



Research Article

Developed Lightweight Endpoint Detection and Response Prototype Using Machine Learning for Efficient Threat Detection and Prioritization

Mary Chiamaka Ekwughaonu^{1*} , Theophilus Aniemeka Enem² , Freeman Bitrus³  and Okoroafor Chinedu David⁴ 

^{1,2,3,4}Department of Cyber Security, Faculty of Computing, Air Force Institute of Technology, Kaduna, Nigeria

Article Information

Article History

Received: 2 October 2025

Revised: 12 February 2026

Accepted: 15 March 2026

Published online: 2 April 2026

Keywords

Endpoint Detection

Machine Learning

Endpoint Detection

MITRE ATT&CK

Response

Correspondence*

maryekwughaonu@gmail.com

ORCID

Mary Chiamaka Ekwughaonu 

<https://orcid.org/0009-0007-3623-0755>

Theophilus Aniemeka Enem 

<https://orcid.org/0000-0002-5245-8828>

Freeman Bitrus 

<https://orcid.org/0009-0005-9964-5128>

Okoroafor Chinedu David 

<https://orcid.org/0009-0007-1782-6334>

Abstract

The rising frequency of zero-day and advanced persistent threats highlights the need for effective cybersecurity solutions for resource-limited endpoints. This study aimed to develop a lightweight endpoint detection and response (EDR) prototype using machine learning for efficient threat detection and prioritization. The prototype integrated Convolutional Neural Networks (CNN) and Random Forest models, with Dockerized Elasticsearch and Kibana for real-time log analysis. Datasets from Windows logs and CICIDS-2018 were used for training and testing. The CNN achieved 91% accuracy on Windows logs, and the Random Forest achieved 97% accuracy on CICIDS-2018 with a 5% false positive rate, while maintaining 15–20% CPU usage and a 50–100 ms response time. The system outperformed traditional EDR tools, offering an efficient, scalable, and resource-friendly solution suitable for both military and enterprise environments.

© 2026 Centre for Research and Innovation (CRI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

With the rise of technology and innovation, there is an urgent need for cybersecurity in the current age, as it is vital for new innovations. Cybercriminals leverage endpoint devices such as computers, laptops, and smartphones because they can easily access critical information and sensitive components. Threat actors develop newer threats (spyware, insider threats, etc.) due to technological advancements; therefore, strategic measures are required to protect against these threats. To overcome these limitations, Endpoint Detection and Response (EDR) is an essential cybersecurity solution that

focuses on securing endpoint devices such as computers, mobile devices, and servers, which are often the targets of cyberattacks [1]. EDR systems regularly monitor, detect, and respond to threats on endpoint devices in real time. However, traditional EDR systems have significant limitations, such as reliance on predefined threat signatures, which may not detect unknown threats, making them less effective against evolving threats, and high computational resource requirements. Moreover, Machine Learning (ML) provides a modern and advanced solution to address these challenges. With the help of ML, Endpoint Detection and Response systems can identify and recognize advanced threats by

examining behavior and detecting suspicious activities. Techniques such as supervised and unsupervised learning can automate threat prioritization and reduce response times, addressing critical gaps in traditional security frameworks. This study aims to develop a lightweight EDR prototype that integrates Machine Learning to improve threat detection and prioritization, with a focus on resource efficiency and real-time response.

As organizations increasingly rely on interconnected systems, endpoints have become a focal point for threat actors. According to the Ponemon Institute, 68% of organizations have experienced one or more endpoint attacks that successfully compromised data and IT infrastructure. Prominent breaches such as the Target breach in 2013 and the Equifax breach in 2017 illustrate the urgent need for endpoint protection, as attackers exploit vulnerabilities in these devices, resulting in substantial financial losses and reputational damage. Statistical evidence indicates that data breaches are becoming more frequent and sophisticated. According to IBM, the average cost of a data breach reached an all-time high of \$4.45 million in 2023, including legal fees, fines, and recovery efforts. This represents a 2.3% increase from 2022 and a 15.3% increase over the last three years. According to the Verizon 2023 Data Breach Investigations Report, the median cost per ransomware incident more than doubled over the past two years to \$26,000, with 95% of incidents that experienced a loss costing between \$1 and \$2.25 million. In modern organizational networks, endpoint devices are essential, serving as critical infrastructure for digital operations; however, they remain among the most vulnerable components in a network and are frequently targeted by cybercriminals. Advanced threats exploit undiscovered security flaws, resulting in severe consequences, including financial losses and reputational harm.

