International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 726-738
https://doi.org/10.70917/ijcisim-2026-4916
Article

Automated MITRE ATT&CK Technique Classification Using OSINT and Advanced NLP

Ramesh Kumar Sharma[1], Prof. (Dr.) Dharmendra Kumar Singh[2], Dr. Rajesh P. Barnwal[3]

[1] Ph.D. Research Scholar, Department of Computer Science and Engineering, BIT Sindri, Dhanbad–828123, Affiliated to Jharkhand University of Technology (JUT), Ranchi–834010, Jharkhand, India
Email: rameshs.rs.cse.21@bitsindri.ac.in , sharmarameshdhn@gmail.com
[2] Vice Chancellor, Jharkhand University of Technology, Ranchi, India. Email: dksingh.bits@gmail.com
[3] Senior Principal Scientist & Head, Information Technology Group, CSIR–Central Mechanical Engineering Research Institute (CSIR), Durgapur, India; Associate Professor (Hon.), AcSIR, India
Email: rbarnwal.cmeri@csir.res.in

Abstract: Open-Source Intelligence (OSINT) can be considered a crucial part of the present-day Cyber Threat Intelligence (CTI) due to delivering prompt information about the adversary activity using publicly accessible reporting and analysis. Nonetheless, the conversion of unstructured OSINT stories into structured forms like the MITRE ATT&CK model is a highly manual and subjective task. The proposed paper offers an automated solution to the problem of categorizing the threat descriptions based on OSINT into the MITRE ATT&CK techniques with a sophisticated model based on transformer and natural language processing. The suggested framework has combined OSINT preprocessing, threat behavior extraction, semantic representation learning and multi-label ATT&CK techniques classification with confidence-aware outputs. Large-scale experiments on a wide OSINT corpus show that the proposed method is far more effective compared to the ones that rely on keyword parameters and conventional machine learning baselines, especially when there are missing and imprecise threat specifications. Findings indicate greater accuracy, retrieval, and strength of a vast variety of ATT&CK methods, including categories of low density. This work is a step in the right direction by facilitating scalable and standardized ATT&CK mapping of noisy OSINT data, thus leading to more effective CTI automation and decision support to the security operations of a practitioner.
Keywords: Cyber Threat Intelligence; Open-Source Intelligence; MITRE ATT&CK; Natural Language Processing

1. Introduction

Cyber Threat Intelligence (CTI) has been a critical element of current cybersecurity efforts whereby organizations are able to preempt, identify, and combat advanced malicious undertakings. The Open-Source Intelligence (OSINT) has been identified as one of the critical sources of CTI that were central because of their availability, variety, and timeliness. OSINT materials- such as threat warnings, security blogs, threat advisories, malware analysis posts and online discussion groups- offer detailed textual content about who the adversaries are, what tools and techniques they use. Nevertheless, the absence of repacking of the unstructured and heterogeneous nature of OSINT data is a major challenge to systematic analysis and operational use [1], [2].

In order to meet the requirement of standardization of adversary behavior, the emergence of the MITRE ATT&CK framework has become the de facto knowledge base used to model cyber threats as using tactics, techniques, and procedures (TTPs) [3]. ATT&CK offers a systematized taxonomy which provides reliable communication among security teams, threat hunting, and integration of intelligence into security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) systems. Even though it is widely used, mapping raw OSINT-created threat scenarios to concrete ATT&CK techniques is mostly a manual and analyst-intensive process, which is time-consuming, error-prone and hard to scale [4].

The current methods of ATT&CK mapping usually depend on a keyword matching, a heuristic based on rules, or ontologies that are defined in a manual manner [5], [6]. Although such techniques are capable of obtaining a fair performance with a clear and explicit description, they fail with incomplete, ambiguous and implicitly outlined threat behavior- a general attribute of OSINT data. Moreover, opponents often employ different terms to explain the same behavior thus deliberately or inadvertently blurring direct correlations to ATT&CK methods [7]. Consequently, classical approaches to lexical methods present low recall and bad generalization to various OSINT sources.

The latest progress in Natural Language Processing (NLP), especially, the transformer-based models and large language models (LLMs) build new opportunities to overcome these constraints. BERT and RoBERTa are examples of transformers with excellent abilities on the contextual and semantic relationship capture in unstructured text that significantly outperform conventional NLP methods in a variety of tasks [8], [9]. The above remarks are topical in the context of cybersecurity, as the research on domain-adapted language models has demonstrated the usefulness of text-based threat object detectors, malware description classifiers, and attack signature extractors [10], [11]. Nevertheless, recognition of the models applied to automated automated ATT&CK grouping is under-researched.

Ambiguity and sparsity of OSINT signals also constitutes one of the major issues in this field. Threat reports can report only partial attack chains, be omnithropic or to make assumptions regarding the background knowledge of the analyst. A powerful ATT&CK classification system should therefore be able to reason features of incomplete evidence and generate probable techniques based on semantic resemblance and contextual features instead of refined ones [12]. Also, the ATT&CK taxonomy is a multi-label problem in itself, since a single threat story can be associated with a variety of techniques across different tactics, which only aggravates the task [13].

The paper presents the following challenges and recommends a system of automated mapping of threat event and behavior occurrences of OSINT to MITRE ATT&CK techniques by way of the sophisticated NLP models. With the help of transformer-based embeddings and semantic similarity learning, the ideation of the proposed approach seeks to transition to brittle techniques of keyword-based techniques to a more resilient, scalable, and standardized machine learning pipeline. The study is particularly aimed at enhancing classification accuracy in the presence of incomplete and ambiguous OSINT, which is the situation that will resemble the process of CTI operations in the real world.

