A Self-Knowledge Distillation Approach to the Robustly Optimized BERT Approach for Common Vulnerabilities and Exposures

Gurinder Pal Singh[1], Rohit Bajaj[1] and Manish Kumar Hooda[2]
[1]*Computer Science and Engineering, Chandigarh University, Gharauan, Punjab, India*
[2]*Indian Semiconductor Mission, Ministry of Electronics and IT, India*

Keywords: BiLSTM, Cyber Security, CVE, MITRE ATT&CK, RoBERTa, TTP’s.

Abstract: The ever-increasing threat of cyber assaults on critical infrastructure businesses has prompted a focus on bolstering their cyber security expertise. Common Vulnerabilities and Exposures (CV&E) are the most crucial to understand since they are a collection of flaws that may be discovered in a wide range of software and hardware. However, many vulnerabilities remain unaddressed, making it impossible for an attacker to exploit them against you. A well-known approach for managing cyber security risks called MITRE ATT&CK (Adversary Tactics, Techniques, and Common Knowledge) provides mitigation techniques for a variety of destructive tactics employed by adversaries. Despite the enormous advantages of ATT&CK and CVEs, cyber security stakeholders might benefit from this approach. The CVE model proposed in this study contains a self-knowledge distillation design applied to the pre-trained language model of the highest-caliber model, Robustly Optimised BERT Approach (RoBERTa). A proposed novelty is based on a high-quality dataset that can improve the model's F1-score. The proposed model exceeded the F1 score of 77.20% and improved the accuracy of 75.92% compared to conventional machine learning models. This study's findings the preliminary information from MITRE ATT&CK may be beneficial to cybersecurity stakeholders.

1 INTRODUCTION

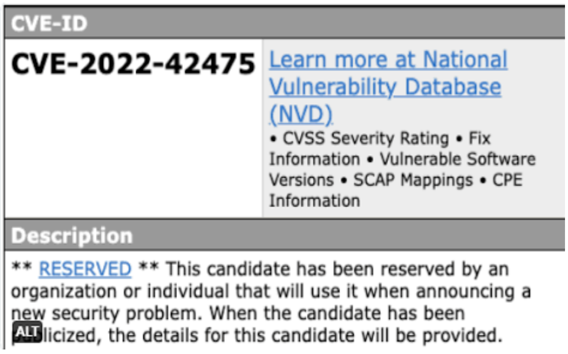
Cybersecurity Ventures predicts that by 2025, the global cost of cybercrime will reach $10.5 trillion annually. This is up from $3 trillion in 2015 (Freeze, 2020). The economic impact of cybercrime is considerable. In 2020, cybercrime will cost the global economy $5.2 billion. This represents 1% of global GDP. This forecast was unable to predict the COVID-19 crisis. Cybercriminals evolved and ramped up their assaults at an astonishing rate, preying on people's anxiety and uncertainty as a result of the pandemics’ fragile social and economic conditions. Corona virus-related fraud complaints soared over 350 times in March 2020, affecting over £800,000 people within one month, as per the UK National Fraud & Computer Security Agency (*Coronavirus-Related Fraud Reports Increase by 400% in March | Action Fraud*, n.d.). Data loss and destruction, financial losses, lower productivity, intellectual property theft, sensitive data exposure, and risks of cybercrime include theft, fraud, obstruction of regular companies’ activities following (Berthold et al., 2008) an attack, evidence collection, restoration and eradication of compromised data and systems, and damage to one's reputation. The motives, motivations, and ultimate aims of cybercriminals must be understood to fully assess the damage they have wrought. Cybercriminals communicate through a variety of adversarial patterns and contradictory actions.

