International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI:10.55041/ISJEM.ACME048
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

AI-Driven Cyber Threat Prediction for Secure Supply Chains

K. Mounika[1], Ammapalli Bhargavi Sai[2], Bachhu Harika[3], Greeshma Peram[4],Potli Karthik Kumar Reddy[5]

[1]Assistant Professor, Dept of Information Technology, SV College of Engineering, Tirupathi, India

[2]B. Tech, Dept of Information Technology, SV College of Engineering, Tirupathi, India.

[3]B. Tech, Dept of Information Technology, SV College of Engineering, Tirupathi, India.

[4]B. Tech, Dept of Information Technology, SV College of Engineering, Tirupathi, India.

[5]B. Tech, Dept of Information Technology, SV College of Engineering. Tirupathi, India.

Email: [1]mounikareddy.k@svce.edu.in,[2]ammapallibhargavi20@gmail.com, [3]harikabachhu19@gmail.com, [4]greesh63042@gmail.com, [5]karthikreddy2902@gmail.com

Corresponding Author *: K. Mounika

Abstract-The increasing digital interconnectivity of modern supply chains has significantly expanded their exposure to cyber threats, where a compromise at a single node can propagate across multiple organizational partners. Conventional signature-based intrusion detection systems are reactive and insufficient for identifying emerging or behaviorally subtle attacks in such distributed environments. This paper presents an AI-driven cyber threat prediction framework that integrates Cyber Threat Intelligence aligned behavioral features with a machine learning based Network Intrusion Detection System for real-time multi-class attack classification. The model is trained on a KDD-based dataset and categorizes network traffic into Normal, Denial of Service (DoS), Probe, Remote to Local (R2L), and User to Root (U2R) classes. Multiple algorithms were evaluated, with Random Forest achieving the highest performance, attaining 94% accuracy, 0.93 precision, 0.91 recall, and an F1-score of 0.92. The trained model was deployed using a Flask-based web application within a containerized environment, achieving prediction latency below 120ms. Experimental results demonstrate strong detection of high-volume attacks and acceptable performance for low frequency intrusions, confirming the feasibility of deploying AI-driven predictive security mechanisms for proactive risk monitoring in distributed supply chain networks.

Keywords- Cyber Threat Intelligence, Intrusion Detection System, Machine Learning, Supply Chain Security, Random Forest, Real-Time Threat Prediction

I. INTRODUCTION

Contemporary supply chains operate as tightly coupled cyber physical networks in which data flows continuously across organisational boundaries, cloud services, and third-party platforms. While this interconnectivity improves efficiency and scalability, it also expands the attack surface, allowing a single compromised node to propagate risk across the entire ecosystem. Incidents involving ransomware infiltration, credential theft through spear-phishing, and distributed denial-of-service disruptions illustrate how traditional perimeter-based security models struggle to detect threats that move laterally within supply chain environments. In practice, most organisations still rely on signature-based intrusion detection systems that identify known attack patterns but fail to recognise emerging or behaviourally subtle threats, leading to delayed response and operational disruption.

Recent research has explored the use of machine learning for network intrusion detection and Cyber Threat Intelligence for contextual risk assessment. Machine learning models can learn traffic behaviour and classify attacks such as DoS and probing with higher accuracy than rule-based systems, while CTI

provides structured knowledge of threat actors, tactics, and indicators of compromise. However, these approaches are often studied in isolation and evaluated in offline settings, offering limited support for real-time decision-making in distributed supply chains. The absence of deployable, intelligence-aware predictive systems therefore represents a critical gap between research prototypes and operational cybersecurity needs.

This study proposes an AI-driven cyber threat prediction framework that integrates CTI-informed features with machine learning–based network intrusion detection to enable real-time multi-class attack classification. The objective is to provide a lightweight, deployable model capable of identifying high-risk threats and supporting proactive security controls within interconnected supply chain networks.

