2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Mapping Cyber Threat Intelligence through Active Semi-Supervised Learning (ASSBM)

Sujie Shao[1], Zhiyi Li[1*], Yan Liu[1], Shaoyong Guo[1] and Chao Yang[2]

[1] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 1008764, China
[2] Information and Communication Branch of State Grid Liaoning Electric Power Co., Ltd., Shenyang 110000, China
[*]hexing@bupt.edu.cn

Abstract. Cyber Threat Intelligence (CTI) plays a critical role in enhancing the implementation of cybersecurity programs by offering comprehensive information on attacks, which enables organizations to identify and respond to cyber threats more effectively. However, because most CTI data is presented in natural language and often contains ambiguous content, it requires interpretation and summarization by security experts for effective utilization. To address these challenges, this paper proposes a mapping method for CTI based on active and semi-supervised SecureBERT, aimed at alleviating the scarcity of labeled data and the ambiguities inherent in the CTI mapping task. This method efficiently extracts potential attack stage information from CTI at a minimal cost, ensuring accurate mapping even when labeled sample sizes are insufficient. We introduce an active learning sampling strategy that integrates uncertainty and instance relevance, selecting the most representative samples from unlabeled data to augment the training set. This strategy enhances the interpretability of labeled-scarce and ambiguous CTI, facilitating precise mappings between ambiguous CTI and the accurate phases of cyber attacks. Validation through experiments on the CPTC and CCDC datasets demonstrates that the proposed method excels across various baseline models, considering the influence of labeled data quantity and comparisons with different active learning algorithms. In situations where labeled CTI is limited, the proposed approach significantly improves the interpretive effectiveness of CTI, thereby enhancing the model's classification accuracy and training efficiency.

Keywords: CTI Mapping, Active Learning, SecureBERT, BERT.

1 Introduction

With the rapid development of information and communication technology, the Internet of Things, and Industry 5.0, modern cyber attacks have become increasingly diverse and covert, rendering traditional defenses like firewalls and rule-based intrusion detection insufficient against complex unknown threats [1]. Cyber Threat Intelligence (CTI) provides critical information for threat prediction, incident response, and security

enhancement. It enables analysts to understand threats more effectively and detect abnormal behavior within networks [2].

However, as network scale and threat complexity grow, Cyber Security Analysis Centers face mounting challenges. For instance, Venezuela's power grid suffered repeated cyber attacks in 2019–2020, leading to widespread blackouts [3], and in 2020 alone, a Chinese province's power grid experienced over 420,000 cyber attacks, with 65.4% deemed high-risk [4]. Analysts are overwhelmed by the vast volume of CTI reports [5], and inconsistencies across CTI sources often result in vague or disconnected information, limiting both accuracy and timeliness in threat response [6].

To address these challenges, researchers have explored methods to extract useful insights from scarce and ambiguous CTI. Semi-supervised learning, for example, leverages unlabeled data to generate pseudo-labels, enhancing classifier training [7]. Yet, its effectiveness depends heavily on the availability of quality labeled samples. Meanwhile, active learning improves labeling efficiency by selecting the most informative samples for expert annotation [8], although it still faces cost-related limitations. Dor et al. found that integrating active learning with BERT significantly boosts performance in real-world tasks [9].

Inspired by this, we propose a CTI mapping method based on Active and Semi-Supervised SecureBERT (ASSBM), designed to address the dual challenges of ambiguous CTI and limited labeled data. Our approach employs SecureBERT, a state-of-the-art pre-trained model trained on 1.1 billion cybersecurity-related tokens [10], and integrates both active and semi-supervised learning to expand the training set. Active learning selects uncertain and relevant samples for expert labeling, while semi-supervised learning generates pseudo-labels for high-confidence data, enhancing classification performance.

In summary, the main contributions of this work are as follows:

1. We propose a method for CTI mapping based on active and semi-supervised SecureBERT(ASSBM), addressing the issue of limited labeled data in CTI mapping tasks. The proposed method effectively interprets intelligence with minimal labeling while managing ambiguous sample data, thus achieving cost-efficient results.

2. We propose an active learning sampling strategy that integrates uncertainty and instance relevance to select the most representative samples from unlabeled data, effectively mitigating the challenges associated with data scarcity.

3. The proposed method has been validated through experiments using the CPTC and CCDC datasets, demonstrating superior performance. The approach is capable of efficiently completing the mapping of attack stages even with a limited amount of CTI data.

2 Background

Machine learning has been widely used for CTI interpretation by modeling logs and network traffic to classify behaviors. Common methods include SVM, Naive Bayes, Decision Trees, and Random Forests, often accelerated by Apache Spark. Given CTI's high-dimensional features, models tailored to specific attacks are necessary—for

example, improved AdaBoost has been used for DDoS detection [11]. However, these approaches rely heavily on large labeled datasets, which are scarce in practical cybersecurity contexts. To overcome this, recent studies have focused on few-shot learning, transfer learning, and active learning to enable effective generalization with limited labeled data.