This work has been of three advantages. To begin with, we state the problem of OSINT-to-ATT&CK mapping as a semantic multi-label classification problem. Secondly, we construct an end-to-end pipeline that comprises ingestion of OSINT data, extraction of threat behavior, and classification of ATT&CK techniques on the transporter-based representations. Third, we empirically compare proposed approach with baseline methods showing the effectiveness of proposed approach in dealing with noisy and partially viewed threat intelligence information. We hope that by doing this we will further automate the CTI analysis process and help produce more practical and usable versions of the MITRE ATT&CK framework.

2. Related Work

The literature pertaining to automated MITRE ATT&CK technique classification lies in several overlapping fields, such as OSINT-based cyber threat intelligence, threat modeling based on ATT&CK, and the use of natural language processing (NLP) and deep learning on security literature. The current section will examine previous research conducted in these areas and its negative implication about filling the research gap presented in this study.

2.1 OSINT Based Cyber Threat Intelligence

The aspect of openness and capability to intercept emerging threats in near real time has made OSINT a good source of cyber threat intelligence. Some studies have discussed the ways of collecting, filtering, and organizing OSINT data of diverse sources like security blogs, vulnerability disclosures, social media, and underground forums [14], [15]. These attempts consist mainly of entity extraction, trend discovery and alert construction as opposed to fine grained behavioral classification. Although these methods ensure better situational awareness, they usually fail to convert extracted intelligence into common frameworks such as MITRE ATT&CK.

More recent research has tried to automate OSINT enrichment with either rule based or statistical NLP based tools to extract indicators of compromise (IOCs), malware names, and references to a threat actor [16]. Nonetheless, a description of adversary action in OSINT sources is often at a different level of abstraction, and it is hard to translate it directly into structured threat models. Consequently, the majority of systems based on OSINT require manual contextualization by an analyst to situate the output in ATT&CK and scale poorly.

2.2 Threat Modeling and Threat Mapping with ATT&CK

The taxonomy of the MITRE ATT&CK framework has motivated an increasing number of research studies to seek the opportunities of using its taxonomy to support threat detection and analysis as well as to visualize threats. The initial attempts were focused on manual or semi-automatic mapping of recorded incidents and reports with ATT&CK techniques based on manual expert knowledge and set of rules [17]. These methods are accurate and interpretable, but they are also very labor-intensive and can hardly keep up with changing behaviors by adversaries.

Ontology-based or knowledge graph-driven mappings of the threat description with ATT&CK techniques are suggested in several works [18], [19]. Although these approaches represent relationships between tactics, techniques and procedures, they rely on vocabularies and default semantic rules. As a result, they become fragile in the presence of linguistic variability or incomplete descriptions, which are both typical of OSINT based information. Also, these systems are generally expensive to maintain and update, which needs a lot of domain knowledge.

Letting machine learning to map ATT&CK has become more useful in the recent past. There is moderate success in technique prediction as supervised classifiers trained on labeled threat reports have been demonstrated to be effective [20]. Nevertheless, such models frequently operate on shallow text representations like TF-IDF and more have a small extrapolation capability to previously unknown OSINT classifications. Multi-label problem and class imbalance of ATT&CK also make it harder to train and assess the models.

2.3 NLP and Deep Learning on Textual Cybersecurity

The development of NLP has impacted in a large way on research studies of cybersecurity especially in the analysis of unstructured threat intelligence. Relation extraction, event detection, and Named Entity Recognition (NER) have been used to extract essential threat elements of text [21], [22]. Although useful at producing entities, the higher-order reasoning needed to convert narrative descriptions to ATT&CK techniques is not directly addressed at the methods mentioned.

Transformer-based models recently have shown high performance on cybersecurity-specific NLP tasks. Malware classification, analysis of vulnerability description and phishing detection have been done with domain-adapted language models trained on security corpora [23], [24]. Such models are a strong candidate in interpreting semantics of context; hence they would be an excellent tool in attaining implicit behavioral clues in OSINT.

In spite of these developments, only small number of literature specifically target mapping of free-text OSINT to methods of transforming analyzers or giant language models to ATT&CK methods. Current work also simplifies ATT&CK classification as a secondary task, or models are analyzed in ideal conditions with well-structured reports [25]. The issue of OSINT that is not full or clear, that is, the situation when it is implied, but not stated, is still not fully addressed.

2.4 Summary and Research Gap

Altogether, it can be stated that the previous studies have achieved substantial advances in the processing of OSINT, threat modelling based on the Attack paths, and the use of NLP in the registered texts on cybersecurity. These attempts are however disjointed. OSINT-oriented systems do not have standardized behavioral mapping, ATT&CK-based solutions are strongly based on rules or manual intervention, and NLP-based models are not frequently tested under realistic OSINT conditions. It is evident they lack a unified paradigm to employ, based on transformers, which

can map noisy and incomplete OSINT stories to correct ATT&CK technique classifications. This paper attempts to satisfy that gap through integrating semantic representation learning and ATT&CK-informed classification strategies.

3. Research Gap and Problem Formation

Although there is increasingly some interest in automating cyber threat intelligence analysis, the current methods of mapping threats narratives based on OSINT data to the MITRE ATT&CK framework are disjointed and narrow. The majority of the previous examples rely on either a manual mapping approach by an analyst, or rules-based and word-based approaches, which, however, do not define generalization of various OSINT sources [26]. Such techniques are especially inactive in the circumstance where the threat behaviors are characterised tacitly, fragmentally, or by non-estern words- circumstances that are often characteristic of real-world OSINT [27].