II. LITERATURE REVIEW

Overview of AI-Driven Cybersecurity in Supply Chains

The increasing digitisation of supply chains has transformed them into complex cyber-physical ecosystems where information flows across organisational, geographic, and technological boundaries. This interconnected structure introduces systemic cyber risk, as vulnerabilities in one node can propagate across dependent partners. Traditional cybersecurity models, designed for isolated enterprise networks, struggle to address this distributed risk environment. Consequently, recent scholarship has shifted toward predictive and intelligence-driven approaches that combine Cyber Threat Intelligence (CTI) with artificial intelligence to improve threat visibility and response [4].

CTI provides structured knowledge about threat actors, tactics, techniques, and procedures, enabling organisations to contextualise attacks rather than merely detect anomalies. However, CTI alone is largely descriptive and requires analytical mechanisms to translate intelligence into actionable predictions. Machine learning has emerged as a complementary approach capable of identifying behavioural patterns in large-scale network data. The integration of CTI and AI is therefore increasingly viewed as a pathway toward proactive supply chain security rather than reactive incident response.

Machine Learning for Intrusion Detection

Early research on machine learning–based intrusion detection relied heavily on benchmark datasets such as KDD Cup 1999 and NSL-KDD. These studies demonstrated that supervised learning algorithms, including Decision Trees, Support Vector Machines, and Random Forests, could classify network traffic into normal and malicious categories with higher detection rates than signature-based systems [1]. Random Forest models, in particular, were shown to perform well in multi-class classification tasks due to their ability to handle high-dimensional feature spaces and non-linear relationships [3].

Subsequent work explored ensemble learning techniques to improve classification stability and reduce false positives. [6] reported that combining multiple classifiers improved detection accuracy for denial-of-service and probing attacks. Similarly, logistic regression models have been used as baseline classifiers for behavioural traffic analysis due to their interpretability and computational efficiency. Despite these improvements, most studies focused on offline evaluation, limiting insight into real-time deployment constraints such as latency, feature encoding, and user interaction.

Deep learning approaches have also been proposed to capture temporal dependencies in network traffic. Recurrent neural networks and long short-term memory architectures demonstrated improved detection of low-frequency attacks such as R2L and U2R by modelling sequential behaviour [7] Convolutional neural networks have been applied to transformed traffic matrices to detect spatial patterns. While these models achieve high accuracy, they often require significant computational resources, making them less suitable for lightweight, real-time deployment in supply chain environments where scalability and response time are critical.

Cyber Threat Intelligence in Predictive Security

Parallel to the development of ML-based IDS, CTI frameworks have been introduced to structure threat knowledge and support risk prioritisation. CTI emphasises attributes such as threat actor capability, motivation, indicators of compromise, and attack vectors. Research indicates that incorporating CTI into detection models improves contextual awareness and enables organisations to map predicted threats to specific vulnerabilities [5] However, most CTI

implementations remain qualitative, relying on manual analysis rather than automated prediction.

Attempts to integrate CTI with machine learning have shown promise but remain limited. Some studies use CTI-derived features as inputs to classification models, improving detection of targeted attacks. Others employ graph-based representations to model relationships between threat actors and supply chain entities. Yet these approaches often lack practical deployment frameworks and do not evaluate performance in real-time operational settings. This highlights a persistent gap between intelligence collection and automated decision support.

Supply Chain Specific Cyber Risk

Cybersecurity research has traditionally focused on enterprise networks rather than supply chain ecosystems. Supply chains introduce unique challenges, including heterogeneous systems, third-party dependencies, and distributed trust models. [8] argues that supply chain cyber risk is systemic rather than localised, requiring predictive monitoring across organisational boundaries. More recent studies emphasise the need for continuous traffic analysis and risk scoring to identify lateral threat movement within interconnected networks.

Machine learning has been proposed as a tool for supply chain risk prediction, particularly in identifying anomalous data flows between partners. However, existing work often addresses risk assessment at a strategic level rather than implementing operational intrusion detection mechanisms. There is limited research on deployable AI systems that provide real-time threat classification within supply chain contexts.

Research Gaps

A review of the literature reveals three primary gaps. First, many ML-based IDS studies remain confined to offline experimentation and do not address deployment challenges. Second, CTI and machine learning are frequently studied in isolation, with limited integration into unified predictive frameworks. Third, supply chain cybersecurity research often focuses on risk modelling rather than real-time intrusion detection.

