



SCReeD Dataset Declaration Form

Dataset name:* Vulnerable functions in C/C++ for vulnerability detection

Dataset version:* 1.0

Dataset URL:* <https://github.com/CSCRC-SCREED/vulnerable-functions-c-cpp>

Creation date:* 23 May 2023

Last update:* 23 May 2023

Author(s):* Roland Croft, M. Ali Babar, M. Mehdi Kholoosi

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

Author contact(s):* roland.croft@adelaide.edu.au, ali.babar@adelaide.edu.au, mehdi.kholoosi@adelaide.edu.au

Keywords:* Software Vulnerability, Data Quality, Machine Learning

Description/background:* A dataset that can be used to train models to detect vulnerabilities in C/C++ code.

Dataset funding:

Attribute details:* CSV data, Vulnerable C/C++ functions

Intended target (if specified):

Format:* CSV

License:* Open Access

Standard compliance:

Type:* CSV files

Size:* 273 MB

Availability:

Data status:

Data provenance:* The datasets are processed versions of those already available openly.

Source computing infrastructure:* Simulated environment

Accompanying program(s)/script(s):

Software installer or VM for replication:

Generated or captured via:* Software

Category/categories:* ML training data

Published in: Data Quality for Software Vulnerability Datasets

Open research question(s) (if any):

Potential use case(s) or application area(s): Development of vulnerability detection models.

Data access control:* Global access

Data retention period:* 20 Years

Data validation/checksum:* N/A

GDPR compliance:* Yes

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: M. Ali Babar

(signed on behalf of all contributors)



Date: 05/03/2024