2.1 Learning with Limited Samples

Few-shot learning and data augmentation are effective in scenarios with limited labeled data. Few-shot learning enables models to learn from small datasets and quickly adapt to new attack types. For example, Wang et al. [12] proposed ID-FSCIL to adapt to new attacks while maintaining performance on known ones. Yu and Bian [13] achieved 92.34% accuracy using under 1% of the NSL-KDD Train+ dataset. Lu et al. [14] developed a meta-learning-based IoT intrusion detection model, while Xu et al. [15] introduced FC-Net, a deep network that classifies traffic efficiently using prior knowledge. Data augmentation techniques improve performance by expanding the training set with synthetic or transformed data. Yash Madwanna et al. [16] used SMOTE to address sample scarcity, achieving 82.19% on UNSW-NB15 and 98.87% on NSL-KDD.

Unlike methods focusing solely on model-level improvements, our approach targets both model and data layers. It enhances diversity via data augmentation and integrates transfer learning for knowledge reuse, enabling better use of limited and unlabeled data and improving generalization.

2.2 Transfer Learning and BERT

Transfer learning, regarded as a form of few-shot learning [17], enables the reuse of pre-trained model weights for new tasks with few labeled samples. ULM-FiT [18] introduced this in NLP, dividing training into pre-training and fine-tuning phases to optimize task-specific performance with minimal data.

BERT [19] has proven effective in low-data NLP tasks like sentiment analysis and classification [20–23], though fine-tuning still requires significant data, and domain mismatch can reduce effectiveness in cybersecurity [24]. To solve this, SecureBERT [10] was developed as a domain-specific language model, enabling better transfer to cybersecurity tasks such as attack stage classification. It provides domain-relevant knowledge, outperforming general models in few-shot scenarios and improving generalization to novel threats.

2.3 Active Learning and Semi-Supervised Learning

Active learning and semi-supervised learning (SSL) are valuable for reducing reliance on labeled data. Active learning selects the most informative samples for annotation, improving model performance with minimal labeling. For example, Boukela et al. [25] proposed a hybrid active learning system for incremental intrusion detection, while Li et al. combined active learning and MMD-based transfer learning to reduce negative

transfer and enhance detection accuracy. SSL uses large pools of unlabeled data to boost model learning. Liu et al. [26] applied adversarial autoencoders for intrusion detection with minimal labels, and Vahidi et al. [27] introduced a collaborative SSL model validated on robust datasets. Despite its promise, SSL may suffer from misclassification due to the absence of human oversight.

By combining active learning with SSL, we leverage the strengths of both: active learning reduces annotation cost, while SSL increases training data and generalization capability [28–29]. This hybrid strategy leads to improved performance and training efficiency.

3 Methodology

In this section, we will employ the SecureBERT model, which is tailored for strong security-related applications, to perform the basic task of CTI mapping. We have designed an active semi-supervised learning framework to accelerate model training and address the scarcity of labeled CTI. First, we will provide an overview of ASSBM in Section 3.1, followed by a detailed description of the other components within the framework in subsequent sections.

3.1 Overview

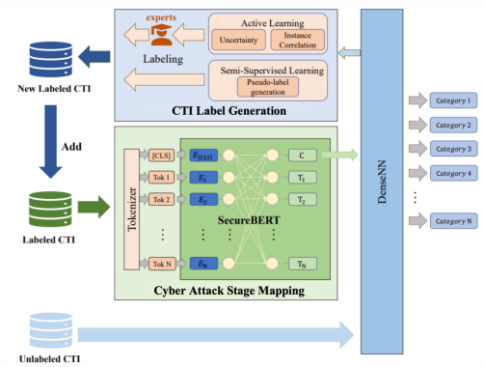

Fig. 1. The Architecture of ASSBM

The architecture of ASSBM is shown in Fig. 1. It comprises two core components: a transfer learning-based model for cyber attack stage interpretation and an active semi-supervised labeling algorithm for CTI generation. ASSBM operates cyclically through multiple rounds of data augmentation and model training. In each iteration, the training set includes both existing labeled CTI and newly labeled data generated by the labeling module.

The labeling process combines active learning and semi-supervised learning. Active learning selects high-value samples from the unlabeled CTI dataset using a strategy based on uncertainty sampling and instance relevance. These samples are then manually

annotated and added to the labeled dataset. In parallel, semi-supervised learning identifies high-confidence unlabeled instances, assigns pseudo-labels, and incorporates them into the training set. The CTI mapping model is built on transfer learning, where SecureBERT is fine-tuned using the evolving labeled dataset.

3.2 Cyber Attack Stage Mapping

Attack Stage Model. The proposed model maps ambiguous CTI to clear cyber attack stages, helping security analysts better interpret attack intent. Choosing a concise and intuitive stage model is crucial for overall effectiveness. We compare three widely used models—Cyber Kill Chain [30], MITRE ATT&CK [31], and AIF Attack Stage Model [32]—and construct a CTI-oriented classification standard. The Cyber Kill Chain provides a high-level view of attacker objectives but lacks the granularity required for CTI interpretation. Conversely, MITRE ATT&CK defines nearly 300 detailed techniques, from "Active Scanning" to "Kerberos Ticket Forgery," which, while comprehensive, can overwhelm analysts and degrade performance in low-resource scenarios.