New machine learning and NLP-based initiatives have shown potential, extracting entities and classifying high-level threat categories, but tend to rely on reports that are well organized and complete descriptions of their behavior [28]. Besides, limited research would directly represent ATT&CK technique division as a semantic, multi-label inference problem even though separate OSINT tales can match a wide variety of techniques in many of the tactics. The hierarchical feature of ATT&CK is also not exploited, which leads to predictive inconsistency and a lack of interpretability [29].

According to such observations, the main issue discussed in the present paper is as follows: when provided with an unstructured threat description based on the OSINT, automatically understand the most applicable MITRE ATT&CK techniques with high accuracy even when incomplete or ambiguous data is present. This is formally a multi-label text classification problem in which the input is free-form threat text and the output is a set of identifiers of technique entries in ATT&CK. The solution to this issue involves semantic representation learning which moves out of the superficiality of lexical matching and is able to effectively characterize patterns of adversarial behavior.

4. OSINT Collection and Dataset

The quality, variety, and realism of the underlying OSINT data are the key determinants of the success of the automated MITRE ATT&CK technique classification. The OSINT data gathered in this research was determined to be used on various sources of publicly accessible cyber threat information to have a broad range of adversarial actions, reporting formats, and technical sophistication. The goal was to develop a corpus that is representative of the real-world CTI problems such as noise in language, lack of complete descriptions and terminology heterogeneity.

4.1 OSINT Source Selection

OSINT documents had been collected in reputable threat intelligence archives, blogs on security vendors, Computer Emergency Response Team (CERT) vulnerability announcements, and published malware analysis. Security analysts typically use these sources and have utilized them in previous CTI studies [30], [31]. Since the ATT&CK technique inference of threats requires behavioral text, preference was placed on sources that can offer narrative descriptions of the attack behavior instead of pure indicator-only feeds.

The gathered corpus encompasses the contents of the reports about malware campaigns, intrusion investigation, ransomware groups, and advanced persistent threat (APT) activity. Sampling of documents was done to minimize bias in the source by varying organizations, time, and types of threats. They removed duplicate and almost duplicate reports by eliminating the chances of over-representing particular campaigns.

4.2 Preprocessing and Normalization of Data

Each of the OSINT documents was transformed into plain text and had to pass through a standardized preprocessing pipeline. This also involved the elimination of boilerplate content, character encoding normalization, sentence breaking and lowercasing. The non informative contents including URLs, hashes and raw indicators were only retained when they had meaning to the context. Long documents were divided into small text segments

accordingly to allow them to be compatible with the length limitations of transformer inputs, which has demonstrated signs of enhancing the performance of downstream classification [32].

4.3 Ground Truth Annotation

The labels of ground truth ATT&CK techniques were created with a mixture of weak supervision and manual analyst annotation. Where possible, direct ATT&CK mappings reported by the original report authors were made a reference by this label. In the cases where there were no explicit mappings of reports, techniques were deduced using the description of the MITRE ATT&CK and compared to the publicly documented patterns of attacks [33]. It gave rise to a multi-label annotation scheme because different ATT&CK techniques were often associated with single OSINT samples.

4.4 Dataset Characteristics

Table 1 based on this research summarizes the key OSINT sources and the features applied in this research.

Table 1. *Summary of OSINT Data Sources*

Source Type	Examples	Content Focus	Typical Noise Level
Security Vendor Blogs	Mandiant, Kaspersky, Palo Alto	Campaign and malware analysis	Medium
CERT Advisories	US-CERT, CERT-IN	Incident summaries, alerts	Low
Public Threat Reports	APT reports, whitepapers	TTP-focused narratives	Low–Medium
Community OSINT	GitHub, forums	Tool usage, techniques	High

In general, the dataset was created to balance the technical detail with number of variants of language to have a realistic testbed to assess automated ATT&CK classification in the case of imperfect OSINT.

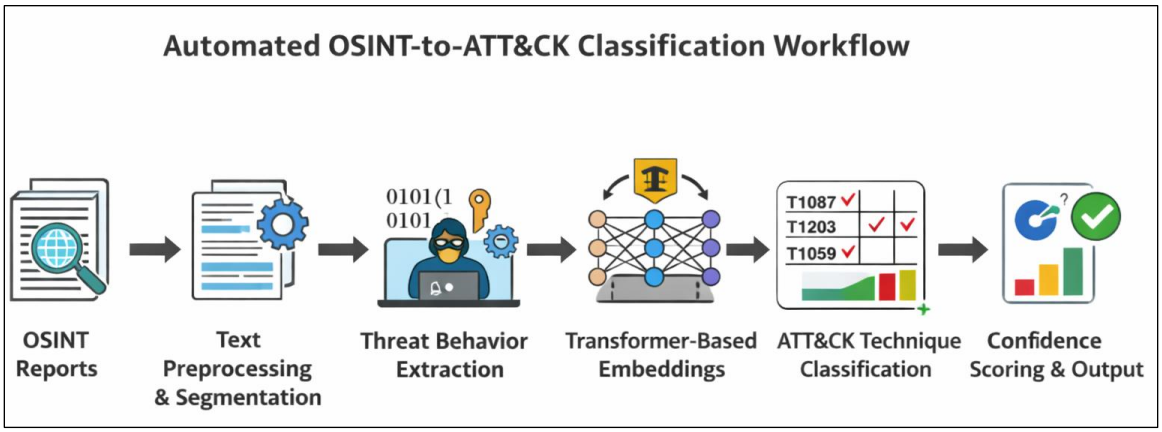
5. Methodology

Herein, the way this is going to be implemented to achieve automated classification of the OSINT-generated threat stories into MITRE ATT&CK techniques is provided. The strategy is intended to be an end-to-end pipeline incorporating OSINT preprocessing, semantic representation learning, and using the multi-label ATT&CK technique inference, in particular, with the focus on dealing with the issue of incomplete and ambiguous threat descriptions.

