Threat Based Vulnerability Management: Mapping CVEs to the Mitre ATT&CK Framework

Logan McMahon ○ Oluwafemi Olukoya

School of Electronics, Electrical Engineering And Computer Science (EEECS) ○ Queen's University Belfast

About the Presenter

Logan McMahon.

- BSc Computer Science Graduate from Queen's University Belfast.

- Current Software Engineer I at Rapid7, Belfast.

www.linkedin.com/in/logan-mcmahon-419b77255

lmcmahon25@qub.ac.uk

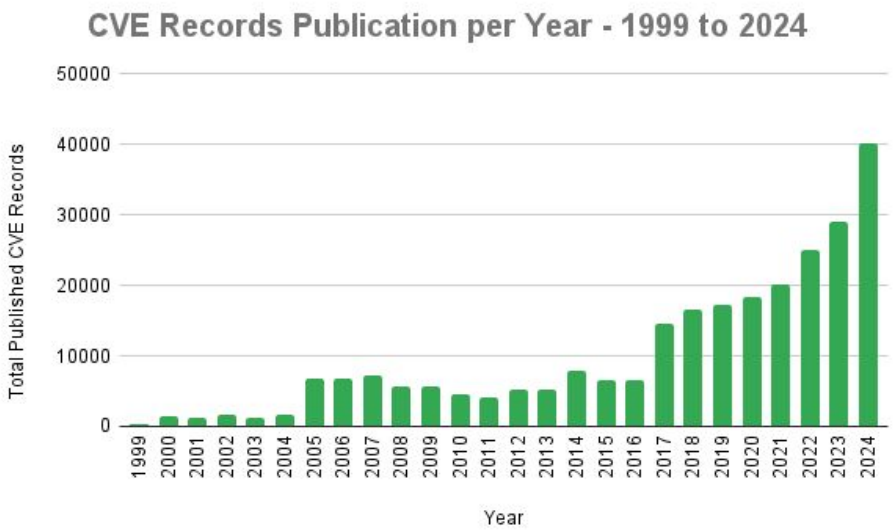
Motivation - Problem Statement

- In 2024 there were a total of 40,077 new CVEs, according to CVE.org which is an increase of 39% from 2023.

- Studies show that companies, each month can only remediate around 10-15% of open Vulns, leaving a persistent backlog.

- As it is not possible to remediate all vulnerabilities, there is a clear need for risk-based vulnerability prioritization.

- Therefore, this approach seeks to further threat-informed defense by mapping CVEs to Mitre ATT&CK, enabling defenders to prioritise High Impact attacks.



Why Mitre ATT&CK?

- Mapping CVEs to MITRE ATT&CK allows defenders to better understand the potential impact of vulnerabilities in the context of adversarial tactics and techniques, which enables better risk-based prioritisation of vulnerability remediation efforts.

- This mapping would enable companies to shift from reactive to proactive vulnerability remediation, preventing MCEs and exploitation by better understanding the attack surface of new CVEs.

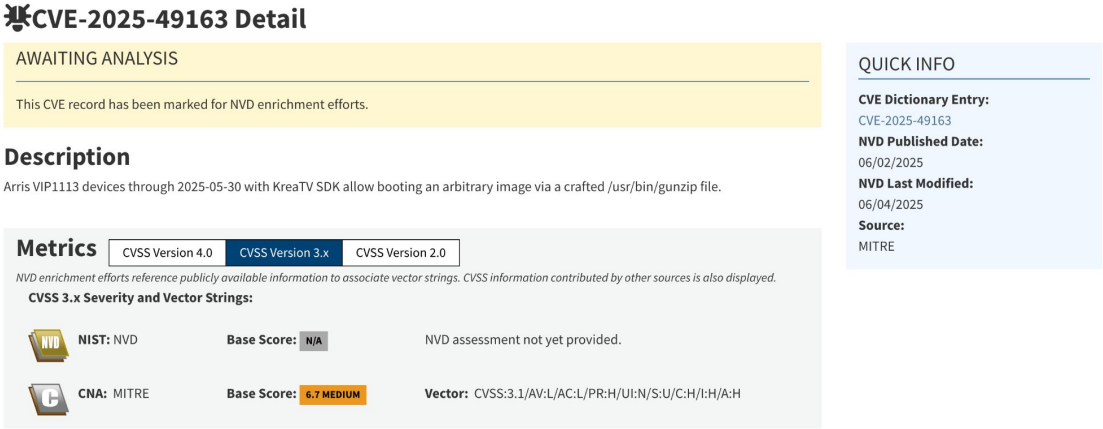
Research Gap

The research conducted in my paper found:

1. Prior works rely heavily on CVE Descriptions.
    a. These approaches struggle when descriptions are incomplete, poorly written, or lack sufficient detail about exploitation methods.