To balance interpretability and detail, inspired by the CTI-oriented perspective proposed in [33], we design an attack stage model that characterizes adversarial behavior from both macro and micro levels. The macro level reflects what the attacker intends to achieve, while the micro level captures how the attack is executed. The model defines four macro stages: Reconnaissance and Scanning, Exploit, Maintenance Access, and Final Attack, along with sixteen micro behaviors. To maintain clarity and relevance, we exclude components with limited CTI observability, such as Passive Reconnaissance and Zero-Day Attacks. As shown in Table 1.

Table 1. Attack Stage Model

Macro Stages	Micro Stages	Description
Reconnaissance and Scanning	Host Discovery	Reconnaissance of the location/IP
	Service Discovery	Reconnaissance of services
	Vulnerability Discovery	Reconnaissance of vulnerabilities
	Information Discovery	Reconnaissance of technical information
Exploit	Privledge Esc.	Actions that gain user privileges
	Brute Force Access	Brute force cracking techniques
	Exploit Public Application	Attacking services that are open
	Exploit Remote Services	Connect to network using vpn etc.
	Arbitrary Code Execution	Arbitrary code execution
Maintenance Access	Defense Evasion	Techniques to evade detection
	Command & Control	Establishing a channel to control target
Final Attack	End Point Dos	Exploiting the system to cause crashes
	Network Dos	Depleting critical network bandwidth
	Service Stop	Stop services
	Data Exfiltration	Remove files and information
	Data Delivery	Data theft in the form of backdoors, etc.

Cyber Attack Stage Mapping Model. The proposed Transfer Learning-Based Cyber Attack Stage Mapping Model combines SecureBERT and DenseNN, aiming to leverage the pre-trained SecureBERT model for feature extraction and representation learning, while employing DenseNN layers to further process these feature representations and execute classification tasks. Specifically, the architecture encompasses the following key parameters and functions:

SecureBERT Model: The SecureBERT model is a domain-specific language model for cybersecurity, derived from the BERT framework. Within the architecture proposed in this paper, SecureBERT is responsible for converting input text sequences into high-dimensional semantic representations, encapsulating rich semantic features and contextual information.

DenseNN Layer: Building on the semantic representations output by the SecureBERT model, DenseNN (Dense Neural Network) functions as a classifier. It receives the semantic representations from SecureBERT as input and applies a multi-layer fully connected structure to further abstract and non-linearly transform the features. This process culminates in an output that corresponds to the number of classification labels.

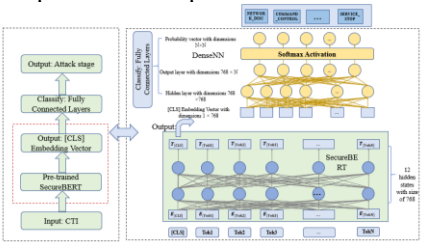

Fig. 2. Cyber Attack Stage Mapping Model

3.3 Active Semi-Supervised Algorithm for CTI Label Generation

While fine-tuning a pre-trained cybersecurity language model can interpret cyber attack stages (Section 3.2), its performance is constrained by limited labeled data, risking overfitting. To address this, we propose an Active Semi-Supervised CTI Label Generation Algorithm, combining active learning (AL) and semi-supervised learning (SSL). Early on, AL reduces labeling costs by selecting the most informative samples from limited labeled data. As labeled data grows, SSL leverages unlabeled data to improve generalization. This hybrid approach bridges the gap between labeled and unknown data, enhancing training efficiency and classification performance.

Active Learning Sampling Strategy. AL typically relies on uncertainty/entropy-based methods [34], where selecting high-value subsets can achieve strong performance with minimal data. However, when applied to BERT, while uncertainty can help identify the samples with the highest model prediction uncertainty for labeling, thereby assisting the model in exploring unknown knowledge domains, this approach may overlook the correlation and representativeness among samples. This could lead to selected samples being skewed in data distribution or repeated selection of similar yet information-poor

samples, resulting in wasted labeling resources. By integrating uncertainty with instance relevance in active learning, a more comprehensive evaluation of sample value can be achieved. This method considers not only the model's prediction uncertainty for individual samples but also the associations and diversity among samples. Through this approach, it is possible to select samples that are both challenging and representative, thus maximizing the utility of labeling resources.

Thus, this paper proposes an active sampling strategy integrating uncertainty and instance relevance for instance selection. This strategy no longer relies on a single approach, but rather simultaneously considers both sampling uncertainty and relevance. The details are as follows:

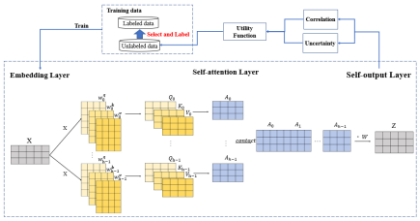

Fig. 3. Sampling Strategy Integrating Uncertainty and Instance Relevance

The fig.3. illustrates the training process of ASSBM under the integrated active learning strategy, which begins with a small amount of labeled data and a large pool of unlabeled data. ASSBM consists of multiple encoding layers, and we utilize the final encoding layer to extract instance relevance among samples. Meanwhile, we aggregate uncertainty measures to form a utility function. This function is applied to select informative data from the unlabeled data pool to supplement the training datasets.