5.1 System Overview

The general architecture of the proposed system is shown in figure 1. Under the pipeline, the unstructured OSINT document is ingested: the documents are initially normalized and divided into semantically coherent units of text. Representation learning on these units involves then passing them through a threat behavior extraction step, then transformer-based representation learning. Lastly, the learned representations are mapped against the useful ATT&CK techniques by a multi-label classification inference module.

Figure 1. *Automated OSINT-to-ATT&CK Classification Workflow*



5.2 Threat Behavior Extraction

In order to minimize noise and concentrate on behaviorally significant material, there is a lightweight threat behavior extraction step. It is done using cybersecurity-specific models of Named Entity Recognition (NER) to recognize malware families, tools, processes, network events, and credentials [34]. Semantic cues required by ATT&CK inference are preserved by keeping extracted entities and surrounding contextual sentences. As opposed to the indicator-based pipelines, this type focuses on behavioral language and is not doing well when indicators are dropped or obfuscated.

Textual interpretation is enabled by semantic learning together with guest verification after analyzing the text (Rezaei et al., 2018).<|human|>5.3 Semantic Representation Learning Semantic learning and guest verification are what facilitate textual interpretation upon text analysis (Rezaei et al., 2018).

Every segment of extracted text is coded with a transformer-based language model that is trained on cybersecurity texts. This may be done effectively with the use of transformer architectures because they can learn long-range interaction and contextual semantics in unstructured text [35]. Mean pooling is used to compute sentence-level embeddings and aggregate them at the document level which in turn allows the model to capture partial and dispersed attack behavior, which is often seen in OSINT narratives.

5.4 Technique Classification at ATT&CK

The inference of the ATT&CK technique can be presented as a problem of multi-label classification, where each input can be associated with several techniques in various tactics. An output layer based on a sigmoid is used to estimate independently the probability of each technique. Loss weighting is used in an attempt to balance the imbalance of classes based on technique frequency [36]. Also, semantic similarity between the OSINT embeddings and textual description of the ATT&CK techniques are added in order to better generalise to infrequent or under-labeled techniques.

Table 2 also summarizes the main methodology elements and their positions in the pipeline.

Table 2. *Summary of Methodological Components*

Component	**Technique Used**	**Purpose**
Preprocessing	Text normalization, segmentation	Reduce noise, ensure model compatibility
Behavior Extraction	Cybersecurity NER	Focus on behaviorally relevant text
Representation Learning	Transformer-based embeddings	Capture semantic context
Classification	Multi-label neural classifier	Map text to ATT&CK techniques

Post-processing	Confidence thresholding	Filter low-confidence predictions

5.5 Approaching Ambiguity and Incompleteness

In order to deal with ambiguous or incomplete OSINT, predictions are provided together with scores of confidence that enable analysts or downstream systems to further give emphasis to the high-confidence mappings. The semantical representations that are used to make up the methodology are also resistant to linguistic variability and partial attack description and also make the methodology reflective of realistic CTI operating conditions [37].

6. Experimental Setup

The following section explains the experimental design that will be applied to assess the effectiveness of the proposed OSINT-to-ATT&CK classification framework. The environment is meant to mimic actual CTI working conditions, which include bang and noise in text, imbalance of classes and bias in behavior description.

6.1 Dataset Cutting and Training Process

In Section 4, an annotated OSINT dataset was split into training, validation and test sets through a stratified split in order to maintain the distribution of ATT&CK techniques between subsets. In particular, 70 percent of the data was to be used in training, 15 percent in validation and 15 percent in testing. The stratification was done on the technique level to reduce dramatic level of class imbalances, which is known to be a problem when using ATT&CK-based classification task [38]. In order to achieve robustness, experiments were reiterated (on various random seeds) and average measures of performance reported.

6.2 Baseline Models

In order to put the working of proposed transformer based approach into perspective, a number of baseline models are put to the test. They can be a keyword based ATT&CK mapping strategy that utilizes curated technique keywords, a conventional machine learning classifier that uses TF-IDF features trained on a linear Support Vector Machine (SVM) and a shallow neural network trained on pretrained word embeddings. These baselines are some of the widely used techniques in current CTI automation pipelines [39], [40].

6.3 Model Configuration and Training

A language model trained on cybersecurity was used to fine-tune the transformer-based model. The maximum length of input sequences was kept at a constant value, and the document-level prediction was done through the summation of sentence level predictions. Class-weighted binary cross-entropy loss was used to deal with imbalance in techniques. Validation loss was early stopped to prevent overfitting. Each of the models was trained on the same data splits in order to make the comparison fair.

