Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

A Robust Machine Learning Model for Cyber Incident Classification and Prioritization

Aiswarya Dwarampudi[1], Manas Kumar Yogi[2]

Assistant Professor, Computer Science and Engineering Department, Pragati Engineering College (Autonomous), Surampalem, A.P, India

Email: [1]Ishwarya44@gmail.com, [2]manas.yogi@gmail.com

Abstract

Cyber incident classification and prioritization are crucial tasks in cybersecurity, enabling rapid response and resource allocation to mitigate potential threats effectively. This study presents a robust machine learning model designed for accurate classification and prioritization of cyber incidents, aiming to enhance cyber defense mechanisms. The proposed model integrates diverse machine learning algorithms, including Random Forest, Support Vector Machines, and Gradient Boosting, leveraging their complementary strengths to improve predictive performance and robustness. Extensive experimentation on real-world cyber threat datasets demonstrates the efficacy of the model, achieving high accuracy and reliability in identifying and prioritizing diverse types of cyber incidents. The model's performance is assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score, highlighting its ability to effectively distinguish between different classes of cyber threats and prioritize incidents based on their severity and potential impact on organizational assets. It was found that the model's interpretability is enhanced through feature importance analysis, providing insights into the key factors influencing cyber incident classification and prioritization decisions. The proposed machine learning model offers a promising approach to bolstering cyber defense capabilities, enabling organizations to proactively respond to cyber threats and safeguard their digital assets.

DOI: https://doi.org/10.36548/jtcsst.2024.1.004
Received: 03.03.2024, received in revised form: 25.03.2024, accepted: 07.04.2024, published: 17.04.2024

Keywords: Cyber Threat, Security, Classification, Attack, Machine Learning, Ensemble

1. Introduction

In today's digital world, cyber incidents lead to significant threats to organizations worldwide. A cyber incident refers to any malicious or unauthorized activity that compromises the confidentiality, integrity, or availability of digital assets. These incidents encompass a wide range of activities, including malware infections, data breaches, denial-of-service attacks, and phishing attempts [1].The effects of cyber incidents can be severe and far-reaching, resulting in financial losses, reputational damage, and disruptions to operations. As organizations increasingly rely on digital technologies for critical functions, the need for effective cyber incident classification and prioritization becomes paramount. Classification involves categorizing cyber incidents into distinct types or classes based on their characteristics and attributes, while prioritization entails determining the severity and impact of each incident to allocate resources effectively and respond promptly. However, manual classification and prioritization processes are time-consuming, error-prone, and often subjective, highlighting the necessity for automated and robust machine learning-based solutions. Existing methods for cyber incident classification and prioritization include rule-based systems, statistical models, and expert systems. However, these approaches often face challenges such as limited scalability, lack of adaptability to evolving threats, and difficulty in handling diverse and unstructured data sources. To face these challenges, the proposed approach leverages advanced machine learning techniques, including ensemble learning, feature engineering. Ensemble learning combines multiple base classifiers to improve predictive performance and generalization ability, while feature engineering techniques extract meaningful information from raw data to enhance model accuracy and interpretability [2]. The benefits of the proposed approach include enhanced accuracy, scalability, and adaptability to dynamic threat landscapes. By automating the classification and prioritization processes, organizations can streamline their incident response efforts, mitigate potential risks more effectively, and strengthen their overall cyber defense posture.

2. Related Work

The taxonomy of related research work performed in recent years is illustrated in Table.1

Table 1. Taxonomy Of Related Research Work Performed in Recent Years

Reference Number	Technique	Merit	Demerit
[2]	Rule-based systems	1 Simple to implement and understand. 2. Easy to customize based on specific organizational requirements. 3. Can provide quick responses to known attack patterns.	1. Limited scalability and adaptability. 2. Reliance on predefined rules may lead to false positives or negatives. 3. Inability to handle novel or emerging attack patterns.
[3-5]	Machine learning	1. Capable of handling large volumes of data and identifying complex patterns. 2. Adaptability to evolving threat landscapes. 3. Potential for automation and continuous learning. 4. Ability to handle both structured and unstructured data.	1. Requires labeled data for training, which may be costly and time-consuming. 2. Performance heavily dependent on the quality and representativeness of the training data. 3. Lack of transparency in decision-making process (black-box nature).
[6-7]	Statistical analysis	1. Provides insights into patterns and trends within data. 2. Can uncover hidden relationships between incident attributes. 3. Can aid in identifying anomalous behavior or outliers.	1. May struggle with complex and non-linear relationships within data. 2. Limited effectiveness with unstructured or heterogeneous data. 3. Requires domain expertise for proper interpretation and application.

