Malware continues to increase in prevalence and sophistication.
Traditional detection mechanisms including antivirus software fail to adequately detect new and varied malware.
Artificial Intelligence and Machine Learning (AI/ML) provide advanced capabilities that can be applied to the cybersecurity domain to enhance detection and classification of malware and malware families. Building a robust and automated artificially intelligent malware analysis framework and producing new, standardized malware datasets for future classification in AI/ML, however, are not trivial.

The principal objectives of this research are to deliver a new malware analysis framework, create a new ML heterogeneous classification model to analyze malware, produce uniform datasets for additional AI/ML analysis, and increase classification accuracy across complete and highly diverse malware corpora.

To our knowledge, this is the first research that:
- Releases a complete framework that automates malware analysis and produces new and standardized malware description datasets (ready for AI/ML analysis)
- Develops a classification ensemble applied to malware analysis (named the Malware Ensemble Classification Facility) that improves traditional ML model selection.
- Releases trained ensemble classifier providing enhanced classification results of an entire 200+ gigabyte, malware family corpus consisting of 80K+ unique malware samples and 70+ malware families.

18 ML Classifications Models Trained → Random Forest had best score

A new framework has been developed that currently automates static malware analysis, performs feature extraction, and standardizes malware dataset production for AI/ML analysis. 18 ML classification models were trained on a sampling (3100 binaries representing 111 unique malware families) of our entire malware corpus of 80K+ binaries. Random Forests produced the best performance, however, we identify models with superior classification performance across discrete malware families. Future work includes enriching dataset production and AI/ML analysis and completing development of our new ensemble classifier named the Malware Ensemble Classification Facility.