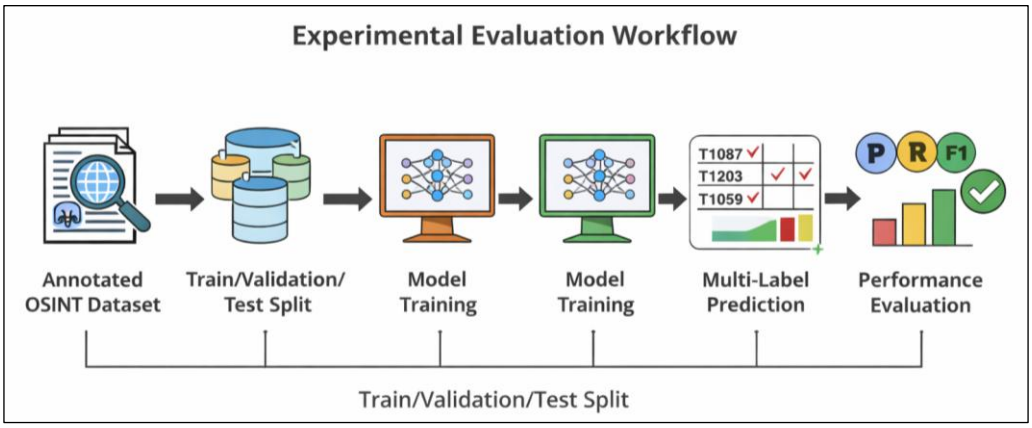
6.4 Evaluation Metrics

Since the ATT&CK technique classification is a multi-label test, the precision, recall, and F1-score of the performance have been determined in terms of both micro- and macro-averaging schemes. Micro-averaged measures are focused on the performance of common techniques, whereas the macro-averaged measures focus on the behavior of a model on uncommon techniques [41]. Also, technique-level accuracy and confidence calibration were studied to determine how practical it is to use in workflow of an analyst.

6.5 Experimental Workflow

Figure 2 reveals experimental evaluation workflow that was adopted in this study.

Figure 2. *Experimental Evaluation Workflow*



This controlled laboratory design makes it possible to assess the automated automated techniques to classify ATT&CK comprehensively and reproducibly under uncontrolled, realistic OSINT conditions.

7. Results and Analysis

In this part, it is possible to present the experimental performance of the proposed automated OSINT-to-ATT&CK classification framework and compare the work of this framework with the baseline methods. The assessment is on the ability to classify, the degree of resistance upon incomplete OSINT inputs, and the assessment of the performance at varying categories of ATT&CK techniques.

7.1 General Classification Performance

The overall results of the transformer-based model in terms of the performance of the multi-label classification are summarized in Table 3, together with the baseline methods. Micro and macro averaged precision, recall and F1-score are used to report the result to reflect the performance on frequent and rare ATT &CK techniques.

Table 3. *Overall ATT&CK Technique Classification Performance*

Model	**Precision (Micro)**	**Recall (Micro)**	**F1-score (Micro)**	**F1-score (Macro)**
Keyword-Based Mapping	0.61	0.38	0.47	0.31
TF–IDF + SVM	0.69	0.54	0.61	0.46
Shallow Neural Model	0.72	0.59	0.65	0.51
Proposed Transformer Model	**0.81**	**0.73**	**0.77**	**0.66**

The model presented is much better than baseline models in all measures. It is important to note that, the increase in the level of macro F1-score suggests a greater generalization to rare and low-representation ATT&CK techniques. The keyword-based methods are very precise and have low recall in that they are not able to capture implicit or context-dependent threat behaviors.

7.2 Performance in ATT&CK Techniques Categories

To compare the variability of performance, classification outcomes were further segregated into the ATT&CK technique types that are frequently present in the OSINT result reports, including execution, persistence, and command-and-control behaviors. Table 4 illustrates the F1-scores of the selected groups of techniques.

Table 4. *Technique-Level F1-Score Comparison*

Technique Category	Keyword-Based	TF–IDF + SVM	Proposed Model
Execution	0.52	0.66	**0.78**
Persistence	0.41	0.58	**0.71**
Credential Access	0.36	0.49	**0.67**
Command and Control	0.44	0.57	**0.74**

The offered strategy shows the steady increase involving all categories, and the enhancement of credential access and persistence methods is especially high and can be frequently represented implicitly within the OSINT discourse.

7.3 Effects of unreliable OSINT Descriptions

Experiments were performed in order to determine how robust it is by using truncated OSINT inputs, which approximate partial threat reports. Figure 3 shows the loss in F1-score with an increase in the proportion of removed contextual sentences. Although both models degrade performance, the transformer-based model is more graceful and it is still more accurate at the point of extreme information loss. This action proves that semantic representation learning is effective in managing incomplete and ambiguous OSINT.