Furthermore, as the number and sophistication of cyberattacks increase, so does the state of cybersecurity. The current level of advancement in cyber security measures is insufficient. Antivirus software, firewalls, and other security measures are available. We have access to security operation centres, intrusion detection systems, and a multitude of other data. Organizations are the primary focus of security technologies and responses. Antiattack detection or prevention. What is the best way to do business? The importance of investigating attack action connections and anticipating malevolent conduct, which enables proactive detection and mitigation of intrusions, is often overlooked. To address new and changing cyber threats, it is critical to expand our cyber security knowledge base. The Common Vulnerabilities and Exposures (CVE) list maintained by the MITRE Group is a valuable

Singh, G. P., Bajaj, R. and Hooda, M. K.
A Self-Knowledge Distillation Approach to the Robustly Optimized BERT Approach for Common Vulnerabilities and Exposures.
DOI: 10.5220/0013634900004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 636-643
ISBN: 978-989-758-763-4

resource for anyone interested in cyber security. The use of CVEs to organize efforts to fix problems is common practice among security analysts. A new CVE is created each time a vulnerability in the system is found and reported to MITRE. Figure 1 is an illustration of a CVE and its ID.



CVE-ID

CVE-2022-42475 Learn more at National Vulnerability Database (NVD)

• CVSS Severity Rating • Fix Information • Vulnerable Software Versions • SCAP Mappings • CPE Information

Description

** RESERVED ** This candidate has been reserved by an organization or individual that will use it when announcing a new security problem. When the candidate has been ALT licized, the details for this candidate will be provided.

Figure 1: An example of common vulnerabilities and exposures taken from cve.mitre.org.

Each vulnerability or exposure has a specific identification assigned by the CVE. CVE identifiers (also known as CVE names or numbers) allow security experts to search for information on specific cyber risks across numerous sources using a single name. The product is CVE-compliant, and its reports include CVE IDs. This allowed us to search for any CVE-compliant vulnerability database for repair information.

MITRE AT&CK founded it in 2013 to describe and categorise attacker techniques, methods, and protocols/procedures (TTPs) against Microsoft Windows programs (Unit 42 Threat Intelligence and IoT Security Experts, 2021) with the aim of enhancing the detection of hostile behaviour after system penetration. Over the years, ATT&CK has developed a classification of hostile activities and a level of expertise in cyber adversary behaviour by studying numerous platforms and systems. This strategy, along with an opponent simulation scenario, can uncover analytical monitoring and defensive gaps in target networks (Campbell et al., 2003). A few key components form the foundation of ATT&CK, a behavioural model in which a tactical adversary seeks to launch an attack. It answers the question, "Why?" Tactics serve as contextual categories for specific approaches, encompassing conventional attacker operations such as data collection, privilege escalation, and defense avoidance and providing a more detailed description. Techniques: The technique involves elucidating how adversaries achieve their tactical (Hasan et al., 2019) objectives during an activity. In other words, the technique concentrates on the "who" and, in certain situations, the "what" an enemy gain from their actions. To achieve tactical objectives, there may be a variety of methods or approaches, each of which has several techniques and sub-techniques.

The text provides a detailed explanation of how competitors can accomplish tactical goals at a level (Hasan et al., 2019) beneath tactics. Procedures define the precise implementation of an adversary's tactics or sub-techniques. They also serve to illustrate the application of these methodologies, or sub-techniques, in the field. Demonstrate various extra behaviors in the manner in which they are executed.

Given the significant benefits that CVEs and the ATT&CK methodology can offer to critical cyber security participants such as experts, academics, trainers, and executives, these groups continue to remain distinct. As of the beginning of 2021, there were (Berthold et al., 2008) over 156,000 CVEs in existence, making it challenging to collect mitigation strategies for each individual CVE and link them individually to the ATT&CK framework. In this study, we devised a unique method that assigns an ATT&CK tactic label to CVEs not included in our gold standard CVE dataset, utilizing textual elements from CVEs previously linked to an ATT&CK technique in previous research.

