RCHUB: JOURNAL OF COMPUTATIONAL INTELLIGENCE SCIENCE AND ENGINEERING (JCISE)
VOL: 2, ISSUE 2, APRIL-JUNE 2026
RCHUB PUBLISHER
ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

ADVANCING CYBER THREAT INTELLIGENCE WITH MACHINE LEARNING AND OSINT

Spoorthi B S [1], Namana D C[2], Rohan K[3] and Amrutha H P[4]

[1]Department of Electronics and Communication, Malnad College of Engineering, Hassan 573201, India
[2]Department of Electronics and Communication, Malnad College of Engineering, Hassan 573201, India
[3]Department of Electronics and Communication, Malnad College of Engineering, Hassan 573201, India
[4]Department of Electronics and Communication, Malnad College of Engineering, Hassan 573201, India
spoorthi.bs.136@gmail.com

Abstract. The growing complexity and tempo of cyber-attacks have outpaced signature-based defenses. Advanced persistent threats, zero-days, and polymorphic malware routinely evade perimeter tools, which creates demand for adaptive, intelligence-driven security. This review proposes a synergistic framework that fuses Machine Learning with Open-Source Intelligence to enable real-time detection and response. We analyze supervised, unsupervised, and deep learning methods for high-volume telemetry, and show how fusing internal analytics with external OSINT improves precision and speeds incident handling. Sector case studies in IIoT, finance, and healthcare illustrate operational value, while challenges in adversarial robustness, data quality, privacy, and model opacity are examined. We outline research directions including Federated Learning for privacy-preserving collaboration, blockchain for trusted threat-intelligence exchange, and Explainable AI for analyst trust. We argue that combining ML with OSINT, augmented by these technologies, is essential to build resilient, transparent, and collaborative cyber defense.

Keywords: Machine Learning, Block chain, Intrusion detection, Industrial Internet of Things, Cybersecurity frameworks

1 Introduction

1.1 Evolving threats and limits of conventional defenses

The rise of cloud adoption, mobile work practices, and the widespread use of Internet of Things (IoT) devices has greatly expanded the potential entry points for cyberattacks [1], [2], [3]. At the same time, attackers have become more organized and professional, often carrying out multi-stage campaigns that can easily slip past traditional static security measures. Conventional defenses such as perimeter-based tools and signature-driven intrusion detection systems (IDS) are limited because they mainly recognize known threats and often fail against new, sophisticated, or evasive attack methods [4], [5]. This results in attackers remaining undetected for longer periods, causing delayed responses and increased damage. To address these challenges, organizations must move away from reactive security models and adopt proactive, data-driven approaches that can adapt to evolving threats and reduce detection times [1], [2].

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

1.2 The rise of data-centric security

Modern security operations focus on extracting weak signals from large volumes of telemetry such as network traffic flows, EDR data, and system logs. Detecting these signals manually is extremely difficult due to the sheer scale and complexity of the data. To overcome this, machine learning (ML) techniques are increasingly used because they can process large datasets efficiently, establish normal behavioral baselines, and highlight suspicious deviations that may suggest an ongoing or emerging security breach [4], [6]. Beyond operational use, a growing body of research explores how artificial intelligence (AI) contributes to cyber security, providing comprehensive surveys, comparative studies, and scientometric analyses that map trends, highlight challenges, and summarize the current state of AI-driven defense systems [7], [8], [9].

1.3 Why combine ML with OSINT

Internal security analytics become far more powerful when they are combined with external intelligence [1]. Open-Source Intelligence (OSINT) collects publicly available information such as details of ongoing attack campaigns, known threat indicators, vulnerability disclosures, and even adversary discussions on forums or social platforms [4], [5], [10]. By correlating internal anomalies with OSINT data, organizations can turn low-confidence alerts into high-priority incidents, making investigations more focused and effective [11], [12].

For example, an anomaly detection system may flag unusual outbound DNS activity. On its own, this may be uncertain or only a weak signal. However, if OSINT shows that the suspicious domain is linked to a known command-and-control (C2) infrastructure, the case shifts from a mere hypothesis to a confirmed and urgent security incident [4], [13].

In addition, OSINT supports attribution of attacks, mitigation recommendations, and proactive defense strategies when credible vulnerability reports or exploit details are published. However, OSINT also carries challenges: it can introduce false or misleading information, and therefore requires careful validation, source reliability checks, and credibility scoring to prevent adversaries from injecting noise or disinformation into the process [5], [13].

