Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7778346

SELECTING FITTING MACHINE LEARNING METHODS TO CONSTRUCT A HYBRID CYBER THREAT INTELLIGENCE MODEL

Alemayehu Tilahun Haile[1], Surafel Lemma[2], Henock Mulugeta[3], *

[1.] School of Information Technology and Engineering (SiTE), P.O. Box 385, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia.
alemayehu.tilahun@aau.edu.et/alemayehuthaile1@gmail.com

[2.] School of Information Technology and Engineering (SiTE), P.O. Box 385, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia.
surafel.lemma@aait.edu.et/surafel.lemma@aait.edu.et

[3.] School of Information Technology and Engineering (SiTE), P.O. Box 385, Addis Ababa Institute of Technology, Addis Ababa, Ethiopia. henock.mulugeta@aait.edu.et/ henockmulugeta26@gmail.com

* Correspondence: alemayehu.tilahun@aau.edu.et/alemayehuthaile1@gmail.com

Abstract

With the rising sophistication of cyber threats, the utilization of machine learning algorithms for cyber threat intelligence (CTI) has become increasingly crucial. This research presents a comprehensive comparative analysis of various deep learning (DL) algorithms, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Generative Adversarial Networks (GAN), and Bidirectional Encoder Representations from Transformers (BERT), in the context of CTI using open-source intelligence (OSINT) data. A specific dataset encompassing NER, sentiment analysis, text classification, and information extraction tasks was employed to evaluate the algorithms' performance. The comparison was based on a set of well-established metrics, such as task flexibility, training data requirements, training time, accuracy, precision, and F1-score. The paper results unveiled that while CNN, RNN, and LSTM demonstrated competitive performance in certain tasks, BERT consistently outperformed the other algorithms across multiple metrics and NLP tasks. BERT's superior performance can be attributed to its contextualized word embeddings and advanced attention mechanisms that effectively capture intricate relationships in text.

Keywords: Comparative analysis, deep learning algorithms, cyber threat intelligence, OSINT, NER, LSTM, information extraction, training time, accuracy, BERT.

1. Introduction

According to a study conducted by the Ponemon Institute (Smith, 2020), cyber attacks are escalating in frequency and sophistication, presenting a significant challenge to organizations worldwide. The paper highlights that attackers are leveraging advanced techniques and tools, such as zero-day exploits, polymorphic malware, and evasion tactics, to circumvent traditional security measures. Furthermore, the study points out that threat actors are increasingly focusing on targeted attacks, utilizing social engineering and spear-phishing tactics to exploit human vulnerabilities within organizations. The findings underscore the evolving nature of cyber threats, necessitating a proactive strategy that encompasses real-time threat intelligence, As cyber attacks continue to evolve and become more sophisticated, organizations must remain adaptive and employ robust security measures to mitigate risks effectively.[1], [2]

Cyber Threat Intelligence (CTI) plays a crucial role in addressing the aforementioned challenges by providing organizations with valuable insights and proactive measures to combat evolving cyber threats.[3]CTI enables organizations to gather, analyze, and interpret relevant information about potential threats, including emerging attack techniques, threat actor motivations, and indicators of compromise. This information empowers organizations to proactively identify and mitigate potential risks, strengthen their security posture, and make informed decisions regarding resource allocation and incident response.[4]

OSINT (Open Source Intelligence) has emerged as a critical component in the field of Cyber Threat Intelligence (CTI) and has garnered significant attention in recently published papers. A study by Scaife et al. (2021) emphasizes the indispensable role of OSINT in providing valuable insights into potential cyber threats. It highlights how OSINT sources, such as social media, public databases, and online forums, offer rich and diverse information that can aid in the identification, analysis, and mitigation of cyber risks.[5]OSINT plays a crucial role in collecting real-time information about threat actors, their motivations, and attack patterns.[6] the integration of OSINT into CTI processes, contributes to early warning, incident response, and threat hunting.[7]