Uncertainty. Calculating uncertainty in deep learning is challenging due to the models' lack of interpretability and unreliable class probabilities. To address this, Gal et al. [35] introduced "Dropout," which randomly drops neurons during training to improve generalization. In this paper, we apply this by dropping a percentage of neurons across iterations and using the standard deviation of the resulting classification probabilities as a measure of uncertainty.

Let the model be denoted as f_{nn}, and consider a data instance x. Let T represent the number of Dropout iterations, with the i Dropout configuration denoted by d_i. We compute the average output probability of the model across T iterations as follows:

$$p = \frac{1}{T}\sum_{i=0}^{T} f_{nn}^{d_i}(x) \tag{1}$$

The uncertainty measure is defined as the dispersion of the predictive probabilities across all model iterations:

$$R_{uncertain}(X) = \frac{1}{T}\sum_{i=0}^{T}\left[f_{nn}^{d_i}(x) - p\right]^2 \tag{2}$$

For a model, an uncertain prediction is associated with a higher dispersion value in its predictive probabilities (i.e., the uncertainty measure). This indicates that different parts

of the neural network exhibit conflicting activations in response to a given input. The uncertainty measure can assist in selecting samples that present challenges to the model for labeling, thereby enhancing overall model performance.

Instance Relevance. Regarding the relevance measure, we explore instance relevance by examining relationships between words in SecureBERT's final self-attention layer. By calculating the average variance of multi-head attention, we assess word correlations. Each word's importance is represented by its attention weights, and summing these weights gives insights into its significance in the sentence. The variance across attention heads reflects the level of correlation. Strong correlations suggest higher model confidence in predictions. The relevance measure is computed by analyzing the attention matrices of multiple heads to evaluate word interactions [36].

Let AM_i denote the attention matrix for head i, which reflects the influence of all words in the text on the updating of word representations. The intra-correlation is measured by jointly considering the interactions among words within each text. The calculation formula for intra-correlation is as follows:

$$AM_i = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}_{d_{vocab} \times d_{vocab}} \tag{3}$$

where n is the dimensionality of the words in the text.

We calculate the sum of attention weights for each word by summing the rows of AM_i, resulting in S_i. The intra-instance relevance $R_{\text{intra}}(X)$ is defined as the average variance across all attention heads:

$$\begin{aligned} S_i &= (s_1, \ldots s_n)^T = \begin{bmatrix} a_{11} + \cdots + a_{1n} \\ \vdots \\ a_{n1} + \cdots + a_{nn} \end{bmatrix} \\ \mathrm{s}' &= \frac{1}{n} \sum_{j=1}^{n} S_j \\ R_{\text{intra}}(X) &= \frac{1}{h} \sum_{i=0}^{h-1} \left(\frac{1}{n} \sum_{j=1}^{n} \left(s_j - \mathrm{s}' \right)^2 \right) \end{aligned} \tag{4}$$

Objective Function. Based on the definitions provided above, the selected instances satisfy the total objective function:

$$X_{sel} = \arg\min_{X} \left(\frac{R_{\text{intra}}(X)}{R_{uncertain}(X)} \right) \tag{5}$$

This consideration unfolds from two aspects:

1. Uncertainty Measure: High variance in predicted probabilities reflects ambiguity in the model's response.

2. Instance Relevance Measure: Low relevance indicates unclear meaning and weak logical connections within the instance. Lower confidence suggests ambiguity and inaccurate relationships.

Semi-Supervised Learning. In the Active Semi-Supervised Learning (ASSBM) framework, Semi-Supervised Learning (SSL) is employed to enhance the training dataset by generating pseudo-labels for unlabeled data. Initially, the model is trained using the labeled dataset. Subsequently, predictions are made on the unlabeled data, and samples with high prediction confidence are selected. These high-confidence samples are assigned pseudo-labels and incorporated into the training set alongside data selected through active learning. The model is then retrained on the augmented dataset to improve its performance.

Algorithm 1: ASSBM Algorithm for CTI Label Generation

Input: The labeled data L; The unlabeled data U; Classifier θ
Output: Indicators of classifier θ

function ASSBM(L, U, θ)
  while not meeting the stop criterion **do**
    $D_1, U = selectAL(L, U, \theta)$
    $D_2, U = selectSSL(L, U, \theta)$
    $L = D_1 + D_2$
    Train the classifier r model θ based on L ;
  end while
  return the result
end function
function selectAL(L, U, θ)
  for $x_i \in U$ **do**
    Select instance x_i according to Eq.(5)
    Query the label y_i of x_i from the expert
    Remove x_i from U
    Merge (x_i, y_i) into D_1
  end for
  return D_1, U
end function
function selectSSL(L, U, θ)
  for $x_i \in U$ **do**
    Select instance x_i from U based on high confidence levels
    Assign pseudo label to x_i
    Remove x_i from U
    Merge (x_i, y_i) into D_2
  end for
  return D_2, U
end function

Fig. 4. Active Semi-Supervised Algorithm for CTI Label Generation Algorithm

Algorithm. The proposed active and semi-supervised algorithm for CTI label generation operates in rounds. In each round, the newly added training dataset is split into two parts: one selected by active learning, the other by semi-supervised learning. The training set is composed of the existing labeled CTI data and new data generated by the label

generation module. Active learning picks critical data from the unlabeled CTI dataset using a strategy that combines uncertainty and instance relevance. After expert manual labeling, the data is added to the labeled dataset. Simultaneously, semi-supervised learning selects high-confidence unlabeled data, generates pseudo-labels, and adds it to the labeled set.