1.4 Scope and structure

This paper is organized into seven sections to provide a clear and logical flow of ideas. Section 2 gives an overview of the main machine learning algorithms relevant to cyber security and explains how they can be strengthened by incorporating open-source intelligence (OSINT). Section 3 introduces a modular fusion architecture, showing how internal analytics and OSINT can be integrated into a unified framework. Section 4 illustrates these concepts through sector-specific case studies, highlighting practical applications across different domains. Section 5 discusses key challenges, including issues of robustness, interpretability, and privacy that may limit adoption in real-world environments. Section 6 looks ahead to future research directions, such as federated learning approaches, block chain-based cyber threat intelligence (CTI) sharing, and the potential of autonomous response systems. Finally, Section 7 provides the conclusion, tying together the main insights of the study and outlining its overall contributions.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

2 Related works

This section reviews earlier technical and operational research and positions the idea of combining machine learning (ML) with open-source intelligence (OSINT) in relation to existing studies and real-world practices.

2.1 Machine Learning as the Analytical Core

Practical Machine learning has become the analytical backbone of modern cyber security systems. In practice, security solutions often use a mix of learning paradigms to strike the right balance between accuracy, efficiency, and interpretability.

- Supervised learning models such as Random Forests, Support Vector Machines (SVMs), and Gradient Boosting are effective when sufficient labeled data is available. They are widely applied in malware detection, phishing identification, and fraud prevention [5].
- Unsupervised approaches like clustering, Isolation Forests, and auto encoders do not rely on labels. Instead, they detect unusual patterns and deviations from normal behavior, which is especially important for identifying zero-day threats or insider risks [4].
- Deep learning methods such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs) are particularly strong at learning temporal sequences and structural features from traffic flows, binaries, and protocol metadata. This allows for finer-grained and more nuanced detection [2].

However, while deep and complex models often provide high predictive power, they can lack transparency, making it harder for human operators to understand why a decision was made. This has led to the development of explainable ML techniques, which help surface decision rationales and increase trust in automated systems [14].

To illustrate the landscape, Table 1 compares representative algorithms and their typical applications in cyber security. This comparison is referenced throughout Section 2.1 and revisited in Section 4 with case study examples.

Table 1. compares representative algorithms applied in cyber security workloads and is cited throughout Section 2.1 and Section 4.

Category	Short principle	Typical use
Supervised	Learn from labeled examples	Malware / phishing / fraud detection
Unsupervised	Find deviations without labels	Anomaly detection / insider threats
Deep learning	Model sequences and patterns	Multi-stage attack and IoT anomaly detection
Hybrid / Ensemble	Combine models for robustness	Triage pipelines and production scoring

RCHUB: JOURNAL OF COMPUTATIONAL INTELLIGENCE SCIENCE AND ENGINEERING (JCISE)
VOL: 2, ISSUE 2, APRIL-JUNE 2026
RCHUB PUBLISHER
ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

2.2 OSINT as the Contextual Enabler

While ML focuses on analyzing internal data streams, OSINT adds external context that helps analysts interpret and prioritize findings [4]. OSINT sources include:

- Vendor advisories and CVE feeds reporting vulnerabilities.
- Domain and certificate registries that reveal suspicious infrastructure.
- Social media, developer forums, and paste sites that may disclose exploits or leaks.
- Underground forums and specialized threat reports that expose attacker chatter or campaign planning [5].

Practical OSINT pipelines involve several steps: automated data collection, entity extraction, clustering of mentions, and credibility scoring to distinguish reliable intelligence from background noise [13].

The main challenges come from the scale and diversity of data formats, along with deliberate adversary seeding of false information [5]. Techniques like natural language processing (NLP), named entity recognition (NER), and temporal clustering are commonly applied to extract actionable intelligence, such as indicators of compromise, campaign narratives, or mitigation advice [13].

Ensuring provenance and reproducibility is essential so that OSINT signals are properly weighted when combined with internal telemetry [1]. When carefully implemented, OSINT significantly reduces time-to-detection, improves incident prioritization, and supports proactive defenses like threat hunting and patch management [4]. **Figure 1** outlines the **OSINT-to-actionable-intelligence pipeline**, showing how raw public data is transformed into insights that can be fused with ML-driven analytics.