Although new cyberattack vectors have emerged, Endpoint Detection and Response (EDR) systems have become increasingly important. They monitor activity in real time and collect various types of telemetry data to identify patterns and detect potential threats. While traditional antivirus solutions rely on static signature-based detection (a signature is a unique identifier for a known threat), EDR systems apply behavioral analysis to identify and correlate suspicious activities. However, most EDR systems require substantial computational resources, which can render them impractical for systems with limited resources, such as older PCs or IoT devices [2]. Moreover, these systems generate large volumes of alerts, many of which are false positives, making it difficult for security teams to triage legitimate threats [3]. Machine Learning (ML) is a disruptive technology with the potential to address these challenges. Studies [4] demonstrate that ML enhances the ability of EDR systems to identify abnormal patterns and respond to emerging threats through

sophisticated behavioral analysis. While traditional methods may miss important indicators of compromise (IoCs), integrating ML into EDR systems enables deeper analysis of large datasets to identify patterns correlated with malicious activity. This capability improves detection accuracy and reduces false positive rates.

Despite these innovations, the evolving nature of cyber threats calls for continuous development of security technologies. This project addresses these issues by developing a lightweight EDR prototype that utilizes machine learning for effective threat detection and prioritization, with an emphasis on scalability and resource efficiency. By overcoming the constraints of legacy systems, this study seeks to provide an actionable and adaptable solution to contemporary endpoint security challenges.

II. LITERATURE REVIEW

A. Theoretical Framework

1. Evolution Of Endpoint Detection And Response: The EDR system development journey has evolved through various phases, starting with basic signature-based detection and culminating in sophisticated artificial intelligence systems. In the early 2000s, endpoint security operated mainly using static signature detection, despite its inability to protect organizations against unknown threats or adaptive attacks. The mid-2010s introduced behavioral analytics and sandboxing technologies, which enabled real-time monitoring of process execution, registry changes, and network traffic to detect anomalous patterns. During the 2010s, ML models were developed to learn from historical data and detect signatures of malicious events.

The endpoint detection and response system functions as a cybersecurity tool that tracks and detects security attacks targeting endpoint devices, including desktop computers, servers, and IoT infrastructure (Eventus Security, 2025). Traditional antivirus tools operate differently from EDR, which requires continuous monitoring, behavioral analysis, and automated threat response functions to counter advanced threats. The security solution from Trend Micro (2025) includes EDR, as it integrates machine data collection, real-time endpoint monitoring, and refined risk-correlation detection to identify illegitimate endpoint activities and trigger automated responses that enforce system isolation.

P. Shripad *et al* (2024) describes how its product protects endpoints through real-time data analysis for malware and ransomware threats. Cisco's analysis states that EDR solutions perform environmental threat detection and investigate threat life cycles to reveal incident chronology, entry points, current activity, and remediation options. Endpoints remain protected through EDR because the

solution prevents attacks from progressing. According to Palo Alto Networks (n.d.), Endpoint Detection and Response (EDR) is a security method designed to identify and mitigate threats targeting endpoint devices such as laptops, desktops, and mobile devices.

The integration of advanced machine learning (ML) techniques within Endpoint Detection and Response (EDR) systems has gained significant attention in addressing the complexities of modern cyber threats. This literature review synthesizes various studies focusing on the development of lightweight EDR prototypes, with an emphasis on efficient threat detection and prioritization.

2. Enhancing Malware Detection Efficacy: A Comparative Analysis of Endpoint Security and Application Whitelisting [5]: In this study, [4] conducted a comprehensive analysis of four endpoint security solutions—Endpoint Detection and Response (EDR), Network Detection and Response (NDR), antivirus, and application whitelisting—to evaluate their efficiency in detecting malware. The study illustrates the strengths of EDR, particularly its detailed visibility in detecting fileless malware using machine learning (ML), advanced persistent threats (APTs), and heuristic analysis. In contrast, application whitelisting is recognized for its proactive defense against zero-day threats by allowing only approved applications to execute. The authors elaborate on the essential role of machine learning in enhancing detection accuracy across all solutions, noting its capability to reduce false positives and adapt to emerging threats. The study concludes by recommending hybrid methods that integrate EDR, application whitelisting, and malware analysis to address the challenges posed by evolving threats.