4 Experiment and analysis

4.1 Experiment setup

To evaluate the proposed algorithm, we used intrusion alert data from the CPTC [37] and CCDC [38] competitions to simulate ambiguous CTI. We performed 10-fold cross-validation, splitting the data into training, validation, and test subsets (6:2:2 ratio). The validation set was used to calculate metrics like loss, accuracy, and F1 score to monitor convergence, while the test set, kept separate, was used for final evaluation (metrics in Table 4 and Figures 7, 8).

For active learning labeling, instead of domain experts, we followed Sun et al. [39] to simulate expert annotations, selecting a portion of the data for labeling and the rest for active semi-supervised learning. We tested with thresholds of 10%, 15%, 20%, 25%, 30%, 35%, and 40%.

Table 2. Dataset

Data Source	**Numbers**
CPTC	3242
CCDC	2453
Total	5695

This paper adopts the evaluation metrics: top-1 accuracy, top-3 accuracy, F1 score, and loss value as evaluation indicators. Top-1 accuracy and top-3 accuracy represent the percentage of correctly predicted results among all samples and the percentage of samples for which at least one of the top three predicted results is correct, respectively. The F1 score provides an overall evaluation based on precision and recall.

To verify ASSBM can significantly improve the interpretability of CTI under sparse data conditions, the algorithm is compared with three other baseline models:

Table 3. Baseline Model

Model	**Description**
bert-base-uncased[19]	Basic BERT Model
SecureBERT[10]	Basic SecureBERT Model
SecureBERT-AL	SecureBERT Model with Active Learning
SecureBERT-SSL	SecureBERT Model with Semi-Supervised Learning
ASSBM	**Our work**

To ensure the rigor of the experimental results, this paper will repeat each algorithm experiment ten times and compute the average as the final result. The experiments were

conducted using an NVIDIA RTX 4060 GPU with 32GB RAM, and the software environment included Python 3.9, PyTorch 2.0.0. The optimizer uses Adam, the learning rate is fixed at 2e-5 (determined by grid search), and the batch size is 16.

4.2 Experimental Results

Comparison of Different Baseline Models. We examine the effect of iteration count on training efficacy, where performance ideally improves as iterations increase, with each iteration augmenting the labeled dataset. Figure 1 depicts the impact of iterations on classification accuracy across models. SecureBERT outperforms basic BERT, while the semi-supervised approach exhibits instability due to reliance on pseudo-labeling. The active learning-only model stabilizes early, as less informative samples do not significantly enhance performance. In contrast, ASSBM effectively identifies valuable samples, minimizing labeled data and training costs while enhancing sample evaluation. The proposed algorithm consistently outperforms both active learning and semi-supervised methods, achieving superior accuracy in explaining macro and micro attack phases.

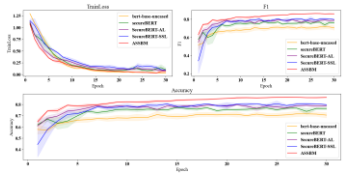

Fig. 5. Comparison across Different Models for Macro Attack Stages

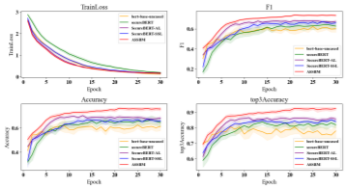

Fig. 6. Comparison across Different Models for Micro Attack Stages

Comparison of Different Labeled Data Proportion. We selected a specific proportion of the complete dataset as labeled data for input into the model training, while utilizing the remaining data for active semi-supervised learning. Seven threshold values were chosen: 10%, 15%, 20%, 25%, 30%, 35%, and 40%. To account for the impact of the initial amount of labeled data on the experimental results, the specific results are displayed in the table 4. As observed, when the amount of labeled data is limited, all models perform poorly. The SecureBERT model shows a slight improvement in accuracy and F1 score compared to the basic BERT model. The BERT model that incorporates active learning demonstrates a more significant enhancement in accuracy—averaging around a 5% increase—compared to models without active learning. Moreover,

models based solely on semi-supervised learning yield lesser performance compared to those employing active learning.

The proposed ASSBM framework significantly enhances the classification accuracy of the models. Unlike methods that rely solely on active learning or semi-supervised learning, the model presented in this paper achieves an additional 10% improvement.

Table 4. The Impact of Different Proportions of Labeled Data on Classification Accuracy

The impact of annotated data proportion on **macro** attack stage classification

Methods	Proportion of labaled data															
	10%		15%		20%		25%		30%		35%		40%			
	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy		
bert-base-uncased	0.549	0.546	0.593	0.592	0.640	0.640	0.666	0.656	0.671	0.677	0.703	0.703	0.718	0.726		
secureBERT	0.651	0.653	0.662	0.667	0.705	0.711	0.718	0.712	0.735	0.736	0.741	0.741	0.774	0.776		
secureBERT-AL	0.646	0.664	0.703	0.699	0.743	0.740	0.754	0.755	0.776	0.776	0.795	0.796	0.804	0.804		
secureBERT-SSL	0.648	0.653	0.698	0.693	0.720	0.720	0.782	0.783	0.798	0.792	0.793	0.789	0.810	0.806		
ASSBM	0.780	0.780	0.809	0.810	0.814	0.814	0.844	0.844	0.849	0.844	0.864	0.865	0.872	0.871		