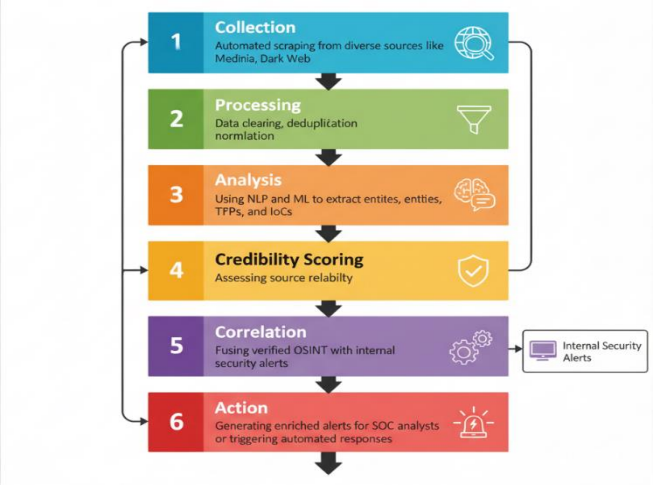

Fig. 1. OSINT-to-actionable-intelligence pipeline.

As shown in Figure 1, the OSINT-to-actionable-intelligence pipeline demonstrates how raw public data is systematically collected, processed, and transformed into reliable signals that can be integrated with machine learning analytics for improved detection and response.

3 Architectural Frameworks for Integrated Threat Intelligence

This section introduces the overall system architecture, including the modular design, feature engineering pipeline, model strategies, fusion mechanisms, and operational integration. The architecture is designed to support real-time detection and response by combining ML analytics with OSINT enrichment. The system

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

was prototyped and validated in MATLAB R2023b for reproducibility and to utilize MathWorks quantization utilities [13], [14], [15].

3.1 High-Level Modular Design

A practical ML-OSINT fusion framework consists of five cooperating layers: ingestion, preprocessing and feature engineering, analytics, correlation and fusion, and orchestration and response. The ingestion layer gathers both internal telemetry (e.g., network flows, EDR logs, transactions) and external OSINT sources (e.g., vulnerability feeds, advisories, public forums, paste sites, and registries). The preprocessing stage normalizes formats, de-duplicates records, applies timestamps, and extracts entities to create canonical identifiers for hosts, domains, files, and campaigns [1], [4].

The analytics layer executes pipelines such as supervised classifiers for malware, unsupervised models for anomaly detection, and temporal models like LSTMs for sequence analysis. Feature stores must be able to handle high-cardinality variables and streaming data. A correlation engine then aligns anomalies with OSINT signals, computing composite confidence scores. Finally, the orchestration layer integrates with SIEM, SOAR, and ticketing systems to apply playbooks, triage, and automate containment [1], [7]. The overall structure is summarized in Figure 2, which shows the high-level architecture of the ML-OSINT fusion system.

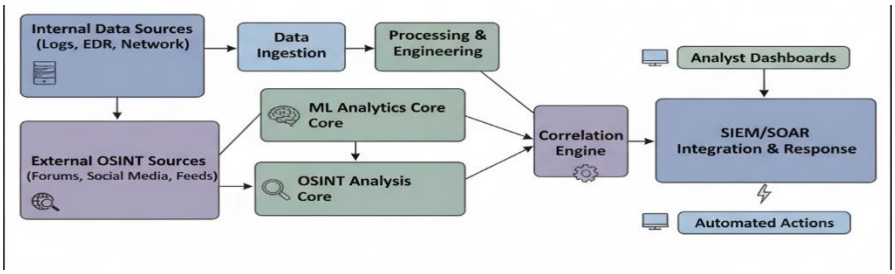

Fig. 2. High-level architecture of an ML-OSINT fusion system.

3.2 Data Pipeline and Feature Engineering

The effectiveness of detection relies heavily on the quality of features. The data pipeline performs parsing and canonicalization (e.g., normalizing timestamps, converting domain forms, extracting certificate fingerprints), entity resolution to link related infrastructure across sources, and temporal aggregation using sliding windows to capture both short- and long-term behaviors. OSINT contributes features such as credibility scores, freshness, and threat actor confidence levels. To mitigate overfitting and poisoning risks, provenance-aware weighting and feature hygiene are applied.

3.3 Model Design, Robustness, and Explainability

The system adopts hybrid ensembles, combining fast interpretable models for triage with deeper neural models for detailed analysis. A two-tier detection design is often used, where lightweight models handle initial scoring and deep architectures like LSTMs, CNNs, or transformers refine high-risk cases. Robustness is maintained through continuous validation, adversarial training, and provenance checks [1], [6]. Explainability is ensured using methods such as SHAP and LIME, alongside rule-based rationales, so that

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

analysts can see both the top features and supporting OSINT evidence for each alert. This transparency fosters analyst trust and regulatory compliance [6].