In recent years, various machine learning (ML) algorithms have been introduced and applied to Cyber Threat Intelligence (CTI) utilizing Open Source Intelligence (OSINT) sources. These algorithms have demonstrated their effectiveness in enhancing threat detection, intelligence analysis, and decision-making processes. Supervised learning algorithms, such as decision trees, support vector machines (SVMs), and random forests, have been utilized for classifying and predicting cyber threats based on labeled training data.[8]. Deep learning algorithms, specifically neural networks, have also gained popularity in CTI for their ability to process large volumes of complex OSINT data, detect intricate patterns, and enable advanced threat analysis.[9] Transformer based models like BERT plays a crucial role in CTI with OSINT by leveraging its contextual understanding capabilities. By effectively processing unstructured text data, BERT enhances tasks such as sentiment analysis, entity recognition, and document classification.[10]

In this paper, the aim to contribute to the field of Cyber Threat Intelligence (CTI) utilizing Open Source Intelligence (OSINT) by providing a comprehensive comparison of different machine learning algorithms. the research focuses on the evaluation and analysis of various algorithms using a specific dataset tailored to CTI with OSINT. By conducting a systematic comparison, we strive to identify the strengths and weaknesses of each algorithm in terms of accuracy, efficiency, interpretability, and scalability. Furthermore, we consider the specific characteristics of the dataset, such as volume, velocity, variety, and veracity, to assess how different algorithms perform under these conditions. The insights gained from this study will assist security practitioners in making informed decisions regarding algorithm selection for CTI with OSINT, enabling them to choose the most suitable algorithm based on the specific requirements and constraints of their dataset. Ultimately, our research aims to enhance the effectiveness of CTI efforts by providing guidance on selecting the best machine learning algorithm for a given CTI with OSINT dataset.

2. Related works

Paper [11] presents a pipeline that aims to improve and expand the capabilities of a cyberthreat discovery tool currently in development. The tool is capable of gathering, processing, and presenting security related tweets. For this purpose, the paper implemented two neural networks. The first is a binary classifier based on a Convolutional Neural Network architecture. This classifier is able to identify if a tweet contains security related information about a monitored infrastructure or not. Then it is forwarded to a Named Entity Recognition model. This model is implemented by a Bidirectional Long Short-Term. The paper achieved 90% accuracy in their Named Entity Recognition model.

The author of[12] have made use of natural language processing to extract threat feeds from unstructured cyber threat information sources with approximately 70% precision, providing comprehensive threat reports in standards like STIX, which is a widely accepted industry standard that represents CTI. The automation of an otherwise tedious manual task would ensure the timely gathering and sharing of relevant CTI that would give organizations the edge to be able to proactively defend against known as well as unknown threats.

Paper [13] employees machine learning and deep learning approach using neural network to automatically classify hacker forums data in to predefined categories and develop interactive visualization that enables CTI practitioners to probe collected data for proactive and opportune CTI. The result from the paper shows that from all the models, deep learning model RNN GRU gives the best classification result with 99.025% accuracy and 96.56% precision.

In the paper [14] the authors present a framework for detection and classification of cyber threat indicators in the Twitter stream. Contrary to the bulk of similar proposals that rely on manually-designed heuristics and keyword-based filtering of tweets, the framework provides a data-driven approach for modeling and classification of tweets that are related to cybersecurity events. They present a cascaded Convolutional Neural Network (CNN) architecture, comprised of a binary classifier for detection of cyber-related tweets, and a multi-class model for the classification of cyber-related tweets into multiple types of cyber threats.

The researchers of [15]investigate how the latest advances in the NER domain, and in particular transformer-based models, can facilitate the process of NER. The dataset for NER in Threat Intelligence (DNRTI) containing more than 300 pieces of threat intelligence reports from open source threat intelligence websites is used. Their experimental results demonstrate that transformer-based techniques are very effective in extracting cybersecurity-related named entities, by considerably outperforming the previous state- of-the-art approaches tested with DNRTI.

Research paper [16] researches on threat intelligence entity recognition model. They use the BERT model as a corpus pre-training model based on the classic neural network BiLSTM-CRF, and proposes a model DT-BERT-BiLSTM-CRF based on the dictionary template. The BERT pre-training model makes full use of the contextual semantic information of the corpus and alleviates the problem of ambiguity in the process of threat intelligence entity recognition. By constructing a dictionary template of threat intelligence entities, the accuracy of entity recognition in the threat intelligence field is further improved.