The data set used in this study is the "Gold Standard CVE Data Set", which is a common vulnerability and exposure data set (CVE) manually annotated by cyber security experts (Samtani et al., 2021). Some of the viewpoints of the datasets are as under:

- The data set contains information about CVEs, a unique identification of a known security vulnerability.
- The data sets are considered a "golden standard" because they are manually marked by cyber security experts, which means that each CVE is reviewed and verified by human beings, rather than relying solely on automated methods.
- The data sets include information such as CVE ID, vulnerability description, severity of vulnerability and software or hardware affected by vulnerability.
- The data sets are used to evaluate the performance of different machine learning models to predict the severity of new CVEs.

This research builds a novel cyber security model, the CVE Transformer (CVET), using cutting-edge deep learning-based text classification techniques. We then thoroughly compared CVET to standard models from the CVE data mining and (Khan et al., 2018) cyber security analytics literature to ensure its viability. The rest of the paper follows this structure.

First, we conducted a literature survey on CVE data machine learning, text classification transformers, distilled our own expertise, and coordinated the use of vulnerability disclosure (CVD) data in machine learning. Second, we defined the research issue for examination after identifying gaps in our literature review and proposed the method and its architecture. Thirdly, discuss our recommended strategy for marking CVEs with the MITRE ATT&CK method. The next section presents the results and experiments, followed by a summary and exploration of the implications of the empirical findings. Finally, provide major conclusions and the future scope of the study.

2 LITERATURE SURVEY

Recent research on CVEs shows that it faces several challenges that impact the accuracy, recall, and F1-score. So multiple different approaches have been used in recent years. The researcher (Das et al., 2021) proposes a novel Transformer-based learning framework called V2W-BERT for automating the mapping of observed vulnerabilities in software listed in Common Vulnerabilities and Exposures (CVE) reports to weaknesses listed in Common Weakness Enumerations (CWE) reports. examines the CVE records of known exploited vulnerabilities to identify trends and patterns. The author (Guyon and Elisseeff, n.d.) found that the most common vulnerabilities are those that allow for Remote Code Execution (RCE). RCE vulnerabilities can cause attackers to execute arbitrary code on target systems, causing malicious activities such as data theft, system corruption, and denial of service (Lim et al., 2023). Researchers using the CVE dataset found that the number of known exploited vulnerabilities using new CVE records has increased in recent years. This indicates that attackers have become more sophisticated and are finding new ways to exploit vulnerabilities. It does not provide a complete list of known exploited vulnerabilities but focuses on identifying trends and patterns in data.

The author (Dodge et al., 2020) explores three factors that can affect the fine-tuning process: weight initialization, data order, and early stopping. Weight initialization refers to how the parameters of the large language model (LLM) are initialized before training. The data order refers to the order in which the data are presented to the model during training. Early stopping refers to the stopping of training when the model's performance on a validation set stops improving. They conducted experiments on four datasets from the GLUE benchmark (Liu et al., 2019), fine-tuning Bidirectional Encoder Representations from Transformers (BERT) hundreds of times on each while varying only the random seed. In this research, the author (Sangaroonsilp et al., 2023) investigates the coverage of privacy-related vulnerabilities in the Common Weakness Enumeration (CWE) and Common Vulnerabilities and Exposures (CVE) systems. The model used in the research is a machine learning model that was trained to detect privacy-related vulnerabilities in software. The model was trained on a dataset from CVE of software vulnerabilities that had been manually labeled as privacy-related or not privacy-related. This section contains three parts, namely text classification transformers, self-knowledge distillation, and CVE data machine learning.