- a. *Strengths:* The paper’s strength lies in its support for a hybrid security model, providing a detailed evaluation of four endpoint security solutions integrated with machine learning to improve detection rates. It employs large-scale, real-world datasets, demonstrating the practical application of these solutions in real-world cybersecurity environments.
- b. *Weaknesses:* The study’s drawbacks include the lack of empirical case studies addressing operational challenges such as alert fatigue and the reliance on static features, which limits the model’s ability to adapt to emerging threats. The study also does not address lightweight optimization and overlooks endpoint-specific threats such as fileless malware, which are essential considerations for deploying ML-driven solutions in resource-constrained environments.

3. Semi-supervised Based Unknown Attack Detection in EDR Environment [6]: The authors introduce a semi-supervised machine learning model for detecting unknown EDR attacks. The method applies an Autoencoder in collaboration with a 1D Convolutional Neural Network (1D-CNN) for anomaly detection in event logs without requiring prior signature knowledge. The system is based on historical logs spanning

one month, assuming that any deviations after that period may signal security threats. The model processed commercial endpoint data and detected 37 unknown attacks, of which VirusTotal confirmed 26 as malicious.

- a. *Strengths:* The research presents a newly developed semi-supervised learning system that requires fewer labeled datasets to operate. The approach effectively integrates an Autoencoder with a 1D-CNN for anomaly detection, enabling monitoring of behavioral patterns along with temporal dynamics. Real-world testing demonstrates its practical value, and the method delivers strong performance with minimal false alerts, making it suitable for EDR implementation with reduced analyst workload requirements.
- b. *Weaknesses:* The approach depends on the assumption that all logged historical data is normal; however, this may result in flawed training data if threats remain undetected. The paper does not compare the performance of its approach with other machine learning models, such as Random Forest and LSTM, leaving its relative effectiveness unestablished. Additionally, the lack of attack-type classification reduces its practical value for detailed incident response and threat remediation tasks.

4. Evolution of Endpoint Detection and Response (EDR) in Cyber Security: A Comprehensive Review [7]: The development of EDR technologies is explained in detail through [7], a systematic review that examines the transition from signature-based detection in the past to modern AI- and ML-powered systems. The research analyzes technological developments from 2010 to 2023 through four evolutionary stages, beginning with signature-based early detection systems (2005–2010), followed by behavioral analysis (2011–2015), the integration of machine learning with cloud platforms (2016–2020), and finally the emergence of artificial intelligence and extended detection and response solutions (2021–2023). The study investigates the Deep Ocean Protection System (DOPS) along with XDR to understand how their combined threat detection capabilities enable automated monitoring and multi-domain visibility. It identifies three main challenges faced by contemporary AI-based solutions: excessive false positive alarms, adversarial attacks, and high computational requirements.

- a. *Strengths:* Through this extensive structured review spanning more than a decade, the authors demonstrate how EDR solutions have evolved from reactive to proactive systems. The paper examines XDR and DOPS through practical examples to illustrate real-world applications, along with a comprehensive assessment of AI/ML methods for detection improvement and emerging complexities. It also provides detailed coverage of next-generation trends, including zero-trust and cloud-native EDR, highlighting their relevance for future research.
- b. *Weaknesses:* The study relies exclusively on previously published research and does not include any experimental data. During the literature selection phase,

the use of keywords as a filtering method may have resulted in the omission of relevant studies. The article also fails to address non-technical barriers, such as budget constraints or organizational resistance to change. Although the research focuses on XDR as a primary topic, it does not extensively analyze its scalability and interoperability challenges.

III. METHODOLOGY

This section presents the systematic approach adopted to design, implement, and evaluate a lightweight Endpoint Detection and Response (EDR) prototype that integrates machine learning for efficient threat detection and prioritization. The methodology covers the full workflow, from data acquisition and preprocessing to model development, system architecture, and integration with real-time monitoring tools. The design emphasizes resource efficiency and real-time responsiveness, directly addressing the performance and scalability limitations of traditional signature-based EDR systems.

A. Research Method and Design

The study employs an experimental research design aimed at developing and validating a lightweight EDR prototype that enhances endpoint security through supervised machine learning. The experimental framework tests hypotheses on the effectiveness of ML models in detecting and prioritizing cyber threats while maintaining low computational overhead. The approach follows iterative cycles of data collection, model training, testing, and refinement to empirically evaluate performance using key metrics such as detection accuracy, response time, and resource utilization.