2. Unsupervised methods, while not requiring labeled data, struggle with sparse or ambiguous descriptions because they don't leverage structured attributes like CVSS, CWE, or CPE.

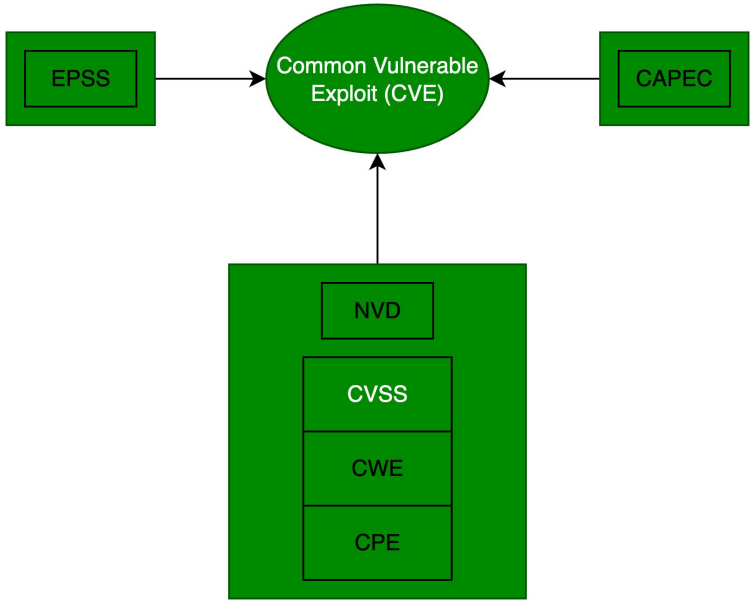
3. Existing supervised methods are constrained by a lack of comprehensive and consistent annotations.



Example of a Poor CVE Description: https://nvd.nist.gov/vuln/detail/CVE-2025-49163

Research Contributions

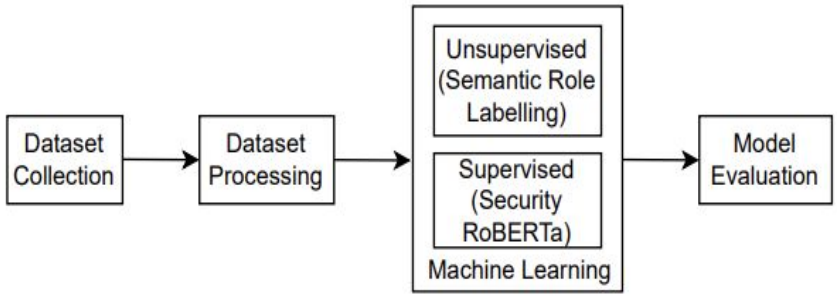
- This paper proposes a comprehensive approach to address the shortcomings of prior works by integrating structured vulnerability data.
- As such, we created an enriched dataset for supervised learning by incorporating structured data from NVD, CAPEC, and EPSS, and performed systematic feature evaluation, showing the impact of added features on model performance.
- Hyperparameter fine tuning was applied to the model which led to significant performance increases.
- We released all datasets, code, and supplementary materials to support reproducibility.



Methodology Overview

The Methodology Proposed by this Paper consists of Four Main Phases:

1. Dataset Collection.
2. Dataset Processing.
3. Mapping (Unsupervised and Supervised approaches)
4. Performance Evaluation.



An overview of the proposed framework for automated mapping of CVEs to MITRE ATT&CK Tactics.

Dataset Collection

- This paper’s initial Supervised dataset was sourced from prior work which included CVE IDs and Descriptions.