The impact of annotated data proportion on **micro** attack stage classification

Methods	Proportion of labaled data															
	10%		15%		20%		25%		30%		35%		40%			
	F1	top1Acc(top3)	F1	top1Acc(top3)	F1	top1Acc(top3)	F1	top1Acc(top3)	F1	top1Acc(top3)	F1	top1Acc(top3)	F1	top1Acc(top3)		
bert-base-uncased	0.423	0.423(0.653)	0.458	0.463(0.692)	0.477	0.487(0.717)	0.518	0.523(0.723)	0.534	0.538(0.728)	0.577	0.577(0.742)	0.579	0.588(0.752)		
secureBERT	0.463	0.469(0.687)	0.516	0.513(0.716)	0.608	0.613(0.770)	0.615	0.623(0.797)	0.632	0.641(0.824)	0.647	0.651(0.824)	0.640	0.651(0.826)		
secureBERT-AL	0.554	0.562(0.772)	0.598	0.630(0.803)	0.638	0.646(0.831)	0.657	0.660(0.838)	0.660	0.668(0.849)	0.669	0.671(0.857)	0.678	0.681(0.856)		
secureBERT-SSL	0.537	0.544(0.738)	0.589	0.596(0.831)	0.619	0.629(0.825)	0.628	0.632(0.831)	0.648	0.651(0.843)	0.656	0.661(0.854)	0.660	0.667(0.851)		
ASSBM	0.618	0.620(0.814)	0.687	0.689(0.871)	0.709	0.717(0.880)	0.726	0.729(0.891)	0.732	0.735(0.898)	0.753	0.756(0.923)	0.757	0.758(0.925)		

As the quantity of labeled data increases, this model can achieve an accuracy of 87.1% for macro attack phases, and for micro attack phases, a top-1 accuracy of 75.8% and a top-3 accuracy of 92.5%. This demonstrates that the proposed active semi-supervised network for CTI label generation effectively integrates the advantages of both active and semi-supervised learning. It reduces labeling costs while maximizing the use of limited labeled data and abundant unlabeled data to train the model, addressing the discrepancies between labeled datasets and unknown data. Consequently, this improves the training process, accelerates model training, and enhances classification performance, as illustrated in the figure below:

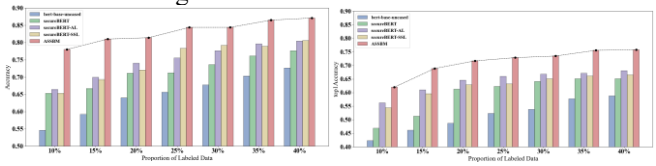

Fig. 7. Impact of Labeled Data Proportion on Accuracy for Macro Attack Stages(left)

Fig. 8. Impact of Labeled Data Proportion on Accuracy for Micro Attack Stages(right)

Comparison of Different Active Learning Algorithms. To validate our proposed active learning strategy combining uncertainty and instance relevance, we compared ASSBM against three baseline methods - Least Confidence, Maximum Entropy, and Random Sampling - under controlled conditions with 40% labeling rate and semi-supervised learning.

The experimental results (Figs. 9-10) demonstrate ASSBM's superior performance across all metrics. While traditional methods like Least Confidence and Maximum Entropy showed slow improvements in BERT before stabilizing around iteration 15, they tend to select uncertain samples without considering their representativeness. This often leads to biased or redundant sample selection and inefficient use of labeling resources.

In contrast, ASSBM achieved more stable performance by iteration 20 through its integrated approach. By simultaneously evaluating both prediction uncertainty and

instance relevance, our method identifies samples that are both challenging to the model and representative of the data distribution. This dual consideration enables more efficient use of labeling resources while significantly improving the model's generalization capabilities.

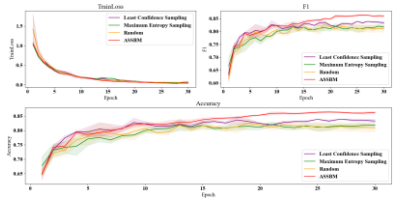

Fig. 9. Comparison of Different Active Learning Algorithms for Macro Attack Stages

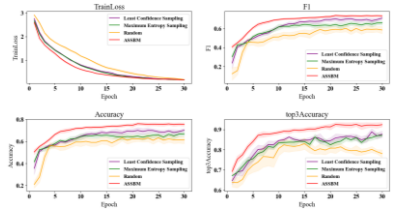

Fig. 10. Comparison of Different Active Learning Algorithms for Micro Attack Stages

Comparison with Existing Methods. We conducted an experiment to compare our proposed method with existing works in the domains of few-shot learning and data augmentation[13-16], which were discussed in Section 2.1. The experiments were conducted using the same datasets and experimental settings described in the respective papers to ensure a fair comparison. The datasets used include NSL-KDD [41], UNSW-NB15 [42], and CICIDS2017 [43]. Considering that the evaluation metrics varied across the selected papers, we chose accuracy as the sole metric for comparison since it is the only common metric reported in all studies. The comparison results are summarized in Fig.11 and Table 5, which highlights the accuracy of our proposed method against the selected prior methods.

