SCReeD Dataset Declaration Form

Dataset name:*   Vulnerability-introducing/fixing commits in Java for vulnerability assessment

Dataset version:*   1.0

Dataset URL:*   https://github.com/CSCRC-SCREED/ UoA_java-vuln-commits

Creation date:*   19 July 2021

Last update:*   19 July 2021

Author(s):*   Triet Huynh Minh Le, David Hin, Roland Croft, M. Ali Babar

Author(s) affiliation(s):   University of Adelaide, Cyber Security Cooperative Research Centre

Author contact(s):*   triet.h.le@adelaide.edu.au, david.hin@adelaide.edu.au, roland.croft@adelaide.edu.au, ali.babar@adelaide.edu.au

Keywords:*   Software Vulnerability, Vulnerability Assessment, Deep learning, multi-task learning, mining software repositories, software security.

Description/background:*   A dataset with details about vulnerability-causing and fixing commits.

Dataset funding:

Attribute details:*   CSV and parquet files, commits and related changes.

Intended target (if specified):

Format:*   CSV

License:*   Open Access

Standard compliance:

Type:*   CSV and parquet files

Size:*   164 KB (for 1,229 VCCs gathered in java_vccs.csv) and approx. 400 MB for preprocessed data in parquet files.

Availability:

Data status:

Data provenance:*   All data was collected from open sources on the internet.

Source computing infrastructure:*   Live system

Accompanying program(s)/script(s):

Software installer or VM for replication:

Generated or captured via:*   Software

Category/categories:*   ML training data

Published in: DeepCVA: Automated Commit-level Vulnerability
Assessment with Deep Multi-Task Learning.

Open research question(s) (if any):

Potential use case(s) or application area(s): Vulnerability Assessment

Data access control:*   Global access

Data retention period:*   20 years

Data validation/checksum:* N/A

GDPR compliance:*   No

Consent: N/A

Ethics approval: N/A

Ethics considerations: N/A

☒   I confirm that I have read, understood, and agreed to the submission guidelines, policies, and
     submission declaration.

☒   I confirm that the contributors of the dataset have no conflict of interest to declare.

☒   I agree to take public responsibility for my dataset's contents.

 Signature of Corresponding Author   Date:
(signed on behalf of all contributors)