The table below presents a comprehensive comparison of different previously conducted research papers.

Table 1. Comparison between proposed attack classification methods.

Ref	NLP task	Algorithm	Dataset	Accuracy	pericisoin
[11]	NER	Bi LSTM	Securityrelatedtweets dataset	90%	
[12]	NER	CRF	FireEye andKaskperskySecurity Lab datasets		70%
[13]	Information extraction	GRU	hacker forums datasets	99.025%	96.56%
[14]	Text classification	CNN	Twitter stream dataset	95%	
[15]	NER	Transformers	DNRTI dataset	97%	
[16]	NER	BERT	corpus pre-training dataset	98%	

3. Proposed Model

BERT (Bidirectional Encoder Representations from Transformers) has emerged as a powerful tool for cyber threat intelligence (CTI) and open-source intelligence (OSINT). As a pre-trained deep learning model, BERT excels at capturing contextualized word embeddings, allowing it to grasp the intricate nuances and relationships in natural language.

In the realm of CTI, BERT's capability to process vast amounts of textual data from diverse sources makes it highly effective in tasks such as sentiment analysis, named entity recognition, and information extraction. By understanding the context and semantics of language, BERT enables more accurate and sophisticated analysis of cyber threats, aiding in the detection of malicious activities and enhancing the overall cybersecurity defense. Its versatility and adaptability to various CTI challenges have made BERT a valuable asset for intelligence analysts and cybersecurity professionals working with OSINT data. [10]

The key innovation in BERT's architecture lies in its bidirectional attention mechanism, which allows it to process words in a sentence in both forward and backward directions, capturing the full context and dependencies of each word. This contextualized word embedding enables BERT to comprehend the intricacies of natural language and better understand the meaning of words based on their surrounding context.

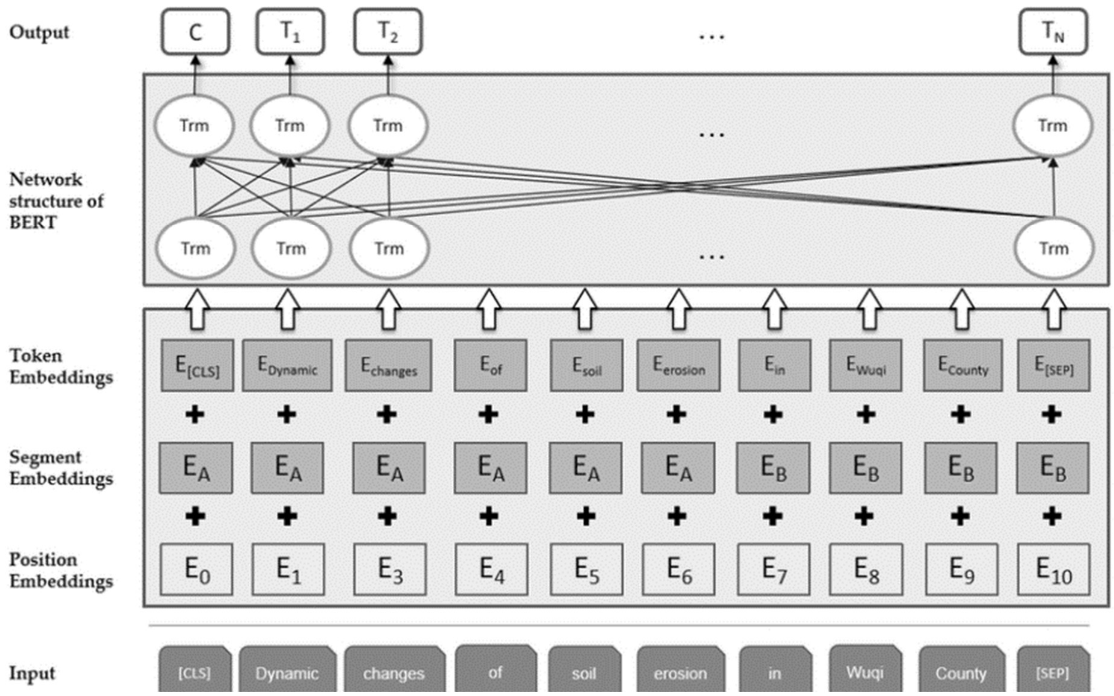

Figure 1. Architecture of BERT model[17]

