Evaluating a Novel Voting System for Malware Dataset Curation with Siamese Networks

Sahil Singh, Stones Dalitso Chindipha

Rhodes University,Department of Computer Science

Grahamstown, South Africa

g21s6211@campus.ru.ac.za, s.chindipha@ru.ac.za

***Abstract*—Replicating research findings is among the most substantial evidence that a scientific claim reflects a real phenomenon rather than chance, bias, or idiosyncrasies of a single laboratory. In malware security, however, replication is especially difficult and is widely regarded as one of the most challenging areas for reproducible research. Those difficulties arise from the domain itself: malware is inherently hostile, constantly evolving, and entangled with legal and ethical constraints that limit data sharing and experimental control. To address these barriers, this study developed a novel malware voting system that aggregates multiple analysis sources to produce robust labels and, in doing so, generated a new, carefully curated dataset. Using that dataset, we successfully replicated the malware classification experiment originally proposed by Hsiao et al. (2019), replicating key performance patterns and verifying central claims about classifier behaviour. To validate the voting system and dataset, we verified visual cluster consistency using perceptual hashing techniques and measured classifier generalisation under varying N-way constraints. The results demonstrate that our voting approach yields consistent labels that enable faithful replication while mitigating some domain-specific obstacles. This work therefore provides both a practical tool for future studies and empirical evidence that replication in malware research, though difficult, can be achieved with thoughtful methodology and rigorous safeguards.**

***Keyword*—Average Hash, Malware Image Classification, One-Shot Learning, Siamese Convolutional Neural Networks.**

Sahil Singh received the B.Sc. (Hons.) degree in computer science from Rhodes University, Grahamstown, South Africa, in 2025.

He has been recognized for academic merit and technical innovation. His work includes research in machine learning, with a particular interest in siamese networks and one-shot learning.

Mr. Singh was a recipient of the TeenTech Gold Award in 2017 and was selected to exhibit his work at The Royal Society in London.

Stones Dalitso Chindipha received his BSC honours in Computer Science and Applied Statistics from University of Malawi, MSc and PhD in Computer Science from Rhodes University in 2012, 2018 and 2022 respectively.

From 2013 - 2015 he worked as a database administrator for a revenue authority organisation before pursuing his MSc which was immediately followed by PhD. He was appointed lecturer in 2020 before being promoted to Senior lecturer in 2024. His research interests include Malware analysis, cryptography and cybersecurity analytics.