The methodology draws inspiration from contemporary cybersecurity research emphasizing lightweight and scalable architectures. Specifically, insights from [8] and [9] influenced the selection of a hybrid ML strategy that balances accuracy with efficiency. The study integrates real-world datasets and endpoint logs, including malware data from Kaggle, Windows logs collected via Winlogbeat, and network traffic captured through Packetbeat, to ensure realistic model evaluation. Two supervised learning algorithms were employed: Convolutional Neural Networks (CNNs) and Random Forests (RF). CNNs excel at identifying complex temporal and behavioral patterns within endpoint logs, while Random Forests leverage ensemble learning to improve classification accuracy and reduce overfitting. Together, these models provide a robust and adaptable solution for threat detection and prioritization.

B. Data Acquisition

Data collection was performed from multiple reliable sources to capture diverse cyber threat characteristics. The CIC-IDS2017 and CIC-IDS2018 datasets were used because of their comprehensive network traffic logs and detailed intrusion behavior records, providing a strong foundation for malicious activity modeling. Additionally, Winlogbeat and Packetbeat were used alongside the loghub-master repository from GitHub to gather endpoint logs containing authentication events, user activity, and potential security risks. This combination of network and endpoint data ensured that the system was trained and validated using realistic attack patterns and normal operational behaviors.

C. Data Preprocessing and Feature Extraction

The acquired datasets contained irregularities such as noise, missing values, and redundant features, which required extensive preprocessing to improve model reliability. The cleaning process involved removing duplicate entries, correcting errors, and imputing or discarding incomplete records. To optimize efficiency and enhance model interpretability, Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) were applied for feature selection, reducing dimensionality and computational cost. Finally, data normalization using Min-Max scaling or Z-score standardization was performed to ensure consistent value ranges across features, thereby improving the convergence rate and overall accuracy of the machine learning models.

D. System Architecture and Design

The proposed lightweight EDR prototype follows a modular and scalable architecture composed of several key components. First, Docker Desktop was used for containerization, allowing Elasticsearch and Kibana to be deployed efficiently for centralized log management, storage, and visualization. Winlogbeat and Packetbeat served as data collection agents, with the former capturing endpoint event logs and the latter monitoring network traffic; both transmitted data to Elasticsearch for real-time analysis. The backend module processed incoming logs, executed ML model predictions, and forwarded classified results to the frontend interface. The frontend provided an interactive dashboard displaying live threat detections, severity levels, and system activity, enabling security analysts to respond swiftly to alerts. This architecture ensures a responsive, portable, and resource-efficient EDR system optimized for both real-time detection and scalability.

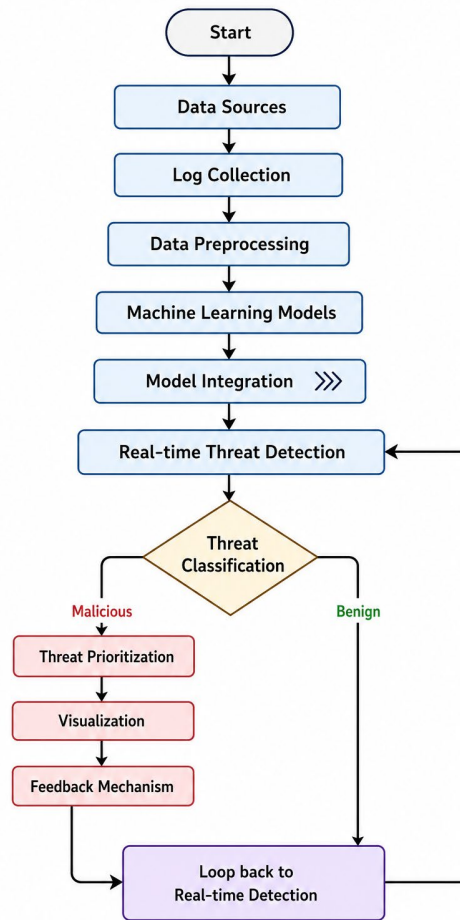


Fig.1 Architectural Workflow of Lightweight EDR (Author,2025)

E. Machine Learning Model Training and Integration

The effectiveness of the proposed EDR prototype depends heavily on the robustness of its machine learning models. To ensure efficient processing of large datasets, model training was conducted in Google Colab, leveraging its GPU/TPU acceleration for faster computation and optimization. The training process began with merging related datasets into two main categories: network intrusion and anomaly detection data (from CICIDS-2017/2018) and endpoint log data (from Winlogbeat and Packetbeat).