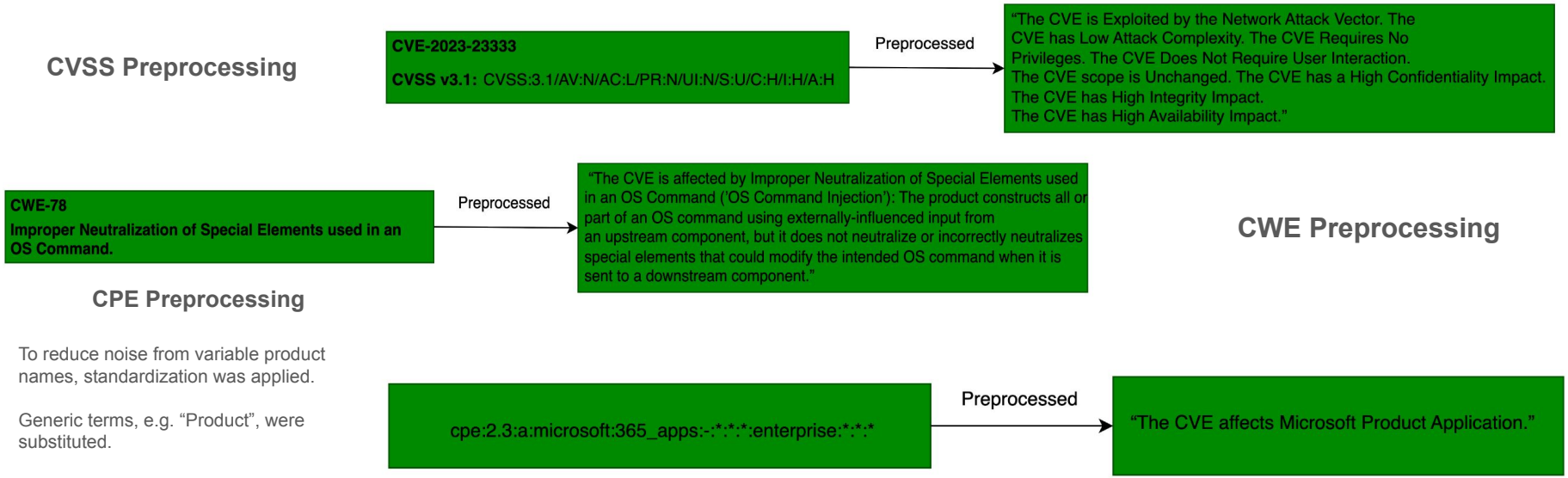
- This Initial Dataset was then enriched with a myriad of Data Features.

- The Final Dataset was extended to include:
  - CWE
  - CVSS
  - CPE
  - EPSS

- The supervised dataset was reduced from 9,986 to 7,328 entries after filtering out CVEs lacking sufficient attributes.

Feature Enrichment

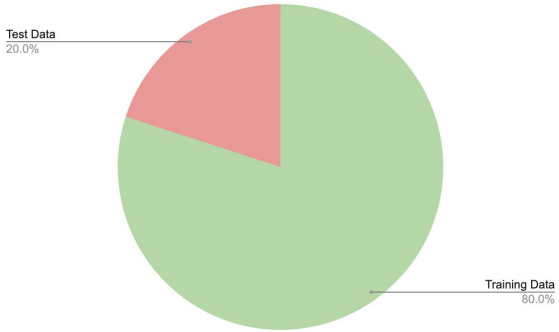
- A key design decision was to enrich the CVE dataset with structured fields to improve mapping accuracy.
- All extended features (except EPSS scores) were pre-processed into natural language format to ensure consistency.



Machine Learning Setup

- While we conducted some research into Unsupervised Methods, ultimately we utilized a supervised approach with SecRoBERTa, a transformer-based model fine-tuned on cybersecurity-specific corpus.

- This was framed as a multi-label classification problem for mapping to ATT&CK tactics.



- The dataset was split into an 80/20 train-test split, consistent with best practices from prior work.

- We used Optuna to fine-tune key hyperparameters, including the learning rate and dropout rate, to optimize model performance.

OPTUNA

Evaluation Metrics

We evaluated our models using several key metrics: accuracy, validation loss, macro F1 score, and weighted F1 score:

Accuracy: Overall proportion of correct predictions.

Validation Loss: Indicator of generalization performance (lower is better).

Macro F1 Score: Unweighted average of per-tactic F1 scores, emphasizing performance on less frequent classes.

Weighted F1 Score: Our primary metric, accounting for class imbalance by weighting each tactic's F1 score by its frequency.