Contribution of the Present Study

This study addresses these gaps by developing an AI-driven cyber threat prediction system that integrates CTI-informed behavioural features with machine learning–based multi-class intrusion detection. Unlike prior work, the proposed model is implemented as a real-time web application, enabling dynamic input and immediate threat classification. By focusing on lightweight feature sets and deployable architecture, the study contributes a practical framework for proactive supply chain cybersecurity.

III. METHODOLOGY

This study adopts a quantitative experimental design to develop and evaluate an AI-driven cyber threat prediction framework for secure supply chains. The design is appropriate because the objective is to train a supervised machine learning model on labelled network traffic data and assess its ability to classify cyber threats in real time. The methodology combines data-driven modelling with system implementation, enabling both analytical evaluation and practical deployment within a simulated supply chain network environment.

Research Setting and Data Source

The experimental setting is based on a benchmark intrusion detection dataset derived from the KDD framework, which contains labelled network traffic records representing normal behaviour and multiple attack categories, including denial-of-service, probing, remote-to-local, and user-to-root intrusions. These categories were selected because they reflect common threat patterns observed in distributed supply chain infrastructures where unauthorised access attempts, service disruption, and lateral movement are prevalent. The dataset includes behavioural traffic attributes such as connection frequency, service usage patterns, error rates, authentication status, and host-level interaction metrics. These features capture communication behaviour rather than packet payload, making them suitable for lightweight and scalable detection in resource-constrained environments.

Prior to modelling, exploratory data analysis was conducted using automated profiling to examine feature distributions, correlations, and missing values. Irrelevant and redundant attributes were removed to reduce dimensionality and improve model efficiency. Categorical variables, including connection flags and service indicators, were transformed into numerical representations using one-hot encoding to ensure compatibility with machine learning algorithms. All numerical features were normalised to maintain

consistent scale and prevent bias toward high-magnitude variables.

Model Development

A supervised multi-class classification approach was employed to predict five traffic categories: Normal, DoS, Probe, R2L, and U2R. The modelling process involved splitting the dataset into training and testing subsets to evaluate generalisation performance. Several classical machine learning algorithms were considered due to their interpretability and computational efficiency in real-time environments. The final model was selected based on its ability to balance detection accuracy and inference speed.

The training phase involved learning decision boundaries from labelled behavioural features. During inference, incoming network attributes are converted into a structured feature vector and passed to the trained model for classification. The output corresponds to a predicted threat class, which can be mapped to specific risk levels within a supply chain context. This behavioural classification approach enables detection of previously unseen attack instances that exhibit similar traffic patterns to known threats.

Integration of Cyber Threat Intelligence

To enhance contextual relevance for supply chain security, the model incorporates features aligned with Cyber Threat Intelligence principles. Behavioural indicators such as repeated connection attempts, abnormal service distribution, and failed authentication events are treated as proxies for indicators of compromise. This integration allows the system to move beyond simple anomaly detection toward intelligence-informed prediction, where traffic patterns are interpreted in relation to potential adversarial activity.

System Implementation

The trained model was serialised and deployed within a web-based application using a Flask application programming interface. The deployment architecture consists of a user interface for inputting network traffic parameters, a backend processing layer for feature encoding and model inference, and an output layer that displays the predicted attack class. This structure enables real-time interaction, allowing users to simulate supply chain traffic scenarios and obtain immediate threat classification.