2.1 Transformers for Text Classification

Recurrent cells are replaced with multi-head attention mechanisms in well-known deep learning (Devlin et al., 2019) models for classifying text (such as Bidirectional Long Short-Term Memory (BiLSTM) and Long Short-Term Memory (LSTM). While the first design (for machine translation tasks) contained an encoder-decoder structure, multiclass text categorization calls for an encoder stack. In order to get a soft maximum probability score, the encoder transformer model takes the input, builds an embedding from it, and then passes it through the transformer block. The embedding layer uses both positional encoding and a one-hot encoding technique. In comparison to recurrent models on benchmark tasks, the transformer block has been shown to significantly improve recall, F1-score, accuracy, (Gong et al., 2019) precision, and accuracy. The feed-forward layers and multi-head attention mechanism make up the building block. Recently, huge Pre-Training Language Models (PTLMs) have been built utilizing transformers, achieving state-of-the-art performance on a variety of text classification tasks (Chalkidis et al., 2020). These models (like BERT and Generative Pre-trained Transformer (GPT-2) are often trained on millions of data points, and their parameters might number in the hundreds of millions. PTLMs can be enhanced and simplified for greater performance on particular tasks, despite the fact that the majority of researchers lack the (Furlanello et al., 2018) technology or data required to develop them. Extracting crucial information from a PTLM's parameters to fine-tune the training of a specific model is called "knowledge distillation," and it's a relatively new method.

2.2 Self-Knowledge Distillation

Through the process of Knowledge Distillation (KD), the expert knowledge of one model (the teacher) is combined with the developing knowledge of another model (the (student). As a result, compared to an uninstalled model (Guyon & Elisseeff, n.d.), the trained student model frequently performs better on data that have not yet been seen. This architecture makes it possible for researchers who lack the computational power to build a large PTLM to produce highly customized, cutting-edge models. Self-Knowledge Distillation (Self-KD) is a technique that is becoming more and more popular. Since both students and instructors use the same architecture, Self-KD facilitates smooth knowledge transfer. This technique enhances feature significance weighting, adjusts regularisation, and separates latent variables from deeper to (Dodge et al., 2020) shallower parts of the network to create without the need for further data, a new model that usually goes above and beyond the original teacher model. In natural language processing tasks (such as text categorization) when target labels are given, such as the weighted sum of (Seif, 2022) Cross-Entropy (CE) loss with the proper labels and CE loss with the soft target, Self-KD is often used.

2.3 CVE Date Machine Learning

Recently, large-scale efforts have been made to leverage CVEs to improve the security of various cybersecurity information systems using conventional machines and (Devlin et al., 2019) deep learning architectures. In order to predict the severity of CVE vulnerabilities, the authors constructed knowledge graphs using the Common Weakness Enumeration (CWE) and the Common Attack Pattern Enumeration and Classification (CAPEC) lists (Lim et al., 2023). However, CNNs struggle to grasp the interconnected nature of words and phrases. BiLSTM models with a self-consideration component have been used in the writing to further develop weakness seriousness and weakness type, (for example, limit condition error) prediction. Recently, researchers have extracted data from the well-known vulnerability database (Chalkidis et al., 2020) Exploit DB to supplement the textual descriptions of new CVEs using the pre-trained transformer model known as Bidirectional Encoder Representations from Transformers (BERT). To develop a model capable of accurately labeling CVEs using ATT&CK methods with textual descriptions, an algorithm must be capable of representing the lengthy text sequences common in (Guyon & Elisseeff, n.d.) CVE descriptions. In terms of text classification, the (Wang et al., 2021) transformer model (and its extensions) is now cutting-edge and has proven resilient to attacks from opponents. To further grasp how the transformer model may help us with our objective job, looked through it in great detail.

3 PROPOSED METHOD

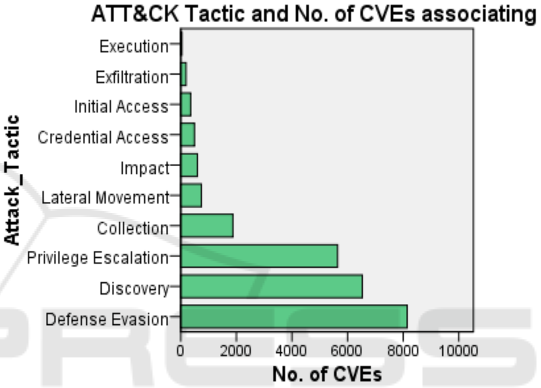


Figure 2: ASSOCIATING ATT&CK TACTIC AND N0. OF CVES.

