Qeios

Peer Review

Review of: "MVD: A Multi-Lingual Software Vulnerability Detection Framework"

Alexandra Sheykina[1]

1. University of Salerno, Italy

#Paper summary

This article explores the development of MVD, a multi-lingual software vulnerability detection framework designed to overcome the limitations of single-language models. Existing approaches predominantly focus on individual programming languages, overlooking the potential benefits of cross-language knowledge. MVD addresses this gap by learning from a curated dataset of over 11,000 real-world vulnerabilities across six languages: Python, Java, C/C++, C#, TypeScript, and JavaScript. The framework integrates incremental learning, allowing it to adapt to new languages without requiring previous training data. It employs CodeBERT to extract syntactic and semantic features, enhancing detection accuracy. Additionally, a multi-class classifier predicts both the presence of vulnerabilities and the programming language. The article uses a custom loss function for class imbalance. Experimental results show that MVD outperforms state-of-the-art models by 83.7% – 193.6% in PR-AUC.

#Strengths

- The topic is interesting and sufficiently original.
- The paper is well-written and well-structured.
- The research questions are clearly defined.
- Threats to external and internal validity are discussed.
- The analysis artifact is provided.

#Weaknesses

doi.org/10.32388/L20839

- Imbalance techniques are not applied.
- There are no box plots.

#Evaluation summary

The paper flows well and is written in clear, effective English. The topic is interesting and relevant, addressing the complex challenge of detecting vulnerabilities in polyglot software environments. The paper's novelty in combining cross-language knowledge with incremental learning and leveraging CodeBERT for feature extraction contributes to the field. The paper ensures replicability through the availability of a replication package, allowing independent verification of results.

The paper presents a well-organized and thorough exploration of MVD, a multi-lingual vulnerability detection framework. The presentation is effective, with a logical structure that makes the methodology and findings easy to follow. The inclusion of tables comparing MVD’s performance with existing models is valuable, but adding box plots would enhance result interpretation, providing a more immediate visual comparison.

Additionally, addressing data imbalance through balancing techniques like oversampling or undersampling, and comparing the results with the custom loss function used by the authors, could further improve the framework’s robustness and the overall soundness of the study.

Overall, the paper makes a significant contribution to the research area, with minor improvements that could further strengthen its impact and completeness.

Declarations

Potential competing interests: No potential competing interests to declare.