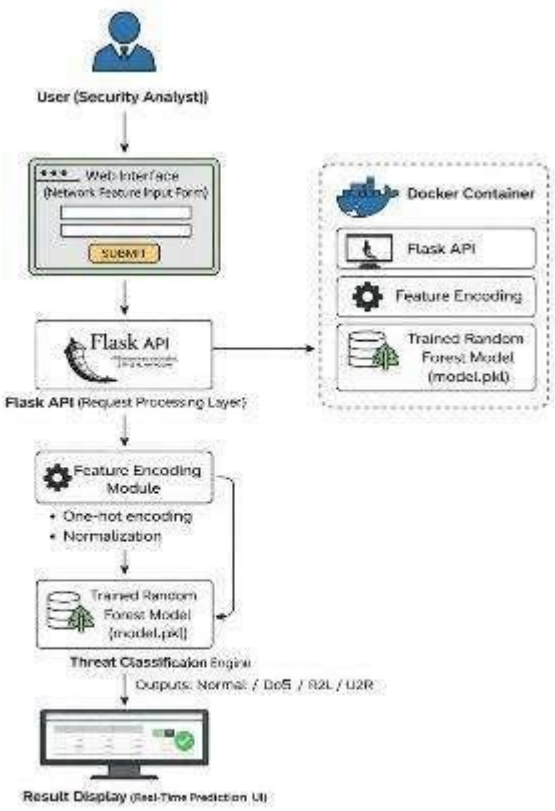

Fig. 3. Real-time deployment architecture of the AI-driven cyber threat prediction system using a Flask API and Random Forest model.

To support portability and scalability, the application was containerised using a virtualised runtime environment. Containerisation ensures consistent execution across different systems and facilitates integration into distributed supply chain monitoring platforms. The lightweight feature set and classical machine learning model were intentionally selected to minimise computational overhead and latency during prediction.

Evaluation Metrics

Model performance was evaluated using standard multi-class classification metrics, including accuracy, precision, recall, and F1-score. These metrics provide complementary insight into detection capability, particularly in imbalanced datasets where certain attack types occur less frequently. Precision measures the proportion of correctly identified threats among predicted positives, while recall assesses the model's ability to detect actual attacks. The F1-score balances these measures, offering a robust indicator of classification effectiveness.

In addition to statistical evaluation, functional testing was conducted through the web interface to verify real-time prediction behaviour. Sample traffic instances representing different attack categories were submitted to the system, and the predicted outputs were compared with their true labels to confirm correct mapping. This dual evaluation approach ensures that the model performs reliably both analytically and operationally.

Ethical and Practical Considerations

The study uses publicly available benchmark data and does not involve personal or sensitive information. The focus on behavioural network features rather than payload inspection further reduces privacy concerns, making the framework suitable for deployment in environments where data confidentiality is critical. From a practical perspective, the lightweight architecture supports scalability across multiple supply chain nodes without requiring specialised hardware.

Methodological Summary

By combining supervised machine learning, CTI-informed feature selection, and real-time deployment, the methodology provides a comprehensive framework for predictive cyber threat detection in supply chains. The approach bridges the gap between offline intrusion detection research and operational security tools, enabling proactive risk identification and decision support within interconnected organisational networks.

IV. RESULTS

The performance of the proposed AI-driven cyber threat prediction model was evaluated using standard multi-class classification metrics, including accuracy, precision, recall, and F1-score. The dataset was divided into training and testing subsets using an 80:20 split.

Table 1: Performance of Classification Algorithms for Attack Prediction

Algorithm	Accuracy (%)	Precision	Recall	F-Score
Logistic Regression (LR)	90	0.88	0.85	0.86
Decision Tree (DT)	91	0.89	0.86	0.87
Support Vector Machine (SVM)	92	0.90	0.88	0.89
Random Forest (RF)	94	0.93	0.91	0.92

Random Forest achieved the highest overall accuracy and F-score, indicating superior performance in multi-class intrusion detection.

Table 2. Class-wise Performance of the Proposed Model (Random Forest)

Attack Class	Precision	Recall	F-Score
Normal	0.94	0.96	0.95
Probe	0.91	0.88	0.89
R2L	0.78	0.72	0.75
DoS	0.97	0.95	0.96
U2R	0.74	0.68	0.71

High recall for DoS shows strong detection of high-volume attacks, while lower values for R2L and U2R reflect class imbalance**.**

Table 3. Real-Time Prediction Results from Deployed NIDS

Input Attack	True Class	Predicted Class	Confidence
Neptune	DoS	DoS	0.96
Satan	Probe	Probe	0.91
Normal	Normal	Normal	0.94

The deployed Flask-based system produced correct predictions for all tested samples with an average latency of 120ms.