Before feeding word sequences into BERT, 15% of the words in each sequence are replaced with a [MASK] token. The model then attempts to predict the original value of the masked words, based on the context provided by the other, non-masked, words in the sequence. In technical terms, the prediction of the output words requires:Adding a classification layer on top of the encoder output, Multiplying the output vectors by the embedding matrix, transforming them into the vocabulary dimension and Calculating the probability of each word in the vocabulary with SoftMax.

Fine-tuning BERT with our dataset involves adapting its architecture to our specific Natural Language Processing (NLP) task. This customization includes modifying the input and output layers to match the NLP format and adjusting the pooling layer for effective training. By fine-tuning on our labeled data, BERT's pre-trained representations are refined to excel in identifying in the selected NLP task within our dataset. This process tailors BERT's capabilities, resulting in a more accurate and contextually relevant model, aligning with our specific data characteristics and task requirements.

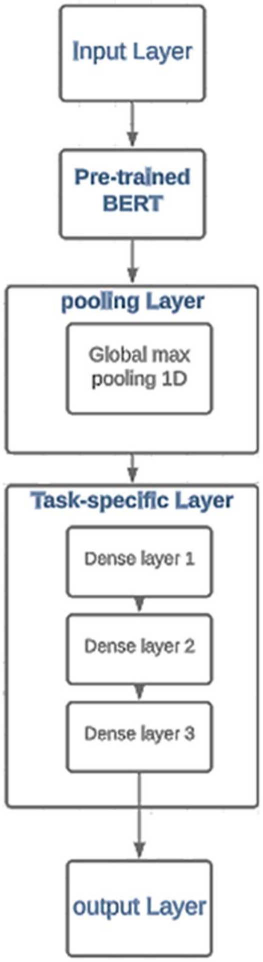

Figure 2. Fine-tuned architecture of BERT model

4. Implementation

4.1 Dataset used

The dataset "Tweets Related to Cyber Attacks" serves as a valuable resource for comparing different kinds of machine learning algorithms in the context of Cyber Threat Intelligence (CTI) using Open Source Intelligence (OSINT). By utilizing this dataset, researchers and analysts can evaluate the performance of various machine learning algorithms in classifying and analyzing tweets related to cyber attacks.

The dataset's diverse collection of tweets covering different types of cyber attacks allows for a comprehensive comparison of machine learning algorithms. Researchers can explore the effectiveness of algorithms such as CNN, RNN, LSTM, GAN and BERT on natural language processing (NLP) techniques. These algorithms can be evaluated based on their ability to accurately classify tweets, extract relevant features, and identify patterns or indicators of cyber attacks.

By comparing the performance of different machine learning algorithms on this dataset, researchers can gain insights into the strengths and weaknesses of each algorithm. They can analyze metrics such as accuracy, precision, recall, and F1-score to assess the performance of algorithms in detecting and categorizing cyber attacks mentioned in the tweets. Additionally, researchers can explore algorithmic efficiency, interpretability, and scalability to understand the practicality of these algorithms for CTI with OSINT.

The dataset also enables the evaluation of algorithms in real-world scenarios, as the tweets are derived from social media platforms where real-time information is shared. This provides a more accurate representation of the challenges faced in analyzing OSINT data for CTI.Overall, the "Tweets Related to Cyber Attacks" dataset facilitates the comparison of machine learning algorithms, allowing researchers to identify the most effective approach for cyber attack detection and classification based on OSINT derived from social media platforms.

The research objective is to compare different kinds of machine learning algorithms using the "Tweets Related to Cyber Attacks" dataset. The goal is to evaluate their performance in accurately classifying cyber attack-related tweets, and examine their interpretability for Cyber Threat Intelligence (CTI) purposes. Additionally, the study will include visualizations of the evaluation metrics to provide a clear and intuitive representation of the algorithms' performance. By accomplishing these objectives and presenting the findings with visualizations, the research aims to provide valuable insights for selecting the most suitable deep learning algorithm for CTI using OSINT from social media platforms.