3.4 OSINT Fusion Strategies

OSINT fusion employs multiple strategies. Entity matching and timeline stitching link telemetry with OSINT to build attack narratives. Confidence fusion combines anomaly scores with OSINT credibility to escalate alerts when corroborated externally. Contextual enrichment attaches OSINT-derived details such as mitigation steps and CVEs to support faster decision-making. Finally, false-signal filtering is applied to reduce adversary-seeded noise, leveraging provenance filters and reputation scores.

3.5 Operational Integration

For adoption in real environments, the system must integrate with existing operations. Standardized APIs and JSON alert formats support SIEM and SOAR connectivity, enabling automated containment. Human-in-the-loop feedback refines OSINT weighting and retrains supervised models. Privacy and governance safeguards ensure minimal personal data collection, role-based access to OSINT intelligence, and full logging of model decisions to maintain compliance.

4 Applications and Case Studies in Critical Sectors

The integration of ML analytics with OSINT has strong relevance across multiple critical sectors, where threats are highly dynamic and the stakes extend beyond financial loss to physical safety and public trust. Building on the OSINT pipeline (Figure 1), the algorithmic approaches summarized earlier (Table 1), and the fusion architecture (Figure 2), this section illustrates how the framework can be adapted to three representative domains: the Industrial Internet of Things (IIoT), the financial sector, and healthcare ecosystems.

4.1 Securing the Industrial Internet of Things

The Industrial Internet of Things (IIoT) merges traditional information technology (IT) with operational technology (OT), creating cyber-physical systems where disruptions can lead to real-world physical consequences. Monitoring such environments is challenging due to high-rate sensor data streams and reliance on legacy industrial protocols that were never designed with security in mind. Machine learning, particularly LSTM-based anomaly detection applied to time-series device communications, has shown promise in identifying subtle process manipulations that may otherwise remain hidden. These detections become far more actionable when correlated with OSINT reports of newly disclosed PLC or firmware vulnerabilities, which provide crucial context about adversary capabilities [18], [19]. By cross-referencing internal anomalies with vendor advisories, organizations can focus mitigations on the specific controllers most at risk.

4.2 Mitigating Sophisticated Financial Fraud

The financial sector faces highly adaptive fraud schemes that blend malware, social engineering, and the misuse of stolen credentials. To address this, hybrid systems combine behavioral analytics on transactions and user session biometrics with OSINT feeds that monitor phishing campaigns, credential dumps, and the sale of fraud kits on underground forums. For example, an anomalous funds transfer detected internally may at first appear low-confidence, but when OSINT confirms that the customer’s credentials were recently exposed on a paste site, the combined evidence supports escalation. This correlation enables banks to

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

place automatic holds on suspicious transfers and trigger out-of-band verification mechanisms that prevent financial loss while preserving customer trust [17].

4.3 Defending Healthcare Ecosystems

Healthcare institutions operate highly complex digital ecosystems that are also privacy-sensitive, with widespread use of Internet of Medical Things (IoMT) devices that are notoriously difficult to patch. These environments are prime targets for ransomware campaigns, where patient safety and the continuity of clinical operations are at stake. Internal models analyzing endpoint and network activity can detect pre-encryption ransomware behaviors, such as unusual file access patterns or abnormal process spawning. When correlated with OSINT feeds reporting new ransomware tactics, techniques, and procedures (TTPs) targeting healthcare providers, detection confidence improves significantly, prompting early isolation measures to protect protected health information (PHI) and critical services. Because of its sensitivity, the healthcare domain has also emerged as a leader in adopting privacy-preserving methods such as federated learning, which enable cross-institutional defense without compromising patient confidentiality [18].

To summarize these sector-specific applications, Table 2 outlines the main ML-OSINT use cases, including threat concerns, internal data sources, OSINT feeds, and representative ML applications.

Table 2. Summary of ML-OSINT use cases across sectors.