Three main elements make up our suggested method: (1) data collection and pre-processing; (2) the Common Valurnability Exposures Transformer (CVET) structure; and (3) benchmark trials. In the subsections that follow, each element is explained in further detail. Made use of the dataset that the BRON conceptual model provided for our investigation (Devlin et al., 2019). As of February 8, 2023, the National Vulnerability Database (NVD) lists more than 176,000 CVEs. (*NVD - Home*, n.d.), and only a small portion of these, or 24599 CVEs, are included in our gold-standard dataset, making this connecting effort necessary. Using pre-existing knowledge bases, the dataset connects 24,863 CVEs to 10 of the 14 ATT&CK techniques. Many ATT&CK strategies (such as "Based on the Analytical" and "Command and Control") don't call for vulnerability. Thus, we can be tied to them. Figure 2 illustrates the ratio of the number of CVEs for each ATT&CK method. The four approach categories of protected evasion (8,452), discovery (6,647), backdoors (5,779), and collection (1,748) account for about 91 percent of the data.

Stop words and non-alphanumeric characters were removed from the CVE textual description as part of the pre-processing. The remainder of the text was lowercase, lemmatized, and padded to ensure that it was appropriate for all the inputs. Dimensionality reduction ttechnique is used to reduce the number of features or variables in a dataset while retaining the important information (Maaten Van Der et al., 2009) . The literature on deep-learning-based text classification frequently follows this order of pre-processing steps (Unit 42 Threat Intelligence and IoT Security Experts, 2021). The data supplied as input for our CVET model was encoded using the prebuilt RoBERTa tokenizer (Campbell et al., 2003). The usage of additional metadata contained in the CVEs was avoided because preliminary testing revealed that it did not improve model performance.

RoBERTa is a type of language model that is used in Natural Language Processing (NLP). It is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, which is a pre-trained NLP model developed by Google. RoBERTa was introduced in (Liu et al., 2019) 2019. The paper describes how RoBERTa was trained using a larger amount of data and a longer training time than BERT, resulting in improved performance on a variety of NLP tasks. In the paper RoBERTa is used as one of the machine learning models to predict the severity of new CVEs. RoBERTa is a machine learning model used in this paper to predict the severity of new CVEs. The authors chose RoBERTa because it proved to be successful in a variety of NLP tasks, including text classification, which is a task that predicts the severity of a CVE. Additionally, RoBERTa was trained on a large amount of data and for a longer period of time than the original BERT model, which may improve its performance on the task at hand (Farooq et al., 2023). Therefore, based on the authors' evaluation of different machine learning models, RoBERTa was found to be one of the most suitable models for predicting the severity of new CVEs. Since the CVE data set contains only English text data, a RoBERTa base case model was used. RoBERTa tokenizers have been pre-trained.

4 CVET ARCHITECTURE

4.1 Model Selection

Employ a PTLM called RoBERTa because of its generality in text-categorization tasks. While hundreds of PTLMs are suitable for our application, RoBERTa is selected because it combines high performance with the ability to fine-tune and Self-KD designs.

4.2 Fine Tuning

The performance of the conventional method for fine-tuning RoBERTa and other PTLMs (such as BERT and GPT-2) depends on the random seed and dataset size and is frequently unstable. Employing equations 1 and 2, we demonstrated that a combination of Adam optimization and bias correction led to a more stable training process and better results than baseline fine-tuning. Adam optimization with bias correction helps to prevent the gradients from becoming too large or too small. Equations (1) and (2) are the mathematical expressions for Adam optimization with bias correction. It is useful because it can help to improve the stability of training and the performance of the model. Adam optimization is a Stochastic Gradient Descent (SGD) algorithm that uses an adaptive learning rate. Bias correction is a technique that introduces a bias that can be corrected by adjusting the learning rate.