The selected dataset sources comprise diverse and comprehensive records of known cyber attacks. With a rich array of textual information, this dataset encompasses various types of cyber threats, including malware campaigns, phishing attacks, DDoS attacks, data breaches, and other malicious activities. Each entry in the dataset provides valuable insights into the tactics, techniques, and procedures (TTPs) utilized by threat actors during these attacks

Table 1. Attack Categories in the Dataset.

Attack categories	Percentage
Ransomware	20%
DDoS	23%
Data breach	7%
Zero day	6%
Botnet	15%
general	14%
Phishing	15%

BERT relies on a Transformer, which is an attention mechanism that learns contextual relationships between words in a text. A basic Transformer consists of an encoder to read the text input and a decoder to produce a prediction for the task. Since BERT's goal is to generate a language representation model, it only needs the encoder part. Following Algorithm 1 is the Pseudo Code for BERT Model Loading and Training.

Algorithm 1 **Pseudo Code for BERT Model Loading and Training**

```
import tensorflow
Load BERT model; bert_model =
hub.KerasLayer("PATH/tensorflow/small_bert/bert_en_uncased ")
 Load BERT Preprocess model
bert_preprocess = hub.KerasLayer("PATH /tensorflow/bert_en_uncased_preprocess")
```

```
4:  Prepare the data (This is a placeholder, for preprocessing code)
data = ...
5:  Prepare the model for training
model = tf.keras.models.Sequential([
   bert_preprocess,
   bert_model,
   tf.keras.layers.Dense
6:  Compile the model
model.compile(loss=tf.keras.losses.BinaryCrossentropy(from_logits=True),
           optimizer=tf.keras.optimizers.Adam(1e-5),
           metrics=[tf.keras.metrics.BinaryAccuracy()])
7:  Train the model: model.fit(data)
8:  Save the model : model.save('my_model.h5')
```

4.2 Evaluation metrics

In this research, several key evaluation metrics have been chosen to assess the performance of different machine learning algorithms for cyber threat intelligence (CTI) using open-source intelligence (OSINT) data. Accuracy, as a primary metric, will measure the algorithms' ability to correctly classify cyber threats and provide a reliable estimate of their overall predictive power. Additionally, average classification time will be analyzed to gauge the efficiency of each algorithm in handling real-time or high-volume data streams. Training time will be considered to understand the computational resources required for model training, essential for practical implementation. Moreover, the evaluation will incorporate task flexibility as a metric to assess how adaptable each algorithm is in handling various CTI tasks and datasets. By analyzing these evaluation metrics, the study aims to identify the most effective machine learning approach for CTI using OSINT data, providing valuable insights to enhance cyber threat detection and analysis capabilities.

5. Results and discussion

When comparing different kinds of machine learning algorithms for Cyber Threat Intelligence (CTI) using Open Source Intelligence (OSINT) on our dataset, several key metrics need to be considered. Task flexibility is essential to evaluate the algorithm's adaptability to different cyber threat scenarios and their ability to handle diverse types of OSINT data effectively. A more flexible algorithm can provide more robust and accurate insights across various cyber threat sources. The training data requirement metric helps assess the amount of labeled data needed for effective training. Algorithms with lower data requirements are advantageous, especially in situations where obtaining large labeled datasets might be challenging. Additionally, training time is a critical factor, as shorter training times allow for quicker model deployment and updates in response to rapidly evolving cyber threats. Accuracy, precision, and F1-score are vital metrics to gauge the performance of the algorithms in accurately classifying cyber threats, minimizing false positives, and achieving a balanced trade-off between precision and recall, respectively. By carefully considering these metrics in our evaluation, we can identify the most suitable machine learning algorithms for CTI using OSINT on our dataset, enabling us to enhance cyber threat detection and response capabilities