Two primary algorithms formed the core of the model selection framework: Convolutional Neural Networks (CNNs) and Random Forests (RFs). The CNN model was trained for pattern recognition and behavioral analysis, effectively identifying temporal anomalies and complex event correlations within endpoint logs. The Random Forest model, based on ensemble learning principles, was employed for classification tasks to minimize overfitting and maximize detection accuracy, particularly in distinguishing malicious traffic within large network datasets.

Model performance was validated using cross-validation techniques to ensure generalization and stability across diverse test sets. Hyperparameter tuning was conducted to achieve an optimal balance between detection accuracy, response time, and computational efficiency. Upon achieving satisfactory performance, the trained models were exported as .h5 (CNN) and .pkl (RF) files and integrated into the backend of the EDR system. These deployed models enabled real-time inference, allowing the system to continuously monitor, classify, and flag anomalous activities without human intervention.

F. Threat Detection, Prioritization, and MITRE ATT&CK Integration

This section represents the operational core of the EDR prototype, encompassing threat detection, prioritization, and visualization through integration with the MITRE ATT&CK framework.

1. *Detection Phase:* The CNN and Random Forest models formed the foundation of the detection engine. Both models analyzed real-time logs retrieved from Elasticsearch, which served as a centralized data repository. The CNN focused on endpoint behavior, identifying anomalous patterns such as irregular logon

activities or process executions, while the Random Forest classifier specialized in network intrusion detection, distinguishing benign from malicious traffic in the CICIDS datasets. The combination of these two models provided dual-layer coverage across both endpoint and network domains, ensuring comprehensive detection accuracy through a binary classification system that categorized data as either benign or malicious.

2. *Prioritization Phase:* Following detection, the system applied a threat prioritization mechanism based on the MITRE ATT&CK framework. Each identified threat was evaluated and scored according to its potential severity, impact, and the sophistication of the adversarial technique employed. Detected anomalies were mapped to relevant MITRE ATT&CK tactics and techniques, such as Initial Access, Execution, Privilege Escalation, or Lateral Movement, to establish structured threat categories. This mapping allowed the system to rank threats by urgency, enabling cybersecurity analysts to concentrate on the most critical incidents first, thereby improving incident response efficiency and reducing investigation time.
3. *Visualization Phase:* The frontend interface acted as the visualization and control layer, designed to provide real-time situational awareness. It featured an interactive dashboard that displayed prioritized threat data enriched with MITRE ATT&CK context, including tactic

descriptions, affected systems, attack vectors, and recommended mitigation strategies. This interface was dynamically updated through backend data processing pipelines, ensuring that security teams could view live system statuses and threat severity levels. The intuitive layout allowed analysts to trace attack origins, review correlated activities, and initiate rapid countermeasures, thereby facilitating a proactive defense posture and improving overall system responsiveness.

IV. PERFORMANCE EVALUATION

Performance evaluation of the prototype takes place in the final stage of the methodology through a comprehensive assessment.

1. The model evaluation relies on standard metrics such as accuracy, precision, recall, and F1-score to measure performance. The evaluation process also examines system operational speed and overall scalability.
2. A portion of unknown data, including real-time endpoint logs, is used for system testing under simulated operational conditions.
3. The testing procedures confirm both the functionality of the prototype and its operational superiority over existing innovative methods.