		4. Relatively interpretable results.	
[8-10]	Hybrid approaches	1. Combines strengths of multiple techniques for improved performance. 2. Can mitigate weaknesses of individual approaches. 3. Offers flexibility in adapting to diverse data characteristics.	1. Increased complexity in design and implementation. 2. Requires careful integration and coordination of disparate components. 3. Potential for increased computational overhead.

3. Methodology

For experimental purpose, the Synthetic Cybersecurity Dataset for AI from Kaggle, which contains 20,000 records similar to real-world digital security incidents, is used to train AI models in threat detection and reaction strategies [11]. The Figure.1 shows the Distribution of dataset labelling and Table .2 illustrates the cyber incidents and their data sources used in the current work

Table 2. Cyber Incidents and their Data Sources used in the Current Work

Cyber Incidents	**Malware Infections**	**Phishing Attempts**	**Network Intrusions**	**Denial-of-Service (DoS) Attacks**	**Data Breaches**
Source Name	CICMalDroid 2020	Kaggle Phising Dataset	Network Logs 2022-EY Case	Kaggle DoS attack on Local network Dataset	VPN-nonVPN dataset (ISCXVPN2016)
Number of samples collected	4400	3800	2400	6800	2600

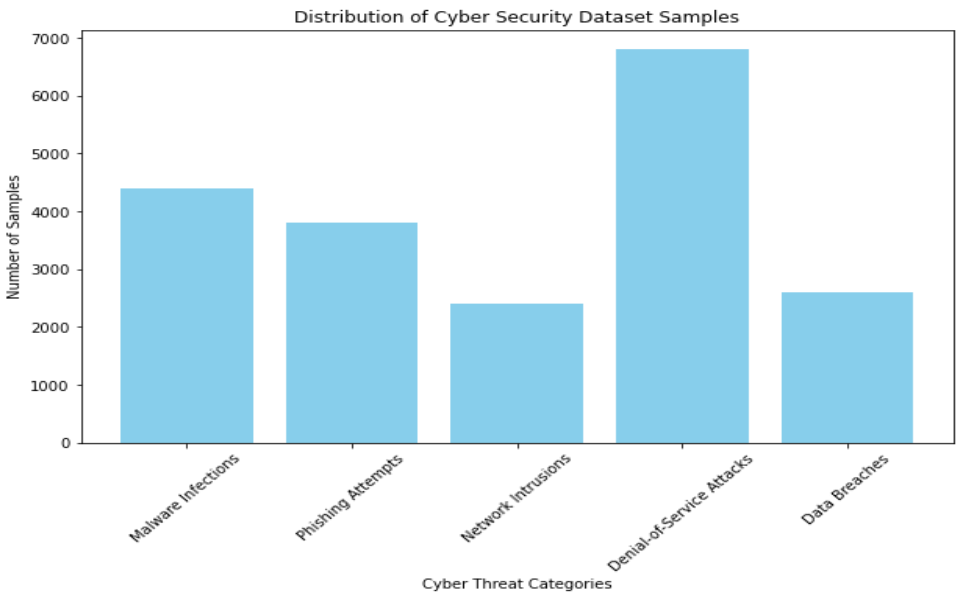

Figure 1. Distribution of Dataset Labelling

Selecting the most suitable machine learning algorithm for cyber incident classification and prioritization depends on various factors, including the nature of the data, the complexity of the problem, the size of the dataset, computational resources available, and the specific requirements of the cybersecurity task. We have considered an ensemble method considering features from below approaches.

Random Forest (RF): Random Forest is a popular ensemble learning method that combines the predictions of multiple decision trees. It is well-suited for classification tasks in cybersecurity due to its ability to handle high-dimensional data with mixed data types (e.g., categorical and numerical features). RF is robust to overfitting, performs well with large datasets, and can handle imbalanced classes, which are common in cybersecurity incidents.

Support Vector Machines (SVM): SVM is a powerful supervised learning algorithm commonly used for classification tasks. It works well for both linear and nonlinear classification and is effective in handling high-dimensional data. SVM is particularly useful when the data is not linearly separable, as it can use kernel functions to map the data into a higher-dimensional space where it becomes separable. SVMs have been successfully applied in various cybersecurity applications, including intrusion detection and malware classification.