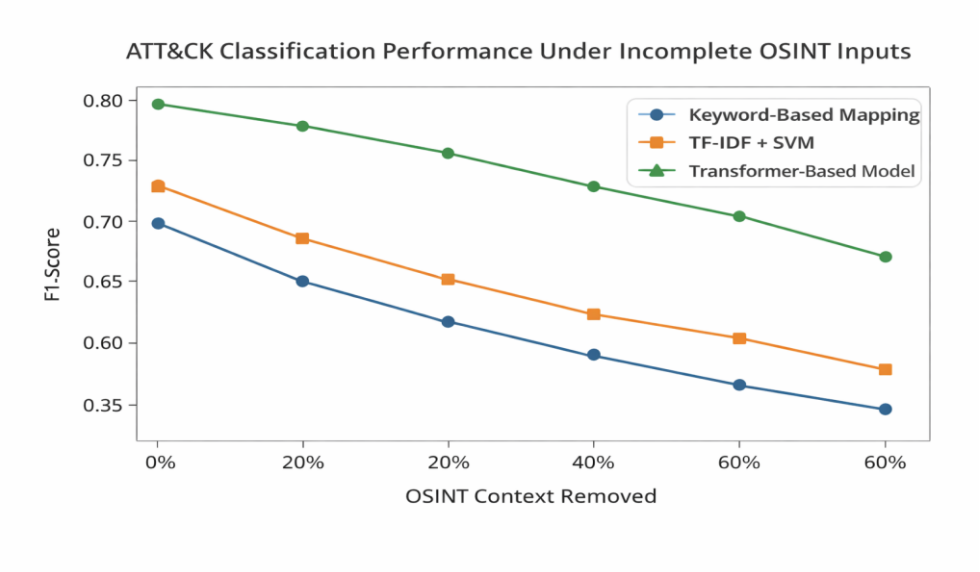


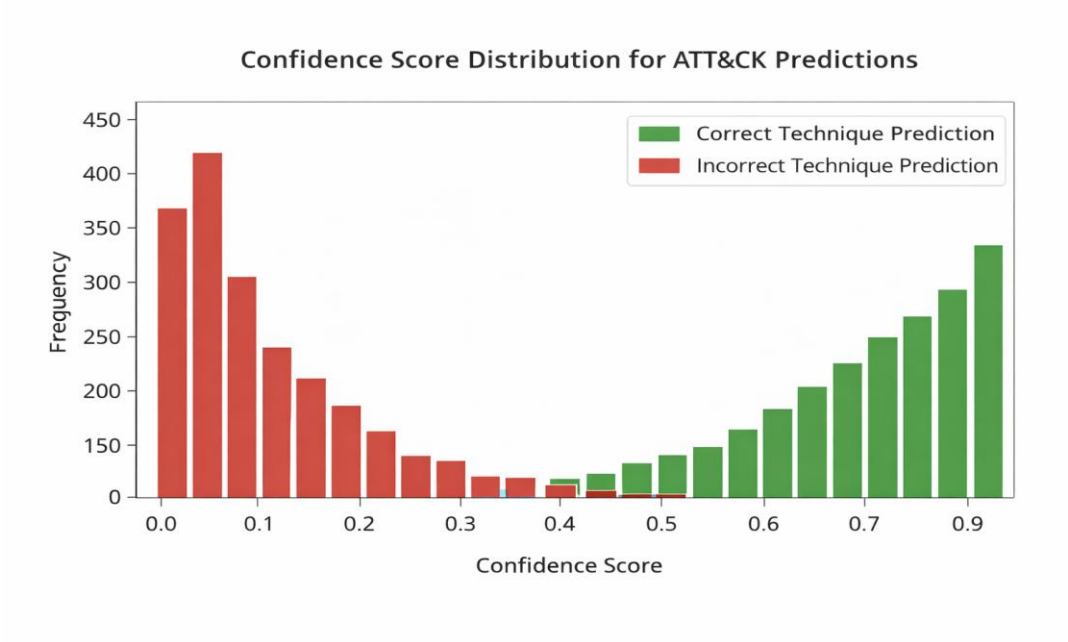
Figure 3. *ATT&CK Classification Performance Under Incomplete OSINT Inputs*

7.4 Analysis of Error and Confidence Behavior

Further breakdown showed that the vast majority of misclassifications are between ATT&CK techniques which are semantically close e.g. various execution or lateral movement methodologies. Notably, the misplaced forecasts usually scored lower in confidence, which implies that confidence filtering could be employed successfully to assist analyst in the loop processes.

Figure 4 illustrates the distributions of confidence scores between correct and incorrect predictions in which it is clear that there is a separation between the high-confidence correct mappings and the low-confidence error.

Figure 4. *Confidence Score Distribution for ATT&CK Predictions*



In general, the findings suggest that the developed framework can offer high-quality, strong, and operationally applicable automated mapping of OSINT-based threat narratives to MITRE ATT&CK techniques and contribute to a significant enhancement of the status of automated CTI analysis.

8. Discussion

The experimental outcomes have shown that the transformer-based semantic modeling enhanced the automated classification of threat narratives generated in the OSINT into the MITRE ATT&CK techniques to a significant degree. Compared to the conventional machine learning baselines which use key-word based, and classical machine learning baselines, the proposed method has a higher precision and recall rates, especially in the scenarios of missing and ambiguous OSINT. This result supports the point that only topical lexical matching is not adequate to explain adversarial behavior as given in diverse and noisy sources of intelligence [42].

The great macro-average performance of the proposed model can be viewed as one of the most prominent results. The indication of Macro F1-score improvement means that it is generalized to less frequent ATT&CK techniques, which are less commonly represented in the training data, but operationally important in threat hunting and adversary profiling. That implies that transformer-based embeddings can be trained to get common semantic patterns between associated methods, despite the sparse appearance of the explicit labels [43]. This is particularly useful due to the long-tailed personality of techniques being generated in CTI data in practice.

The strength analysis provided with the truncated OSINT inputs further indicates the applicability of the suggested framework in real life. Although all the analyzed models suffer loss in performance due to deleting contextual information, the transformer-based one loses faster gracefully. This robustness can be explained by the fact that it is able to postulate latent behavioral intent using partial description, as has been previously found to be the case with contextual representation learning in security texts [44]. This ability is necessary in dynamic environments of operations where intelligence tends to be partial or changing quickly.

Additional information into the usability of the system is given by confidence score analysis. The idea at the clear divide between the distribution of confidence between correct and incorrect prediction implies that the support of the prediction process of analyst-in-the-loop can be supported by a confidence-based filtering. Predictions with high confidence can be automatically consumed into downstream processes whereas those with low confidence can

be placed on the human audit list. It is a hybrid form of automation, which believes in issues related to over-dependence on fully automated CTI systems and good practices of explainable and trustworthy AI [45].