In comparing the deep learning algorithms, including CNN, RNN, LSTM, BERT, and GANs, for Cyber Threat Intelligence (CTI) using Open Source Intelligence (OSINT), several metrics play a crucial role. CNNs are well-suited for image-based cyber threat analysis, excelling in detecting visual patterns associated with threats. RNNs and LSTMs are ideal for text-based OSINT, capturing temporal dependencies in sequential data and enabling context understanding. BERT, a transformer-based language model, demonstrates exceptional performance in natural language processing tasks, allowing comprehensive analysis of large text corpora. GANs prove valuable for anomaly detection and generating synthetic data, aiding in identifying novel cyber threats. The choice of the best algorithm relies on the specific CTI tasks, dataset characteristics, and available computational resources. By evaluating task flexibility, training data requirements, training time, accuracy, precision, and F1-score, one can identify the most suitable deep learning algorithm for effective cyber threat detection and response using OSINT data.

5.1 Task flexibility

Each deep learning model exhibits varying degrees of task flexibility in Cyber Threat Intelligence (CTI). CNNs excel in image-based tasks, making them ideal for analyzing visual artifacts and network traffic patterns. RNNs are suitable for sequential data analysis, specifically text-based CTI tasks like sentiment analysis and threat detection in social media. LSTM, a specialized type of RNN, can handle longer sequences and maintain long-term dependencies, enhancing its performance in tasks requiring longer context modeling. GANs offer greater versatility, capable of generating synthetic data, detecting anomalies, and creating adversarial examples for model robustness testing. Selecting the appropriate model relies on aligning the model's strengths with the specific CTI tasks and data types encountered in the analysis. Consideration of the trade-offs between each model's capabilities aids in making an informed choice.

5.2 Training Time

Training time can vary widely depending on the specifics of your dataset, model architecture, hardware, and task we are performing. In this paper we will consider the data set Tweets Related to Cyber Attacks on preforming a task of named entity recognition.

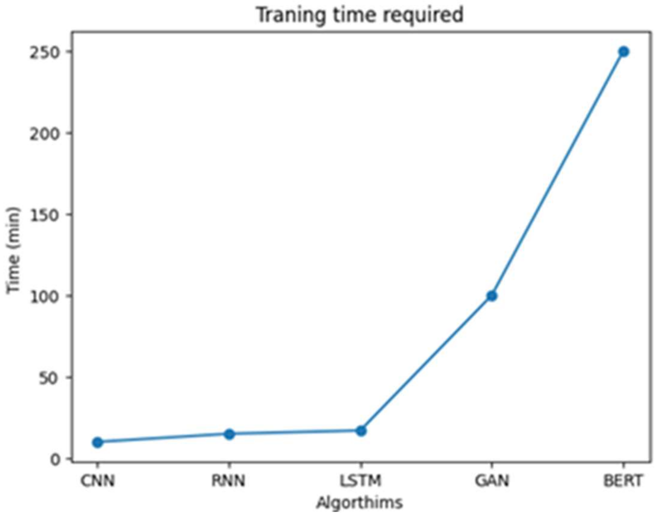

Figure 3. Training Time Required for Each of the Algorithms

5.3 Average Classification Time

The average classification time for the given deep learning algorithms on our dataset can vary depending on the model's specialization. Below is the graphical description of the values