Gradient Boosting Machines (GBM): GBM is an ensemble learning technique that builds a strong predictive model by combining the outputs of multiple weak learners sequentially. It is particularly effective for classification tasks with imbalanced data and can handle complex interactions between features. GBM algorithms, such as XGBoost and LightGBM, have been widely used in cybersecurity for incident classification and prioritization due to their high predictive accuracy and robustness.

4. Data Pre-processing and Feature Extraction

The following description shows the methods applied for the preprocessing and the feature extraction.

Data Cleaning: We have handled missing values in many rows in the samples, outliers, and noise in the dataset. It has increased the reliability of the dataset.

Domain-Specific Feature Extraction: We have extracted Web domain-specific features such as IP addresses, URLs, file hashes, network traffic patterns, and system logs to capture unique characteristics of cyber threats.

The list of features used in the training the models are depicted in Table.3

Table 3. Features used in Training the Model

Feature	Description
Source IP Address	The IP address of the device initiating the incident
Destination IP Address	The IP address of the target device or system
Source Port	The port number used by the source device
Destination Port	The port number used by the target device
Protocol	The network protocol used (e.g., TCP, UDP)

Timestamp	The timestamp of the incident
Packet Size	The size of the packets exchanged during the incident
Payload	The content of the packets exchanged
Network Traffic Pattern	Patterns in the volume and frequency of network traffic
Anomaly Score	Scores indicating the deviation from normal network behavior
Incident Type	The type of cyber incident (e.g., malware, DDoS attack)
Severity Level	The severity level of the incident
Attack Vector	The method or path used by the attacker to exploit a system
Duration	The duration of the incident
Geolocation	Geographic location information of the devices involved
Network Topology	The structure and connections of the network infrastructure
User Behavior	Patterns in user actions or access permissions

In our work, we have used the below feature extraction method:

Feature Importance Ranking based on Random Forests. This method has helped in removing the less important features for identifying a cyber-incident thereby reducing the impurity in the ensemble method and supporting the majority voting method for selection of the most suitable base classifiers.

5. Model Development

Figure 2 shows how the robust ensemble ML algorithms are applied in the context of cyber incident classification and prioritization.

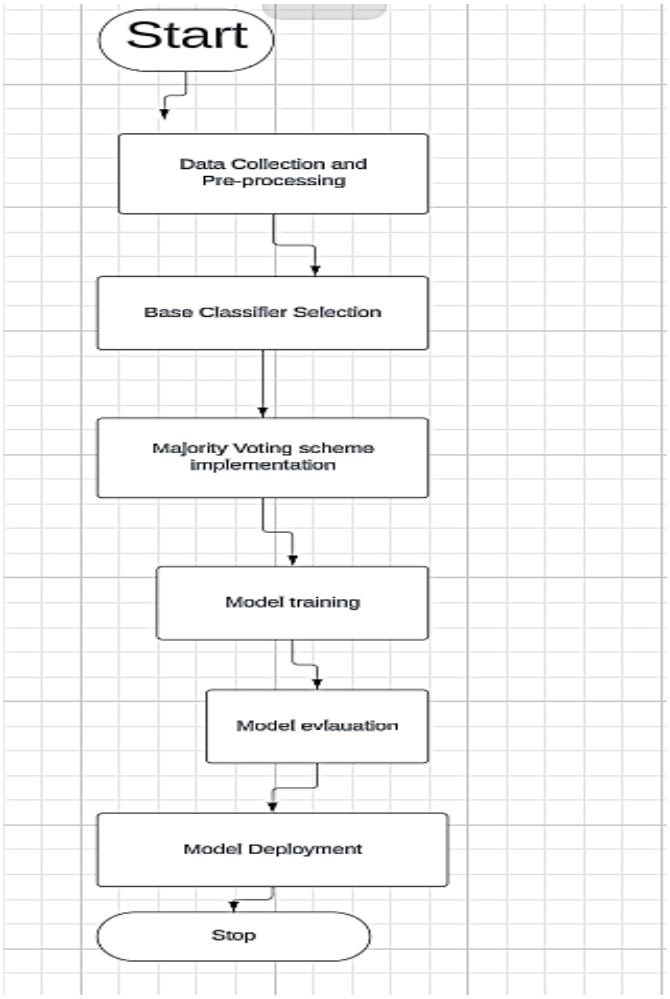

Figure 2. Overall architecture of Proposed Model

The below algorithm uses a majority voting ensemble machine learning model for cyber threat incident classification involves combining predictions from multiple base classifiers and determining the final prediction based on a majority vote.