$$\alpha_t \leftarrow \alpha . \sqrt{1-\beta'_2}(1-\beta'_1) \tag{1}$$

$$\theta_t \leftarrow \theta_{t-1} - \alpha_t . m_t / (\sqrt{v_t - \varepsilon} \tag{2}$$

4.3 Self-Knowledge Distillation

CVET was both the instructor and the pupil as worked to refine the CVET model. While the teacher model is CVET at the coarse-tuning time step tt using Equation 3, the student model is CVET at the fine-tuning time step tt.

$$L\_\theta(x,y) = CE(CVET\_S(x,\theta),y) + \lambda MSE(CVET\_S(x,\theta),CVET\_T(x,\theta) \tag{3}$$

The equation (3) is a loss function for fine-tuning a BERT model for Natural Language Inference (NLI). Loss function measures the ability of the model to perform NLI tasks. The calculation is done by taking the total loss of cross entropy for each input-output pair.

Table 1: COMPARING CVET AGAINST BENCHMARK MODELS (*: P<0.05, **: P<0.01, ***: P<0.001).

Model Type	**Model**	**Accuracy**	**Precision**	**Recall**	**F1-score**
Classical Machine Learning	Random Forest	64.10% ***	34.92% ***	30.53% ***	34.22% ***
	SVM	64.34% ***	49.32% ***	45.22% ***	47.63% ***
	Naive Bayes	65.40% ***	43.05% ***	32.92% ***	37.76% ***
	Logistic Regression	66.10% ***	42.54% ***	35.12% ***	39.31% ***
	Random Forest	64.10% ***	34.92% ***	30.53% ***	34.22% ***
Deep Learning	RNN	68.92% ***	68.85% ***	68.30% ***	60.56% ***
	GRU	69.10% ***	74.59% ***	61.19% ***	71.83% ***
	LSTM	70.15% ***	75.64% ***	**72.09%**	71.40% ***
	BiLSTM	71.25% ***	75.92% ***	71.71%	75.22% ***
	BiLSTM with Attention	71.98% ***	69.52% ***	69.32% *	68.40% ***
	Transformer	73.04% ***	73.22% ***	69.82% *	70.61% ***
Pre-Trained Language Model	GPT-2	71.45% ***	75.03% ***	62.56% ***	70.67% ***
	XLNet	73.02% ***	79.12% *	61.50% ***	72.68% ***
	BERT	74.03% **	80.01% *	69.41% *	73.56% *
	RoBERTa	75.02% *	**80.43%**	67.03% **	71.54% *
Self-Distillation	**CVET**	**75.92%**	78.93%	70.12%	**77.20%**

The researcher's results are summarized in Table 1 and explored in more detail below.

The suggested CVET was evaluated against current and cutting-edge deep learning, classical machine learning, and pre-trained machine translation. Each benchmark model is frequently utilized for machine-learning tasks involving CVE data and/or text categorization. Except when US Customary units are utilized as identifiers in trade, such as 3.5-inch disc drive, the models chosen for each category are as follows.

- **Classical Machine Learning Techniques**: Random Forest, Support Vector Machine, Naive Bayes, and Logistic Regression
- **Deep Learning**: Recurrent Neural Networks, GRU, LSTM, BiLSTM, BiLSTM with attention, Converter.
- GPT-2, XLNet, BERT, and RoBERTa pre-trained language models.

We implemented all the conventional machine learning models using the Python package sci-kit-learn. We developed all the deep learning models in Python using the Keras framework. The hugging-face transformers library was used to implement the pre-trained language models. Using RoBERTa big from the Hugging-face package and our self-distillation and fine-tuning techniques, the CVET model was implemented in PyTorch 1.4.