Table 5. Comparison with Existing Methods

Reference	Method	Dataset	Avg Accuracy (training/testing)
Yu and Bian[13]	DNN+CNN	UNSW-NB15 NSL-KDD	**0.920/0.941** **0.923/0.933**
Lu et al.[14]	Meta-Learning	UNSW-NB15 CICIDS2017	**0.880/0.901** **0.887/0.911**
Xu et al.[15]	Meta-Learning	NSL-KDD	**0.923/0.944**
Yash Madhavan et al. [16]	Deep Learning	UNSW-NB15 NSL-KDD	**0.822/0.843** **0.989/0.984**

Our method consistently demonstrated competitive performance and outperformed most prior methods across the datasets. Specifically, on UNSW-NB15, our method achieved an accuracy of 0.941, surpassing the deep learning-based approach by 2.1%. On NSL-KDD, although one prior methods [16] reported a slightly higher accuracy of 0.989, our method still maintained a comparable performance of 0.984 while showcasing advantages across other methods. On CICIDS2017, our method achieved 0.911, outperforming the meta-learning-based model by 2.4%.

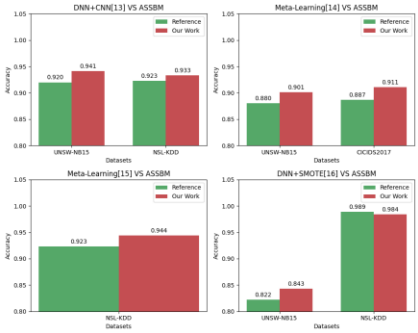

Fig. 11. Comparison with Existing Methods

Our approach achieves superior performance by optimizing both model and data layers. Unlike few-shot learning methods that focus mainly on model structures, we integrate data-level strategies like data augmentation and transfer learning to increase training data diversity and size, improving model generalization to unseen attack types. Additionally, active learning selects the most informative samples for labeling, while semi-supervised learning leverages unlabeled data. These elements collectively enhance the model's robustness and adaptability, demonstrating its effectiveness in scenarios with limited labeled data.

5 Conclusion

In conclusion, this paper addresses the challenges of label scarcity and information ambiguity in CTI by proposing a method that maps CTI effectively to specific attack phases. Utilizing active and semi-supervised SecureBERT, our approach enhances the extraction of relevant attack stage information even with limited labeled data. The active learning sampling strategy prioritizes uncertainty and instance relevance, allowing for the selection of representative unlabeled samples to enrich training datasets. Experimental results on the CPTC and CCDC datasets demonstrate that our method significantly improves interpretability and classification accuracy, paving the way for more effective use of CTI in cybersecurity responses. Future work may consider refining this technique and extending its applications across various cybersecurity contexts.

Acknowledgments. This work is supported by National Key R&D Program of China (2022YFB3105101).

Disclosure of Interests. All authors disclosed no relevant relationships.

References

1. Alaeifar P, Pal S, Jadidi Z, et al. Current approaches and future directions for Cyber Threat Intelligence sharing: A survey[J]. Journal of Information Security and Applications, 2024, 83: 103786.
2. Sun, Nan, et al. "Cyber threat intelligence mining for proactive cybersecurity defense: a survey and new perspectives." IEEE Communications Surveys & Tutorials 25.3 (2023): 1748-1774.
3. Yang S, Lao K W, Hui H, et al. Secure distributed control for demand response in power systems against deception cyber-attacks with arbitrary patterns[J]. IEEE Transactions on Power Systems, 2024.
4. Yang, X. (2023, June 26). Can Virtual Power Plants Become the Optimal Solution for Power Regulation? China Energy News. https://paper.people.com.cn/zgnyb/html/2023-06/26/content_26002578.htm.
5. Allegretta, Mauro, et al. "Are crowd-sourced CTI datasets ready for supporting anti-cybercrime intelligence?." Computer Networks 234 (2023): 109920.
6. Nadeem A, Verwer S, Moskal S, et al. Alert-driven attack graph generation using s-pdfa[J]. IEEE transactions on dependable and secure computing, 2021, 19(2): 731-746.
7. Zhao F, Zhang H, Peng J, et al. A semi-self-taught network intrusion detection system[J]. Neural Computing and Applications, 2020, 32: 17169-17179.
8. Ren P, Xiao Y, Chang X, et al. A survey of deep active learning[J]. ACM computing surveys (CSUR), 2021, 54(9): 1-40.
9. Dor L E, Halfon A, Gera A, et al. Active learning for BERT: an empirical study[C]//Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP). 2020: 7949-7962.
10. Aghaei E, Niu X, Shadid W, et al. Securebert: A domain-specific language model for cybersecurity[C]//International Conference on Security and Privacy in Communication Systems. Cham: Springer Nature Switzerland, 2022: 39-56.
11. Tang D, Tang L, Dai R,et al., “MF-Adaboost: LDoS attack detection based on multifeatures and improved Adaboost[J],” Future Generation Computer Systems, vol.106, pp.347-359,2020.
12. Wang T, Lv Q, Hu B, et al. A few-shot class-incremental learning approach for intrusion detection[C]//2021 International Conference on Computer Communications and Networks (ICCCN). IEEE, 2021: 1-8.
13. Yu Y, Bian N. An intrusion detection method using few-shot learning[J]. IEEE Access, 2020, 8: 49730-49740.
14. Lu C, Wang X, Yang A, et al. A Few-Shot-Based Model-Agnostic Meta-Learning for Intrusion Detection in Security of Internet of Things[J]. IEEE Internet of Things Journal, 2023, 10(24): 21309-21321
15. Xu C, Shen J, Du X. A method of few-shot network intrusion detection based on meta-learning framework[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 3540-3552.