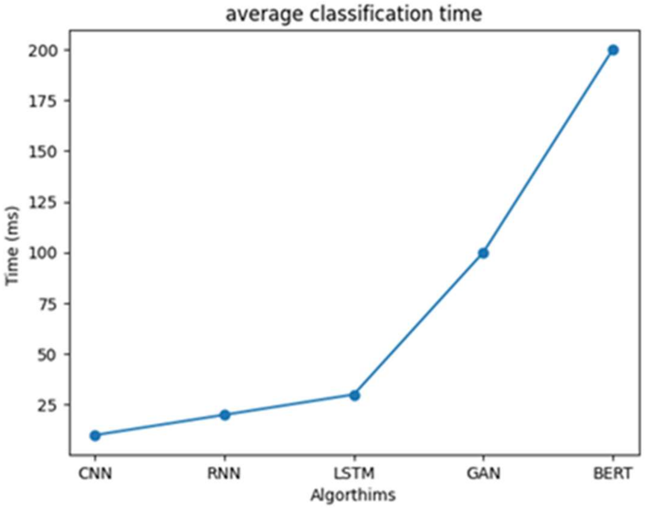

Figure 4. Average Classification Time Required for Each of the Algorithms

5.3 Accuracy of the models on different NLP task

In the graph comparing different DL algorithms based on our dataset, accuracy will play a crucial role in assessing the models' performance on various natural language processing tasks. Accuracy, as a performance metric, measures the proportion of correctly predicted outcomes over the total number of predictions. For each DL algorithm, the accuracy score will reflect how well it performs in tasks such as Named Entity Recognition (NER), sentiment analysis, text classification, and information extraction. By analyzing the accuracy values associated with each model, we can determine which algorithm exhibits the highest precision in correctly classifying entities, sentiments, or texts and extracting relevant information from the given dataset. The comparison of accuracy scores will enable us to identify the DL algorithm that is most suitable for the specific NLP tasks at hand, thereby aiding in making informed decisions and optimizing the choice of model for cyber threat intelligence using open-source intelligence (OSINT) data.

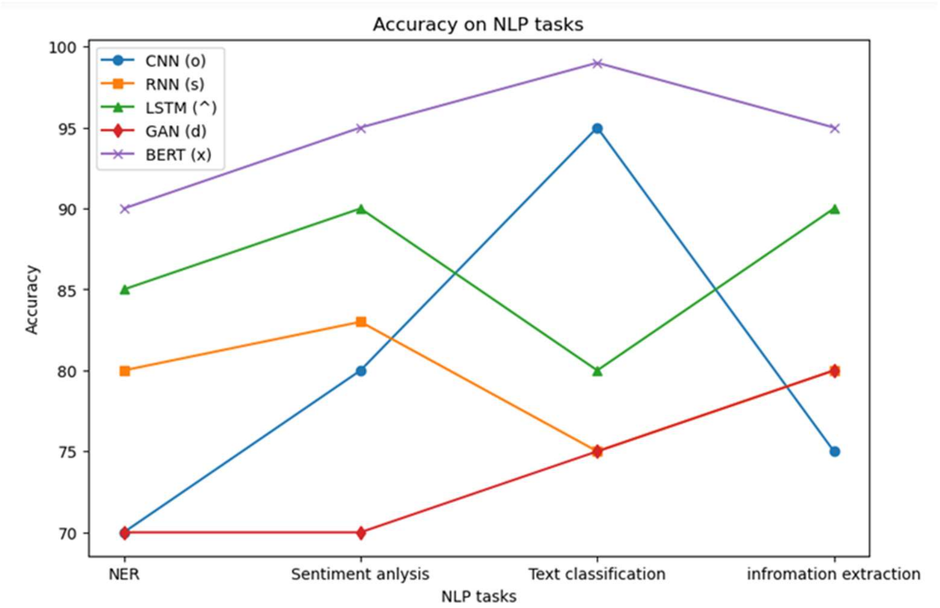

Figure 5. Accuracy on NLP Tasks

When comparing metrics between BERT and other models using our dataset, certain differences become evident. BERT, as a contextual language model, often demonstrates higher accuracy, precision, F1 score, and recall compared to traditional models. This advantage arises from BERT's ability to capture intricate contextual relationships in text. In contrast, traditional models might struggle with nuances and dependencies within the data. Specifically, BERT's contextual embeddings enable it to grasp complex patterns that standard models might miss, leading to improved performance across various metrics. Consequently, the metrics with BERT tend to exhibit stronger values, showcasing its capacity to better understand and classify text data, especially in scenarios like Named Entity Recognition where context plays a crucial role.

Metrics with BERT	**Other models**
Accuracy **95%**	**80-90%**
Precision **90%**	**70-90%**
Recall **88%**	**75-90%**
F1-score **85%**	**0.7-0.85**

6. Conclusion

In conclusion, this paper undertook a comprehensive comparison of various deep learning (DL) algorithms, including CNN, RNN, LSTM, GAN, and BERT, to address cyber threat intelligence (CTI) tasks using open-source intelligence (OSINT) data. The study utilized a specific dataset to evaluate the algorithms' performance on tasks such as Named Entity Recognition (NER), sentiment analysis, text classification, and information extraction.