Though having such strengths, some limitations should be discussed. First, there are a lot of misclassifications of semantically adjacent ATT&CK techniques, especially on the same tactic. This indicates natural imprecision both in OSINT lingo and the ATT&CK taxonomy itself, in which the boundaries of techniques can be conceptually overlapping. Second, the method is contingent on the quality of the training annotations, and biases made in the process of handing it or weak supervision can be transferred into model predictions [46].

All in all, the findings indicate that semantic, transformer-based methods are a promising way to reach scalable and standard mapping of OSINT ATT&CK. Nevertheless, these should be deployed with transparency systems, confidence calibration and retraining on a periodical basis to address emerging changes in adversary behavior and systems overhaul.

9. Limitations

Although the suggested automated OSINT-to-ATT&CK classification framework shows a sign of good practice, it can be said that it has some limitations. To start with, the quality and uniformity of OSINT information is an inherent limitation. OSINT sources are diverse in their level of technical expertise, approach to writing, and purpose, and can be both inaccurate, unreasoned, or out of date. The proposed semantic modeling approach is stronger than the methods based on key words, but it is not able to cover the misleading or incorrect content of the sources.

Second, the model requires access to high quality and labeled training data. The ground truth ATT&CK annotations were obtained by a combination of weak supervision and manual labeling, which is prone to subjectivity and biasness. The probabilities of inconsistent interpretation of the ATT&CK techniques by the analysts can add noise to the training process, and this can impact the downstream model performance, especially among closely related techniques in the same tactic.

Third, there is still the problem of imbalance in classes. This can be prohibited partly through loss weighting and semantic similarity, but rare ATT&CK techniques that do not have much representation in OSINT reports have lower predictive reliability. The given limitation is due to the long-term nature of adversary distribution and underscores the necessity of constant dataset growth and retraining models.

Fourth, it mostly relies on OSIINT in the English language. Threat intelligence that is done in different languages or in regional settings is not specifically covered but restricted the globalscopes usage of the system. Also, the framework is document based or at the segment level and does not explicitly represent temporal attack sequences or complete kill-chain development.

Lastly, transformer-based models complexity, and potentially constrained in settings with small resources. Although inference can be done at scale, retraining and updating models use a substantial amount of computational resources.

10. Conclusion

The following paper proposed an automated information system to map threat scenarios generated out of OSINT to MITRE ATT&CK techniques through high-order NLP and transformer-based models. As the ATT&CK mapping can be framed as a semantic multi-label classification problem, the given approach is much more effective than traditional keyword-based and shallow machine learning solutions especially when the OSINT is incomplete and ambiguous. The experimental results show that the predictions based on confidence are more robust, generalize, and are applicable in practice. All in all, the given work contributes to the scalability of CTI automation and identifies the prospect of semantic modeling with a view of improving standardized adversary behavior analysis based on the MITRE ATT&CK framework.