Supervised Dataset Variant	Validation Loss	Accuracy	Macro F1 Score	Weighted F1 Score
Description Only	0.0747	0.8286	0.7948	0.9232
Description + EPSS	0.0729	0.8335	0.8138	0.9277
Description + CWE	0.0724	0.8407	0.7979	0.9248
Description + CVSS	0.0815	0.8229	0.8024	0.9163
Description + CPE	0.0746	0.8286	0.8050	0.9244
Description + CAPEC	0.0870	0.8179	0.7119	0.9011
Fully Extended (Description + EPSS + CVSS + CPE)	0.0743	0.8383	0.8144	0.9245
Fully Extended + Tuned	**0.0658**	**0.8538**	**0.8401**	**0.9347**

Overall Performance across Supervised Dataset Variants

Supervised Machine Learning Approach Results

- Our fully extended and Optuna-tuned model significantly outperformed prior work, achieving a weighted F1 score of 93.47%.

- Hyperparameter tuning yielded a consistent performance boost of 2%-3%.

Tactics	Description only (Benchmark)	Full Extended Dataset (+EPSS+CWE +CVSS+CPE)	+ Optuna Fine-Tuning	SOTA [14]
Reconnaissance	37.33%	36.73%	**46.15%**	53.84%
Resource Development	51.47%	65.81%	**65.79%**	79.13%
Initial Access	65.27%	61.52%	**67.44%**	37.18%
Execution	89.56%	89.25%	**89.95%**	74.43%
Persistence	94.42%	94.52%	**94.87%**	80.78%
Privilege Escalation	94.66%	94.90%	**95.11%**	80.46%
Defense Evasion	98.67%	98.00%	**98.41%**	91.96%
Credential Access	84.82%	89.39%	**91.81%**	67.27%
Discovery	97.24%	97.29%	**97.92%**	81.55%
Lateral Movement	92.87%	94.59%	**94.97%**	81.37%
Collection	79.44%	81.82%	**84.34%**	51.47%
Command & Control	95.21%	95.81%	**96.43%**	61.79%
Exfiltration	64.23%	74.83%	**81.01%**	88.88%
Impact	67.57%	65.73%	**72.00%**	31.11%

Per-Class F1 Scores across Supervised Dataset Variants

Discussion

- Our performance gains are attributed to the combination of structured feature enrichment and effective Optuna hyperparameter tuning.
- Prior work identified four tactics as 'hard' with F1 scores below 60%.
- While the prior work doesn't explicitly state why, data suggests it's due to a lack of descriptive text and features in the original dataset for these tactics.
- Our work proves this hypothesis by moving most Tactics into the medium and easy categories.
- High-quality descriptions and enriched data are essential for enabling automated systems to aid in vulnerability prioritization and response.

Difficulty Level	ATT& CK Tactics in Branescu et al. [14]	ATT&CK Tactics in our proposed approach
Hard	Reconnaissance, Collection, Initial Access, Impact	Reconnaissance
Medium	Resource Development, Credential Access, Execution, Command & Control	Resource Development, Initial Access, Impact
Easy	Privilege Escalation, Discovery, Persistence, Exfiltration, Defense Evasion, Lateral Movement	Privilege Escalation, Discovery, Persistence, Exfiltration, Defense Evasion, Lateral Movement, Execution, Credential Access, Collection, Command & Control

Difficult to Map Mitre ATT&CK Tactic categories

Limitations

- The dynamic nature of the ATT&CK framework.
- Data bias from excluding CVEs without extended fields, which can omit zero-day threats.
- The Reconnaissance tactic remains difficult to predict, despite improvements.
- Poor representation of the CAPEC feature in the current dataset.

Future Work

- Developing a CAPEC API to improve data integration.
- Extending the approach to map to ATT&CK techniques.
- Exploring the use of machine learning to predict and augment missing key aspects of CVE entries.

Conclusion

- Our research demonstrates that augmenting CVE descriptions with rich, structured features significantly improves mapping accuracy to MITRE ATT&CK tactics. The SecRoBERTa-based model, which utilized this fine-tuned and extended dataset, outperformed the current state-of-the-art models

- We achieved a **93.47% weighted F1 score** for our final model, representing a significant improvement over the baseline. Our approach not only enhanced the overall accuracy but also addressed a long-standing challenge by reducing the number of hard-to-predict tactics from **four** down to just **one**.

- This accurate mapping enables Security Operations Centers to prioritize and mitigate unpatched vulnerabilities more effectively. By shifting from a reactive to a proactive, threat-informed defense, organizations can better understand their attack surface and prevent real-world exploitation.

QUEEN'S UNIVERSITY BELFAST

SCHOOL OF ELECTRONICS, ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

IARIA

Thank you all for your attention. Are there any Questions?

Keep in touch:

www.linkedin.com/in/logan-mcmahon-419b77255

lmcmahon25@qub.ac.uk

Outlook