

# Multimodal Context Enhances LLM Vulnerability Detection on Big-Vul

Assignee Research

May 30, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the inclusion of multimodal context (e.g., commit messages, code diffs) affect the vulnerability detection accuracy of LLMs compared to text-only file context on the Big-Vul dataset. Detecting vulnerability fix commits in open-source software is crucial for maintaining software security. To help OSS identify vulnerability fix commits, several automated approaches are developed. 17 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Code Change Intention, Development Artifact and History Vulnerability: Putting Them Together for Vulnerability Fix Detection by LLM. Research question: How does the inclusion of multimodal context (e.g., commit messages, code diffs) affect the vulnerability detection accuracy of LLMs compared to text-only file context on the Big-Vul dataset?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.0/10.

## 3 Results

12 papers retrieved. 17 claims extracted; 0 independently verified. Quality review score: 3.0/10.

## 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 evaluation metrics used are precision, recall, F1-score, and Matthews correlation coefficient (MCC).	×	0.04
Accuracy is avoided as an evaluation metric due to the highly imbalanced nature of vulnerability fix detection datasets.	×	0.09
LLM4VFD is compared with three PLM-based techniques: VulFixMiner, CoLeFunda, and VulCurator.	×	0.08
CoLeFunda’s implementation is limited to the Java language.	×	0.03
The VulFixMiner dataset is not reused to avoid data leakage.	×	0.02
All models are evaluated using the newly collected dataset BigVulFixes.	×	0.03
LLMs are compared under their vanilla setting without LLM4VFD, defined as the 0-shot CoT setting.	×	0.04
The task instruction for the 0-shot CoT setting is simplified to: 'Determine whether the current patch is intended to fi	×	0.06
An ablation study is conducted to evaluate the individual impact of each component in LLM4VFD.	×	0.03
The ablation study selected two models, Qwen2-72B and Qwen-7B, from the Qwen series.	×	0.02
The Code Change Summary includes sections on Refactoring, Class Adjustments, Purpose of the Change, and Implications of	×	0.06
The Purpose of the Change section addresses Bug Fixes & Enhancements.	×	0.04
The Implications of the Change section mentions Reduced Errors.	×	0.03
The Summary of the report includes sections on Minor Updates and Purpose of the report.	×	0.02
The Purpose of the report section mentions Backporting.	×	0.02
The Implications of the report section mentions Improvement Documentation.	×	0.02
Historical Vulnerability info includes CVE description.	×	0.05

## References

- <http://arxiv.org/abs/2501.14983v1>
- <http://arxiv.org/abs/2509.17337v1>
- <http://arxiv.org/abs/2404.02056v3>