Each model underwent ten validations. We evaluated the reference models using recall, accuracy, precision, and F1-score metrics, widely recognized as the gold standards for multiclass text classification tasks. To determine whether the CVET was significantly different from the various gold standards, researchers employed paired t-tests. Given the uneven distribution of our sample, the analysis primarily focuses on the F1 score, which is less susceptible to outliers.

5 EMPIRICAL RESULTS AND EVALUATION EXPERIMENTS

Based on Table 2, we offer four criticisms of the results of our testing. First, the F1 scores for the four conventional machine learning models ranged from 34.22 to 47.63 percent for Random Forest and SVM. Second, in terms of the F1 score, all deep learning models beat conventional machine learning models. Among the deep learning models, the BiLSTM model achieved the greatest F1-score (75.22 percent). Unlike other models, BiLSTM lacks an inherent recurrent mechanism, implying that its multi-head attention design aids in improving text categorization performance. Third, all PTLMs outperformed the transformer model by a small margin in the F1 score. The performance of the transformer was exceeded by the baseline RoBERTa from 73.61 to 74.57 percent by 0.96 percent. Based on these findings, the transformer might perform better on particular text-categorization tasks if it has been pre-trained.

Our suggested CVET model achieved an F1 score of 77.20%, which is higher than the F1 scores of any of the competing PTLMs, deep learning models, or classical machine learning models. In all models, the differences were significant at $p < 0.05$ or less. achieved the best accuracy (75.92%). These results suggest that our self-KD design and fine-tuning helped to enhance task performance.

6 CONCLUSIONS

This study develops a unique technique for automatically labeling CVEs with their related ATT&CK strategies via self-distillation. We fine-tuned the CVET model using a self-KD architecture and an Adam loss function with bias correction. We tested our model against deep learning, traditional ML, and pre-trained language models. The CVET model worked much better than baseline methods that didn't distillation when it came to tagging CVEs with MITRE ATT&CK strategies. Our approach can significantly benefit the cybersecurity community by directly connecting a widely-used cybersecurity risk-management framework to key vulnerabilities. Important cybersecurity stakeholders may use this to link their scanner-tracked weaknesses to the MITRE ATT&CK method, providing more insights into the most effective vulnerability mitigation strategies.

The author proposes two areas of relevant research on this topic. First, the intention is to strengthen the relationship between CVEs and other well-known Certificate Request Message Formats (CRMFs). There are two potential solutions: the Common Attack Pattern Enumeration and Classification (CAPEC) list and the NIST framework (National Institute of Standards and Technology). When a CVE is uncovered, such linkages can help widen the mitigation options provided in this study. Second, in order to improve the qualities of our textual inputs, we also want to look into various approaches to describe textual data more precisely (such as creative word-embedding schemes, creating synonyms and homonyms, and POS tagging).

Our research has contributed to the following areas:

- We propose a new, more robust and effective self-KD architecture that automatically labels CVEs using ATT&CK technology.
- RoBERTa has evaluated a number of natural language processing tasks and has achieved state-of-the-art results for each.

Using the MITRE ATT&CK framework and CVE, our CVET model lists the vulnerabilities with CVE tags.

7 FUTURE SCOPE

You can use other methods to improve accuracy, recall, and F1 score. The authors use a small number of CVEs in this paper, but the data sets could be larger. We can analyze text data more thoroughly to identify CVEs using MITRE ATT&CK frameworks. Text categorization can be pre-trained.

REFERENCES

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2008). KNIME: The Konstanz Information Miner.