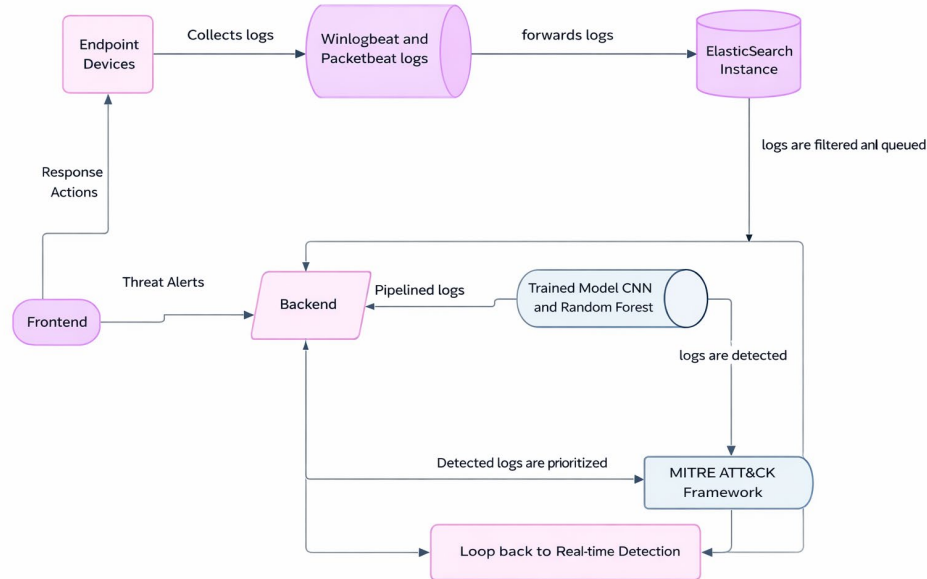


Fig.2 Workflow of UML Diagram (Author,2025)

V. RESULTS AND DISCUSSION

The evaluation of the developed lightweight Endpoint Detection and Response (EDR) prototype demonstrated the effectiveness of integrating machine learning for intelligent and efficient threat detection. The system was tested using the CICIDS-2017/2018 datasets for network intrusion analysis and Winlogbeat/Packetbeat logs for endpoint monitoring. Performance metrics such as accuracy, precision, recall, false

positive rate, and resource utilization were used to assess system efficiency. The Convolutional Neural Network (CNN) model achieved an average detection accuracy of 91% on endpoint logs, effectively identifying behavioral anomalies and suspicious user activities. The Random Forest model, applied to network-based intrusion data, recorded an even higher detection accuracy of 97%, with a false positive rate of approximately 5%. This indicates strong classification capability and robustness against data noise.

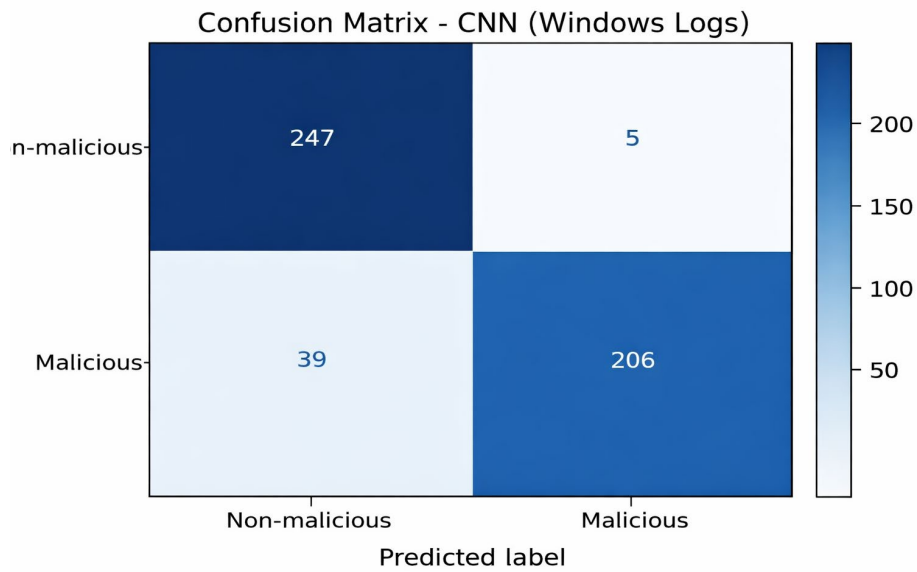


Fig.4 Confusion Matrix

TABLE II RANDOM FOREST MODEL EVALUATION METRICS FOR CICIDS-2017

Class	Precision	Recall	F1-Score	Support
BENIGN	0.99	0.88	0.93	97,371
Bot	0.11	0.71	0.18	414
DDoS	0.94	0.95	0.95	8,001
DoS Slowloris	0.64	0.86	0.74	262
FTP-Patator	0.19	0.97	0.32	1,656
PortScan	0.01	0.50	0.01	2
SSH-Patator	0.01	0.48	0.02	31
Accuracy	-	-	0.89	107,737
Macro Avg	0.41	0.77	0.45	107,737
Weighted Avg	0.97	0.89	0.92	107,737