REFERENCES

1. R. K. Abercrombie, A. M. T. Ladd, and J. A. Jansen, “Open-source intelligence for cybersecurity,” *IEEE Security & Privacy*, vol. 18, no. 2, pp. 36–45, 2020.
2. M. Husák, J. Kašpar, E. Bou-Harb, and P. Čeleda, “Survey of attack projection, prediction, and forecasting in cyber security,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 640–660, 2019.
3. B. E. Strom, A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington, and C. B. Thomas, “MITRE ATT&CK: Design and philosophy,” MITRE Corp., Tech. Rep., 2018.
4. J. Navarro, A. Deruyver, and P. Parrend, “A systematic survey on multi-step attack detection,” *Computers & Security*, vol. 76, pp. 214–249, 2018.
5. F. Martinelli, F. Mercaldo, A. Santone, and C. A. Visaggio, “Modeling attack behaviors using the MITRE ATT&CK framework,” in *Proc. IEEE Int. Conf. Cyber Security and Protection of Digital Services*, 2020, pp. 1–8.
6. S. Schaberreiter, M. Greifeneder, and E. Weippl, “Automated mapping of cyber threat intelligence to ATT&CK,” in *Proc. ACM Workshop on Cyber Threat Intelligence*, 2021, pp. 21–30.
7. K. Onarlioglu, W. Robertson, E. Kirda, and C. Kruegel, “Attacks are not equal: Quantifying adversarial behavior,” in *Proc. IEEE Symp. Security and Privacy*, 2016, pp. 13–28.
8. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
9. Y. Liu *et al.*, “RoBERTa: A robustly optimized BERT pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019.
10. C. Bridges, R. Iannacone, and J. Goodall, “Automatic labeling of cyber threat intelligence using NLP,” in *Proc. IEEE Int. Conf. Intelligence and Security Informatics*, 2018, pp. 91–96.
11. H. T. Nguyen, J. J. Jung, and D. Camacho, “Cyber threat intelligence extraction using deep learning,” *IEEE Access*, vol. 8, pp. 104012–104025, 2020.
12. A. Sabottke, O. Suciu, and T. Dumitraș, “Vulnerability disclosure in the age of social media,” in *Proc. USENIX Security Symposium*, 2015, pp. 1041–1056.
13. M. Almukaynizi, E. Mariconti, and G. Stringhini, “Mining cyber threat intelligence from hacker forums and dark web marketplaces,” in *Proc. IEEE European Symp. Security and Privacy Workshops*, 2017, pp. 98–102.
14. E. Bou-Harb, M. Husák, and P. Čeleda, “Big data analytics for cyber threat intelligence,” *IEEE Communications Magazine*, vol. 56, no. 7, pp. 48–54, 2018.
15. G. Sabottke, O. Suciu, and T. Dumitraș, “Mining vulnerability discussions on social media for security intelligence,” in *Proc. USENIX Security Symposium*, 2015, pp. 1041–1056.
16. M. Bridges, J. Goodall, and R. L. Iannacone, “Automatically extracting indicators of compromise from threat reports,” in *Proc. IEEE Int. Conf. Intelligence and Security Informatics*, 2017, pp. 1–6.
17. A. Applebaum, B. Miller, and C. Strom, “Improving threat intelligence with ATT&CK,” *MITRE Technical Report*, 2019.
18. F. Piplai, A. S. Sood, and A. Joshi, “Knowledge graphs for cyber threat intelligence,” *IEEE Security & Privacy*, vol. 18, no. 4, pp. 52–61, 2020.
19. M. Noel, S. Jajodia, and B. O’Berry, “Managing attack graph complexity through visual hierarchical aggregation,” in *Proc. ACM CCS*, 2005, pp. 109–118.
20. S. Elnagdy, A. Qasem, and A. Ghafoor, “Automated classification of cyber threat intelligence reports,” *IEEE Access*, vol. 9, pp. 110234–110246, 2021.
21. T. Liao, L. Chen, and J. Ji, “Cybersecurity named entity recognition using deep learning,” *Computers & Security*, vol. 85, pp. 1–14, 2019.
22. H. Gasmi, J. Zhang, and C. Caragea, “Event extraction for cyber threat intelligence,” in *Proc. EMNLP Workshops*, 2018, pp. 22–31.
23. J. Han, Y. Li, and L. Wang, “SecurityBERT: A domain-adapted language model for cybersecurity,” *IEEE Access*, vol. 9, pp. 163342–163352, 2021.
24. K. Li, X. Chen, and Z. Li, “Deep learning for vulnerability classification using textual descriptions,” *Future Generation Computer Systems*, vol. 102, pp. 63–72, 2020.
25. S. Schaberreiter, M. Greifeneder, and E. Weippl, “Towards automated ATT&CK technique identification from threat reports,” in *Proc. ACM Workshop on Cyber Threat Intelligence*, 2021, pp. 11–20.
26. P. Mittal, S. Gupta, and A. Joshi, “Automating cyber threat intelligence: Challenges and opportunities,” *IEEE Security & Privacy*, vol. 19, no. 3, pp. 68–76, 2021.
27. M. Almukaynizi, E. Mariconti, and G. Stringhini, “On the accuracy of OSINT for cyber threat intelligence,” in *Proc. IEEE European Symp. Security and Privacy Workshops*, 2019, pp. 1–6.
28. R. Iannacone, G. Bridges, and J. Goodall, “Developing an ontology-driven framework for CTI,” in *Proc. IEEE Int. Conf. Intelligence and Security Informatics*, 2015, pp. 1–8.
29. B. E. Strom, C. Thomas, and A. Applebaum, “Evaluating the use of ATT&CK for adversary behavior modeling,” *MITRE Technical Report*, 2020.
30. M. Husák, M. Čermák, T. Jirsík, and P. Čeleda, “Survey of attack projection, prediction, and forecasting in cyber security,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 640–660, 2019.
31. E. Bou-Harb, M. Husák, and P. Čeleda, “Big data analytics for cyber threat intelligence,” *IEEE Communications Magazine*, vol. 56, no. 7, pp. 48–54, 2018.
32. Y. Liu, M. Ott, N. Goyal, et al., “RoBERTa: A robustly optimized BERT pretraining approach,” *arXiv preprint arXiv:1907.11692*, 2019.

33. MITRE Corporation, “MITRE ATT&CK: Techniques, tactics, and procedures,” 2023.
34. T. Liao, L. Chen, and J. Ji, “Cybersecurity named entity recognition using deep learning,” *Computers & Security*, vol. 85, pp. 1–14, 2019.
35. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
36. H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
37. S. Schaberreiter, M. Greifeneder, and E. Weippl, “Towards automated ATT&CK technique identification from threat reports,” in *Proc. ACM Workshop on Cyber Threat Intelligence*, 2021, pp. 11–20.
38. K. Saito and M. Rehmsmeier, “The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLOS ONE*, vol. 10, no. 3, pp. 1–21, 2015.
39. S. Elnagdy and A. Qasem, “Automated analysis of cyber threat intelligence reports,” *IEEE Access*, vol. 7, pp. 136210–136223, 2019.
40. R. Iannacone, G. Bridges, and J. Goodall, “Applying machine learning to cyber security data,” *IEEE Security & Privacy*, vol. 14, no. 2, pp. 78–81, 2016.
41. G. Tsoumakas and I. Katakis, “Multi-label classification: An overview,” *International Journal of Data Warehousing and Mining*, vol. 3, no. 3, pp. 1–13, 2007.
42. F. Martinelli, F. Mercaldo, and A. Santone, “Cybersecurity text analysis: Challenges and opportunities,” *IEEE Security & Privacy*, vol. 18, no. 5, pp. 76–82, 2020.
43. Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
44. J. Han, Y. Li, and L. Wang, “Context-aware deep learning for cyber threat intelligence,” *IEEE Access*, vol. 9, pp. 121345–121357, 2021.
45. A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on explainable artificial intelligence,” *IEEE Access*, vol. 6, pp. 52138–52160, 2018.
46. M. Almukaynizi, E. Mariconti, and G. Stringhini, “Label noise and uncertainty in cyber threat intelligence,” in *Proc. IEEE European Symp. Security and Privacy Workshops*, 2020, pp. 1–6.