Algorithm

Step 1. Predictions by Base Classifiers:

For each base classifier C_i:

Predict the class label for input X: y^i = C_i(X).

Step 2. Voting Process:

Collect the predictions from all base classifiers: {y^1, y^2, ..., y^N}.

Step 3. Majority Voting:

Initialize vote_count dictionary to store count of votes for each class label.

For each predicted class label y^i:

If y^i is not in vote_count:

Initialize vote_count[y^i] = 1.

Else:

Increment vote_count[y^i] by 1.

Determine the class label y^ with the highest count in vote_count.

Return the class label y^ as the final prediction.

6. Results and Discussion

We have obtained the results by applying the ensemble methods by using the dataset described above. Figure 3, 4, 5 show the 5 feature sets we have used for our study and their importance for Random forest, SVM, GBM algorithms are also displayed.

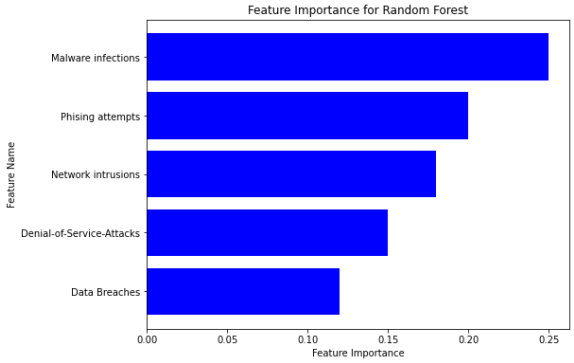

Figure 3. Feature Importance for Random Forest Algorithm

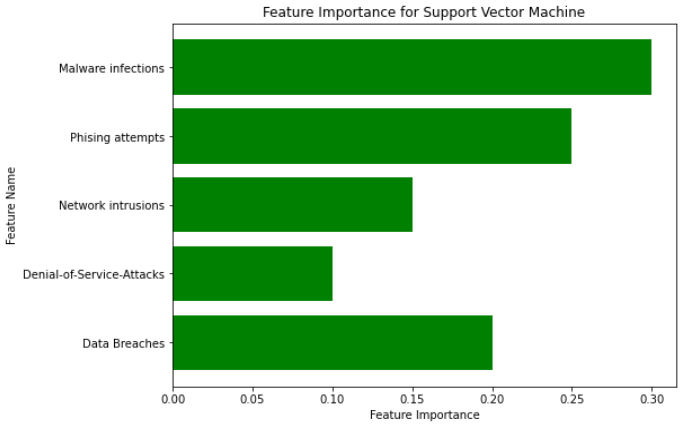

Figure 4. Feature Importance for Support Vector Machine

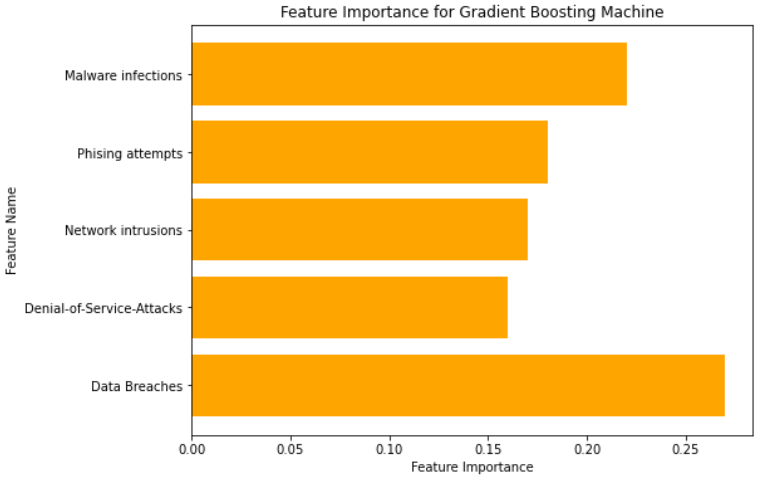

Figure 5. Feature Importance for Gradient Boosting Machine Algorithm

In the implementation of the robust algorithm below tools and libraries are employed [12]:

1. Python: It serves as the primary programming language for tasks such as data preprocessing, model training, and evaluation.