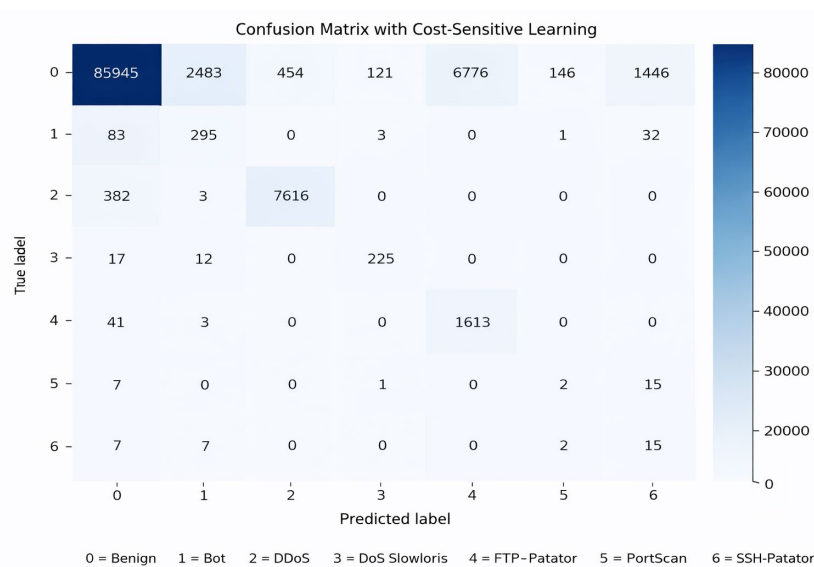


Fig.5 Confusion Matrix of Random Forest Model on CICIDS-2017

VI. CONCLUSION AND RECOMMENDATIONS

The study successfully developed and evaluated a lightweight, machine learning-based Endpoint Detection and Response (EDR) prototype designed to enhance threat detection, prioritization, and system performance in real-time environments. By combining Convolutional Neural Networks (CNNs) and Random Forest algorithms, the prototype achieved a balanced integration of accuracy, speed, and resource efficiency. The hybrid model effectively addressed the limitations of traditional EDR tools, which often depend on static signatures and require substantial computational resources. The research demonstrated that machine learning-driven EDR systems can significantly improve cybersecurity operations by automating detection, reducing false positives, and prioritizing critical threats. Furthermore, the use of Docker containerization, Elasticsearch, and Kibana ensured scalability and portability, making the system adaptable to different organizational environments. Despite its strong performance, the study faced limitations such as reliance on publicly available datasets, limited real-world testing scenarios, and computational constraints during training. Future improvements should focus on expanding dataset diversity, integrating reinforcement learning for adaptive responses, and implementing a distributed detection architecture to enhance scalability and detection precision. In conclusion, this research contributes a practical and efficient approach to modern cybersecurity by demonstrating how lightweight, ML-powered EDR systems can provide robust, scalable, and resource-efficient protection against emerging cyber threats.

The findings affirm that the integration of artificial intelligence into endpoint security represents a sustainable path toward proactive and intelligent cyber defense.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The authors confirm that no AI-assisted technologies were used in the preparation or writing of the manuscript, and no images were altered using AI.

REFERENCES

- [1] P. Shripad, G. Roopesh, and S. Sharma, "Machine learning-based lightweight endpoint detection and response systems," *Int. J. Cyber Defense Research*, vol. 18, no. 2, pp. 54–67, 2024.
- [2] Z. B. Yusof, "Effectiveness of endpoint detection and response solutions in combating modern cyber threats," *J. Adv. Cybersecurity Sci.*, vol. 8, no. 12, 2024.
- [3] N. Rananga and H. S. Venter, "A comprehensive review of machine learning applications in cybersecurity," *Research Square*, Preprint, 2023.
- [4] S. Sewak, X. Deng, and A. Ingle, "Machine learning-driven EDR systems for abnormal pattern detection," *J. Cyber Defense Analytics*, vol. 6, no. 3, pp. 91–108, 2023.
- [5] Eventus Security, "Endpoint Detection and Response (EDR)," 2025.
- [6] Trend Micro, "What is EDR?" 2025.
- [7] SentinelOne, "Understanding EDR Technology," 2025.
- [8] M. Althamir *et al.*, "Enhancing malware detection efficacy," *J. Theoretical Applied Info. Tech.*, vol. 102, no. 6, pp. 2451–2465, 2024.
- [9] Hwang, J. Kim, and S. Lee, "Semi-supervised unknown attack detection in EDR environments," 2020.
- [10] H. Kaur *et al.*, "Evolution of endpoint detection and response in cybersecurity," *E3S Web Conf.*, vol. 556, Art. no. 01006, 2024.