In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications* (pp. 319–326). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-78246-9_38

Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market*. *Journal of Computer Security*, *11*(3), 431–448. https://doi.org/10.3233/JCS-2003-11308

Chalkidis, I., Fergadiotis, M., Kotitsas, S., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). An Empirical Study on Large-Scale Multi-Label Text Classification Including Few and Zero-Shot Labels. *arXiv:2010.01653 [Cs]*. http://arxiv.org/abs/2010.01653

Coronavirus-related fraud reports increase by 400% in March | Action Fraud. (n.d.). Retrieved July 25, 2023, from https://www.actionfraud.police.uk/alert/coronavirus-related-fraud-reports-increase-by-400-in-march

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805 [Cs]*. http://arxiv.org/abs/1810.04805

Dodge, J., Ilharco, G., Schwartz, R., Farhadi, A., Hajishirzi, H., & Smith, N. (2020). Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping. *arXiv:2002.06305 [Cs]*. http://arxiv.org/abs/2002.06305

Farooq, M., De Silva, V., Tibebu, H., & Shi, X. (2023). Conversational Emotion Detection and Elicitation: A Preliminary Study. *2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET)*, 1–5. https://doi.org/10.1109/GlobConET56651.2023.10149922

Freeze, D. (2020, November 10). Cybercrime To Cost The World $10.5 Trillion Annually By 2025. *Cybercrime Magazine*. https://cybersecurityventures.com/cybercrime-damage-costs-10-trillion-by-2025/

Furlanello, T., Lipton, Z. C., Tschannen, M., Itti, L., & Anandkumar, A. (2018). Born Again Neural Networks. *arXiv:1805.04770 [Cs, Stat]*. http://arxiv.org/abs/1805.04770

Guyon, I., & Elisseeff, A. (n.d.). *An Introduction to Variable and Feature Selection*. 26.

Hasan, K., Shetty, S., & Ullah, S. (2019). Artificial Intelligence Empowered Cyber Threat Detection and Protection for Power Utilities. *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)*, 354–359. https://doi.org/10.1109/CIC48465.2019.00049

Khan, M. S., Siddiqui, S., & Ferens, K. (2018). A Cognitive and Concurrent Cyber Kill Chain Model. In K. Daimi (Ed.), *Computer and Network Security Essentials* (pp. 585–602). Springer International Publishing. https://doi.org/10.1007/978-3-319-58424-9_34

Lim, J., Lau, Y. L., Ming Chan, L. K., Tristan Paul Goo, J. M., Zhang, H., Zhang, Z., & Guo, H. (2023). CVE Records of Known Exploited Vulnerabilities. *2023 8th International Conference on Computer and Communication Systems (ICCCS)*, 738–743. https://doi.org/10.1109/ICCCS57501.2023.10150856

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (arXiv:1907.11692). arXiv. http://arxiv.org/abs/1907.11692

Maaten Van Der, Laurens, Eric O, Postma, & H. Jaap van den Herik. (2009). *Dimensionality Reduction: A Comparative Review*. *10*, 66–71.

NVD - Home. (n.d.). Retrieved July 25, 2023, from https://nvd.nist.gov/

Samtani, S., Yang, S., & Chen, H. (2021). ACM KDD AI4Cyber: The 1st Workshop on Artificial Intelligence-enabled Cybersecurity Analytics. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 4153–4154. https://doi.org/10.1145/3447548.3469450

Sangaroonsilp, P., Dam, H. K., & Ghose, A. (2023). On Privacy Weaknesses and Vulnerabilities in Software Systems. *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, 1071–1083. https://doi.org/10.1109/ICSE48619.2023.00097

Seif, G. (2022, February 11). *The 5 Clustering Algorithms Data Scientists Need to Know*. Medium. https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68

Unit 42 Threat Intelligence and IoT Security Experts. (2021, March). 2020 Unit 42 IoT Threat Report 2020 Unit 42 IoT Threat Report. *Unit42*. https://unit42.paloaltonetworks.com/iot-threat-report-2020/

Wang, T., Qin, S., & Chow, K. P. (2021). Towards Vulnerability Types Classification Using Pure Self-Attention: A Common Weakness Enumeration Based Approach. *2021 IEEE 24th International Conference on Computational Science and Engineering (CSE)*, 146–153. https://doi.org/10.1109/CSE53436.2021.00030