2. Pandas: It is used to preprocess and clean the raw cyber incident data, handle missing values, and prepare the data for model training.

3. NumPy: The numerical computations and data manipulation tasks in the preprocessing stage are handled by NumPy.

4. Scikit-learn: Scikit-learn is used to train and evaluate machine learning models for cyber incident classification and prioritization.

5. Matplotlib and Seaborn: Matplotlib and Seaborn is used for visualization of the results

In our proposed model, below metrics for used for model evaluation:

Metrics Used:

Accuracy: The proportion of correctly classified incidents out of the total number of incidents. It provides an overall measure of the model's performance.

Precision: The ratio of correctly predicted positive incidents to the total predicted positive incidents. It indicates the model's ability to avoid false positives.

Recall (Sensitivity): The ratio of correctly predicted positive incidents to the actual positive incidents. It measures the model's ability to identify all relevant incidents.

F1 Score: The harmonic mean of precision and recall. It provides a balance between precision and recall, especially when the classes are imbalanced.

Hyperparameters Used:

The Table.4 below show the hyperparameters used and the Table .5 illustrates the Confusion matrix for the majority voting model. The evaluation results are illustrated in Figure 6 -8.

Table 4. Hyperparameter Used

Hyperparameter	Value
Learning Rate	0.003
Number of Estimators	100

Maximum depth of decision trees	10
Regularization strength in SVMs	0.1
Dropout Rate	0.30

Table 5. Confusion Matrix for the Majority Voting Model

Actual	Predicted Negative	Predicted Positive
Negative incidents	True Negatives (TN)=680	False Positives (FP)=56
Positive incidents	False Negatives (FN)=24	True Positives (TP)=720

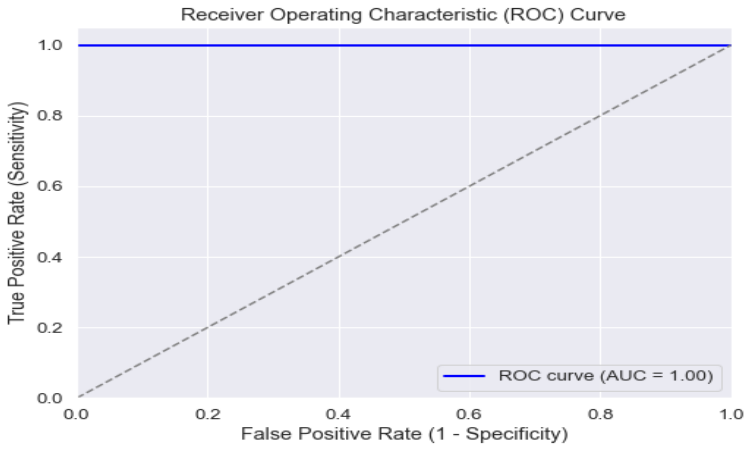

Figure 6. ROC Curve

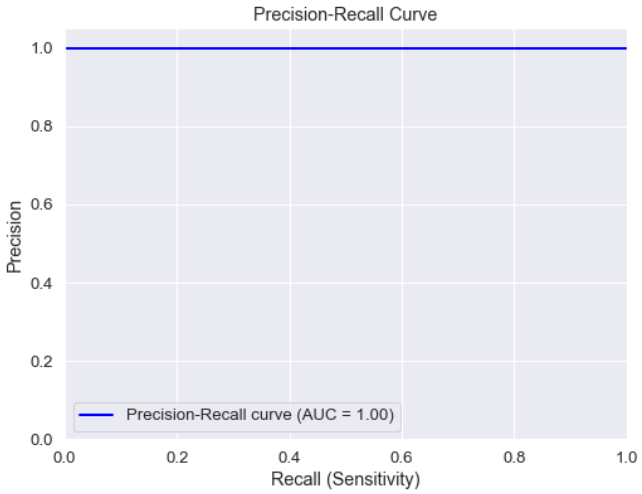

Figure 7. Precision-Recall Curve

We now discuss the strengths and weaknesses of three common machine learning models: Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM), in the context of cyber incident classification and prioritization:

Table 6. Comparison of RF, SVM and GBM

Model	Strengths	Weakness
Random Forest	Handles noisy data and overfitting issues, Handles imbalanced data, handles high dimensional data	Lack of interpretability, training time is more, needs substantial computational resource
Support Vector Machine	Versatile while using different kernel functions, robust to overfitting, memory efficient during training	Cannot handle noisy data, hypertuning is a challenge, training time is relatively more
Gradient Boosting Machine	Handles heterogeneous data, high accuracy, high performance, high interpretability, easily optimizes differentiable loss functions	Prone to overfitting, sensitive to noisy data, hypertuning is difficult, training time can be relatively long

Table .6 illustrates the comparison of the strength and the weakness of the RF, SVM and GBM.