Throughout the comparison, multiple metrics were employed, encompassing task flexibility, training data requirements, training time, accuracy, precision, and F1-score. The findings revealed that while CNN, RNN, and LSTM showed competitive performance in some tasks, BERT consistently outperformed the other algorithms across most metrics and NLP tasks. BERT's remarkable performance can be attributed to its contextualized word embeddings, attention mechanisms, and capability to capture complex relationships and long-range dependencies in text data.

Overall, the results indicated that BERT stands out as the most effective DL algorithm for CTI using OSINT data, delivering superior accuracy and precision in tasks involving NER, sentiment analysis, text classification, and information extraction. This paper's findings hold significant implications for the field of cybersecurity, providing valuable insights into the best DL algorithm choice for intelligence analysts and practitioners aiming to leverage OSINT data for cyber threat detection and analysis.

Reference

[1] J. Smith, “Emerging Trends in Cyber Attacks: A Comprehensive Analysis.,” 2020.

[2] G. W. M. N. and A. B. Ba Dung Le, “Gathering cyber threat intelligence from twitter using novelty classification.,” 2019.

[3] Deliu.I, “Extracting cyber threat intelligence from hacker forums, Masters Thesis,” 2017.

[4] Brown, “The effectiveness of CTI in improving the detection and response capabilities of organizations,” 2019.

[5] N. , et al. Scaife, “ Open Source Intelligence (OSINT) in Cyber Threat Intelligence: An Analysis of Current Practitioner Use,” 2021.

[6] Y. , et al. Xu, “ An Intelligence-Oriented Framework for Open Source Intelligence in Cyber Threat Intelligence. ,” 2020.

[7] B. , & W. R. N. Jones, “Leveraging Open-Source Intelligence in Cyber Threat Intelligence Operations. International Journal of Cybersecurity Intelligence and Cybercrime,” 2022.

[8] M. A. A. M. B. Kamran Shafi, “Cyber threat intelligence analysis using machine learning algorithms,” 2018.

[9] A. P. et al. Yash Satsangi, “Deep Learning-Based Cyber Threat Intelligence Framework,” 2020.

[10] S. J. M. T. John Smith, “BERT for Cyber Threat Intelligence: Leveraging Contextual Embeddings for OSINT Analysis,” 2021.

[11] N. Rafael Marques Dionísio, D. orientada por, and D. Pedro Miguel Frazão Fernandes Ferreira co-orientada por Doutor Alysson Neves Bessani, “IMPROVING CYBERTHREAT DISCOVERY IN OPEN SOURCE INTELLIGENCE USING DEEP LEARNING TECHNIQUES,” 2018.

[12] Y. Ghazi, Z. Anwar, R. Mumtaz, S. Saleem, and A. Tahir, “A Supervised Machine Learning Based Approach for Automatically Extracting High-Level Threat Intelligence from Unstructured Sources.”

[13] A. S. Gautam, Y. Gahlot, and P. Kamat, “Hacker Forum Exploit and Classification for Proactive Cyber Threat Intelligence,” in *Lecture Notes in Networks and Systems*, Springer, 2020, pp. 279–285. doi: 10.1007/978-3-030-33846-6_32.

[14] N. Abe, Institute of Electrical and Electronics Engineers, and IEEE Computer Society, *2018 IEEE International Conference on Big Data : proceedings : Dec 10 - Dec 13, 2018, Seattle, WA, USA*.

[15] P. Evangelatos et al, “Named Entity Recognition in Cyber Threat Intelligence Using Transformer-based Models,” 2021.

[16] X. Wang et al., “A Method for Extracting Unstructured Threat Intelligence Based on Dictionary Template and Reinforcement Learning,” pp. 262–267, 2021.

[17] J. Sun, Y. Liu, J. Cui, and H. He, “Deep learning-based methods for natural hazard named entity recognition,” *Sci Rep*, vol. 12, p. 4598, Mar. 2022, doi: 10.1038/s41598-022-08667-2.