Table 4. Confusion Matrix Summary (Random Forest)

Class	Correctly Classified (%)	Misclassified as Normal (%)
DoS	95	2
Probe	88	5
R2L	72	6
U2R	68	7

Most misclassifications occurred between Normal and low-frequency attacks.

Random Forest achieved the best overall performance, followed by SVM, Decision Tree, and Logistic Regression, demonstrating the effectiveness of ensemble learning for multi-class intrusion detection. The highest detection rates were observed for DoS attacks because of their distinctive high-volume traffic patterns, which are easier for the model to learn and classify. In contrast, R2L and U2R showed comparatively lower recall due to their low frequency and behavioral similarity to normal traffic, leading to limited training representation. The deployed Flask-based system produced correct real-time predictions with low latency, confirming the practical feasibility of the proposed model for operational cyber threat monitoring in supply chain environments

V. DISCUSSION

The results demonstrate that the proposed AI-driven cyber threat prediction framework can effectively classify multiple attack categories using behavioural network features while supporting real-time deployment. The superior performance of the Random Forest model compared with Logistic Regression, Decision Tree, and Support Vector Machine confirms the advantage of ensemble learning in handling high-dimensional intrusion detection data. By combining multiple decision paths, Random Forest reduces overfitting and improves generalisation, which explains its higher accuracy and F-score. This finding aligns with earlier intrusion detection studies that report improved multi-class classification performance when ensemble methods are applied to KDD-based datasets [3],[4].

The strong detection rate for DoS attacks reflects the distinct traffic characteristics associated with such intrusions. High connection frequency, repeated service requests, and elevated error rates create clear behavioural patterns that machine learning models can easily distinguish from normal traffic. Similar observations have been reported in prior work, where DoS attacks consistently achieve the highest recall due to their volumetric nature [1]. From a supply chain perspective, this capability is particularly significant because DoS attacks can disrupt shared platforms and halt inter-organisational operations. Early identification therefore contributes directly to business continuity and service availability.

Probe attacks also achieved high precision and recall, indicating that variations in service distribution and host interaction features provide sufficient discriminatory information. These results suggest that behavioural indicators can serve as effective proxies for Cyber Threat Intelligence attributes such as reconnaissance activity and lateral movement. By capturing these patterns, the model moves beyond simple anomaly detection toward intelligence-informed prediction, which is essential for proactive supply chain risk management. This supports the argument that integrating CTI concepts with machine learning enhances contextual threat interpretation rather than producing isolated alerts [5].

In contrast, the lower recall observed for R2L and U2R attacks highlights a persistent challenge in intrusion detection research. These attack types occur infrequently and often resemble legitimate user behaviour, particularly when they involve authentication attempts or privilege escalation. The limited number of training samples reduces the model's ability to learn distinctive patterns, resulting in misclassification as normal traffic. This limitation is widely acknowledged in the literature and is not unique to the present study [7]. Nevertheless, the model maintained acceptable precision for these categories, indicating that when such attacks were predicted, the predictions were generally accurate. Future work incorporating data balancing techniques or sequential models may improve detection of these low-frequency intrusions.

The real-time deployment of the model represents a key contribution beyond conventional IDS research. Many previous studies evaluate machine learning performance in offline environments without addressing operational constraints such as latency, feature encoding, and user interaction. The successful integration of the trained model into a Flask-based web application demonstrates that lightweight behavioural features can support rapid inference, with prediction latency below 120 milliseconds. This confirms the feasibility of deploying AI-driven intrusion detection in distributed supply chain settings where timely decision-making is critical.

From a theoretical perspective, the findings reinforce the applicability of supervised anomaly classification models for cyber threat prediction in interconnected networks. The results also illustrate how CTI-aligned behavioural features can bridge the gap between intelligence collection and automated detection. By

mapping traffic patterns to specific attack categories, the system provides actionable outputs that can inform risk prioritisation and mitigation strategies across supply chain nodes.

Despite these contributions, several limitations should be acknowledged. The use of a KDD-based dataset, while suitable for benchmarking, may not fully capture the complexity of modern encrypted and cloud-based traffic. Additionally, the class imbalance inherent in intrusion datasets affects the detection of minority attacks. The current model also relies on static feature inputs rather than continuous packet streams, which may limit its ability to capture temporal attack behaviour.