16. Madwanna Y, Annappa B, Sneha H R. YARS-IDS: A novel IDS for multi-class classification[C]//2023 IEEE 8th International Conference for Convergence in Technology (I2CT). IEEE, 2023: 1-6.
17. Bayer M, Frey T, Reuter C. Multi-level fine-tuning, data augmentation, and few-shot learning for specialized cyber threat intelligence[J]. Computers & Security, 2023, 134: 103430.
18. Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 328–339, Melbourne, Australia. Association for Computational Linguistics.
19. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
20. Roy A, Pan S. Incorporating medical knowledge in BERT for clinical relation extraction[C]//Proceedings of the 2021 conference on empirical methods in natural language processing. 2021: 5357-5366.
21. Zhou X, Huang H, Chi Z, et al. RS-BERT: Pre-training radical enhanced sense embedding for Chinese word sense disambiguation[J]. Information Processing & Management, 2024, 61(4): 103740.
22. Garrido-Merchan E C, Gozalo-Brizuela R, Gonzalez-Carvajal S. Comparing BERT against traditional machine learning models in text classification[J]. Journal of Computational and Cognitive Engineering, 2023, 2(4): 352-356.
23. Aftan S, Shah H. A survey on bert and its applications[C]//2023 20th Learning and Technology Conference (L&T). IEEE, 2023: 161-166.
24. Widmann, T., and Wich, M.. 2022. “Creating and Comparing Dictionary, Word Embedding, and Transformer-Based Models to Measure Discrete Emotions in German Political Text.” Political Analysis, 1–16.
25. Boukela L, Zhang G, Yacoub M, et al. A near-autonomous and incremental intrusion detection system through active learning of known and unknown attacks[C]//2021 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). IEEE, 2021: 374-379.
26. Li J, Wu W, Xue D. An intrusion detection method based on active transfer learning[J]. Intelligent Data Analysis, 2020, 24(2): 363-383.
27. Liu X, Luo E, Yang J, et al. Semi-supervised intrusion detection method based on adversarial autocoder[C]//2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). IEEE, 2021: 637-643.
28. Sun X, Tu L, Zhang J, et al. ASSBert: Active and semi-supervised bert for smart contract vulnerability detection[J]. Journal of Information Security and Applications, 2023, 73: 103423.
29. Vahidi J, Ahmadzadeh M. A Comprehensive Semi-Suprvised Model for Collaborative Intrusion Detection Based on Network Behavior Profiling Using The Concept of Deep Learning and Fuzzy Correlation of Alerts along[J]. Electronic and Cyber Defense, 2021, 9(3): 165-186.
30. LockheedMartin.com. (2011) The Cyber Kill Chain. [Online]. Available: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html.
31. Cybotsai.com. (2021) An introduction to MITRE ATT&CK. [Online]. Available: https://cybotsai.com/introduction-mitre-attck/
32. Moskal S, Yang S J. Cyberattack action-intent-framework for mapping intrusion observables[J]. arXiv preprint arXiv:2002.07838, 2020.

33. Moskal S, Yang S J. Cyberattack action-intent-framework for mapping intrusion observables[J]. arXiv preprint arXiv:2002.07838, 2020.
34. Tharwat A, Schenck W. A survey on active learning: State-of-the-art, practical challenges and research directions[J]. Mathematics, 2023, 11(4): 820.
35. Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning[C]//international conference on machine learning. PMLR, 2016: 1050-1059.
36. Zhang A, Li B, Wang W, et al. MII: A Novel Text Classification Model Combining Deep Active Learning with BERT[J]. Computers, Materials & Continua, 2020, 63(3).
37. Rochester Institute of Technology, "Collegiate penetration testing competition," hhttp://nationalcptc.org, 2018, [Online; accessed 19-July-2018].
38. F. Hassanabad, "suricata-sample-data," https://github.com/FrankHassanabad/suricata-sampledata/blob/master/README.md, 2019, [Online; accessed 5-May-2020].
39. Sun X, Tu L, Zhang J, et al. ASSBert: Active and semi-supervised bert for smart contract vulnerability detection[J]. Journal of Information Security and Applications, 2023, 73: 103423.
40. Moskal S, Yang S J. Translating intrusion alerts to cyberattack stages using pseudo-active transfer learning (PATRL)[C]. 2021 IEEE conference on communications and network security (CNS). IEEE, 2021: 110-118.
41. M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2009.
42. Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)[C]//2015 military communications and information systems conference (MilCIS). IEEE, 2015: 1-6.
43. Sharafaldin I, Lashkari A H, Ghorbani A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[J]. ICISSp, 2018, 1: 108-116.