	×	0.06
The AIBugHunter approach achieves 65% multiclass accuracy for CWE-ID classification.	×	0.06
The BERT-base model achieves 59% multiclass accuracy for CWE-ID classification.	×	0.04
The CodeBERT model achieves 50% multiclass accuracy for CWE-ID classification.	×	0.01
The BoW+RF model achieves 27% multiclass accuracy for CWE-ID classification.	×	0.01
The Naive Bayes model achieves 27% multiclass accuracy for CWE-ID classification.	×	0.01
The AIBugHunter approach achieves 74% multiclass accuracy for CWE-Type classification.	×	0.06
The BERT-base model achieves 71% multiclass accuracy for CWE-Type classification.	×	0.04
The CodeBERT model achieves 72% multiclass accuracy for CWE-Type classification.	×	0.02
The BoW+RF model achieves 52% multiclass accuracy for CWE-Type classification.	×	0.02
The Naive Bayes model achieves 51% multiclass accuracy for CWE-Type classification.	×	0.02

## References

- <http://arxiv.org/abs/2305.16615v1>
- <http://arxiv.org/abs/2511.01941v1>
- <http://arxiv.org/abs/2504.00016v1>