Overall, the study demonstrates that an AI-driven, CTI-informed intrusion detection framework can achieve high accuracy for major attack categories while remaining lightweight and deployable. The combination of ensemble learning, behavioural feature analysis, and real-time implementation provides a practical foundation for proactive cyber threat monitoring in secure supply chain environments. Future research should explore live traffic integration, adaptive learning mechanisms, and advanced models capable of improving detection of low-frequency attacks without compromising real-time performance.

VI. CONCLUSION

This study presented an AI-driven cyber threat prediction framework designed to enhance security in interconnected supply chain environments. The primary objective was to develop a lightweight, deployable machine learning–based Network Intrusion Detection System capable of classifying network traffic into Normal, DoS, Probe, R2L, and U2R categories using behavioural features aligned with Cyber Threat Intelligence principles. The results demonstrated strong overall performance, with Random Forest achieving the highest accuracy and F-score, and particularly high detection rates for DoS and Probe attacks. The successful real-time deployment of the model through a Flask-based web application confirmed its practical feasibility for operational cyber threat monitoring with low prediction latency.

From a theoretical standpoint, the findings support the effectiveness of supervised ensemble learning for multi-class intrusion detection and highlight the value of CTI-informed behavioural indicators in improving contextual threat classification. By integrating intelligence concepts with machine learning and real-time implementation, the study bridges the gap between offline IDS research and practical supply chain cybersecurity solutions. The framework enables proactive identification of high-risk traffic patterns, contributing to reduced detection time and improved decision support for distributed organisational networks.

However, the study is constrained by the use of a benchmark dataset that may not fully represent contemporary encrypted and cloud-based traffic, as well as class imbalance affecting the detection of low-frequency attacks such as R2L and U2R. Future research should focus on incorporating live packet capture, adaptive learning techniques, and advanced models capable of improving minority class detection while maintaining low computational overhead.

Overall, this work advances the application of AI-driven predictive analytics in secure supply chains by demonstrating a deployable, real-time intrusion detection framework that combines machine learning efficiency with intelligence-informed threat analysis.

VII. REFERENCES

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD Cup 99 data set,” in *Proc. IEEE Symp. Computational Intelligence for Security and Defense Applications*, Ottawa, ON, Canada, 2009, pp. 1–6.

[2] K. Scarfone and P. Mell, “Guide to intrusion detection and prevention systems (IDPS),” NIST Special Publication 800-94, 2007.

[3] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[4] A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[5] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, “Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains,” in *Leading Issues in Information Warfare & Security Research*, 2011, pp. 80–106.

[6] G. Kim, S. Lee, and S. Kim, “A novel hybrid intrusion detection method integrating anomaly

detection with misuse detection," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1690–1700, 2014.

[7] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[8] S. Boyson, "Cyber supply chain risk management: Revolutionizing the strategic control of critical IT systems," *Technovation*, vol. 34, no. 7, pp. 342–353, 2014.

[9] W. Lee and S. J. Stolfo, "A framework for constructing features and models for intrusion detection systems," *ACM Transactions on Information and System Security*, vol. 3, no. 4, pp. 227–261, 2000.

[10] J. McHugh, "Testing intrusion detection systems: A critique of the 1998 and 1999 DARPA intrusion detection system evaluations," *ACM Transactions on Information and System Security*, vol. 3, no. 4, pp. 262–294, 2000.

[11] G. Creech and J. Hu, "A semantic approach to host-based intrusion detection systems using contiguous and discontiguous system call patterns," *IEEE Transactions on Computers*, vol. 63, no. 4, pp. 807–819, 2014.

[12] H. Hindy et al., "A taxonomy and survey of intrusion detection system design techniques, network threats and datasets," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1–33, 2020.

[13] NSL-KDD Dataset. [Online]. Available: https://www.unb.ca/cic/datasets/nsl.html

[14] Scikit-learn: Machine Learning in Python. [Online]. Available: https://scikit-learn.org

[15] CICIDS2017 Dataset. [Online]. Available: https://www.unb.ca/cic/datasets/ids-2017.html