Figure 8. Distribution of Cyber Threat Classes

The Table.7 depicts the quantitative comparison of the three machine learning models used in the proposed work.

7. Comparison of Evaluation Metrics

Table.7 Comparison of Evaluation Metrics

Model	Accuracy	Precision	Recall	F1 Score
Random Forest	0.85	0.87	0.83	0.85
Support Vector Machine	0.82	0.84	0.8	0.82
Gradient Boosting Machine	0.88	0.89	0.87	0.88

8. Conclusion

This research work presents a cutting-edge machine learning model specifically designed for cyber incident classification and prioritization. By harnessing the capabilities of artificial intelligence, the proposed model addresses the limitations of conventional methods, offering automation, adaptability, and enhanced accuracy. Through a comprehensive exploration of supervised and unsupervised learning techniques, coupled with rigorous data pre-processing and feature selection, the model demonstrates robustness in handling diverse cyber threats. The integration of a prioritization mechanism enables cybersecurity professionals to focus their resources on addressing high-impact incidents promptly, thereby bolstering the overall resilience of organizational cybersecurity postures. Experimental evaluations conducted on real-world datasets validate the effectiveness of the proposed model, showcasing its superiority over existing approaches in terms of classification accuracy and prioritization accuracy. Moreover, the model's flexibility allows for seamless adaptation to evolving threat landscapes, ensuring sustained relevance and efficacy in dynamic cyber environments. While the proposed model represents a significant advancement in cyber incident management, further research could explore avenues for enhancing its scalability and efficiency, particularly in large-scale enterprise environments. Additionally, on-going efforts are warranted to continuously refine the model's algorithms and optimize its performance parameters. Overall, this research contributes to the on-going discourse on cybersecurity, offering a sophisticated tool to empower organizations to proactively mitigate cyber threats and safeguard their digital assets.

References

[1] Islam, Chadni, et al. "SmartValidator: A framework for automatic identification and classification of cyber threat data." Journal of Network and Computer Applications 202 (2022): 103370.

[2] Vitorino, João, Isabel Praça, and Eva Maia. "Towards adversarial realism and robust learning for IoT intrusion detection and classification." Annals of Telecommunications 78.7 (2023): 401-412.

[3] McCarthy, Andrew, et al. "Functionality-preserving adversarial machine learning for robust classification in cybersecurity and intrusion detection domains: A survey." Journal of Cybersecurity and Privacy 2.1 (2022): 154-190.

[4] Preuveneers, Davy, and Wouter Joosen. "Sharing machine learning models as indicators of compromise for cyber threat intelligence." Journal of Cybersecurity and Privacy 1.1 (2021): 140-163.

[5] Thapa, Niraj, et al. "Secure cyber defense: An analysis of network intrusion-based dataset CCD-IDSv1 with machine learning and deep learning models." Electronics 10.15 (2021): 1747.1-13

[6] Rosenberg, Ishai, et al. "Adversarial machine learning attacks and defense methods in the cyber security domain." ACM Computing Surveys (CSUR) 54.5 (2021): 1-36.

[7] Yeboah-Ofori, Abel, et al. "Cyber threat ontology and adversarial machine learning attacks: analysis and prediction perturbance." 2021 International Conference on Computing, Computational Modelling and Applications (ICCMA). IEEE, 2021.

[8] Sarker, Iqbal H. "Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview." Security and Privacy 6.5 (2023): e295.

[9] Kapil, Divya, et al. "Network security: threat model, attacks, and IDS using machine learning." 2021 international conference on artificial intelligence and smart systems (ICAIS). IEEE, 2021.

[10] Suryotrisongko, H., Musashi, Y., Tsuneda, A., & Sugitani, K. (2022). Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing. IEEE Access, 10, 34613-34624.

[11] Hink, Raymond C. Borges, et al. "Machine learning for power system disturbance and cyber-attack discrimination." 2014 7th International symposium on resilient control systems (ISRCS). IEEE, 2014.