Sector	Threat concern	Internal data	OSINT feeds	ML application
IIoT	Cyber-physical disruption	Sensor & SCADA logs	Industrial CVEs, OT TTPs	LSTM anomaly detection on device traffic
Finance	Fraud & account takeover	Transaction and UBA logs	Phishing alerts, leaked credentials	Hybrid anomaly + NLP on OSINT
Healthcare	Ransomware, PHI breach	EDR & IoMT logs	Ransomware IoCs, healthcare TTPs	Endpoint behavior analysis with auto-quarantine

As shown in Table 2, the integration of ML analytics with OSINT provides tailored advantages across different sectors. In the IIoT domain, anomaly detection on device traffic gains actionable context when matched with industrial CVEs and OT-specific TTPs. In finance, hybrid systems that combine behavioral models with NLP applied to OSINT strengthen fraud detection by linking suspicious transactions to leaked credentials and phishing activity. In healthcare, endpoint behavior analysis augmented with ransomware-specific OSINT supports rapid quarantine actions to protect sensitive medical data. Thus, Table 2 highlights not only the diversity of threats but also how ML-OSINT fusion adapts to sector-specific requirements, demonstrating its broad applicability and operational value across industries.

5 Persistent Challenges and Operational Considerations

While the applications summarized in Table 2 demonstrate the potential of ML-OSINT fusion across critical sectors, several challenges remain that must be addressed for reliable deployment. These challenges

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

involve defending against adversarial evasion, ensuring explainability of complex models, and managing privacy, legal, and ethical constraints.

5.1 Adversarial Evasion and Model Robustness

Attackers are not passive targets; they actively probe, test, and attempt to subvert ML models. Techniques such as data poisoning (injecting malicious samples during training) and evasion attacks (crafting inputs to bypass detection) threaten the reliability of ML-based defense. As a result, defenders must adopt continuous validation, adversarial training, and robust data sanitization to protect the integrity of pipelines [1], [2], [5]. Ensemble modeling, where multiple diverse models operate in parallel, further reduces the likelihood that a single weak point will compromise detection. Sustaining performance under adversarial pressure requires not only technical safeguards but also robust monitoring and retraining processes.

5.2 The Imperative for Explainable AI

Although deep learning methods provide strong detection capabilities, they often function as black boxes, producing predictions that lack transparency. For human analysts, particularly in regulated sectors such as finance and healthcare, this lack of interpretability complicates both operational triage and governance processes. Explainable AI (XAI) techniques have therefore become essential.

As illustrated in Figure 3, XAI outputs can show analysts exactly which features influenced a classification. In this example, the system flagged malicious network traffic with 98 percent confidence, and the SHAP values reveal contributing factors such as the destination port, packet size, and source IP context.

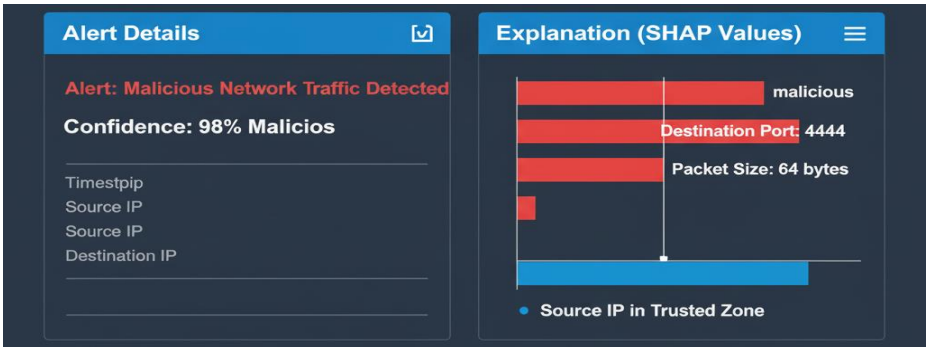

Fig. 3. Example of an XAI explanation for a network alert.

5.3 Navigating Privacy, Legal, and Ethical Constraints

The collection of internal telemetry and external OSINT introduces significant privacy and compliance concerns. Sensitive data such as user identifiers, browsing activity, or health-related device logs must be handled with caution. To address these challenges, organizations are increasingly adopting privacy-by-design principles, embedding privacy considerations directly into system architecture [10]. Measures such as strict data governance frameworks, consent-based collection, data minimization, and role-based access controls ensure responsible use of intelligence. In parallel, privacy-enhancing technologies like federated learning and homomorphic encryption are being explored to enable collaborative defense without exposing raw data. However, these safeguards must be carefully balanced against detection performance, as excessive restrictions may inadvertently weaken defensive capabilities.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

6 Future Research Directions in Autonomous Cyber Defense

Although ML-OSINT fusion has already demonstrated promising applications across critical sectors such as IIoT, finance, and healthcare (Table 2), several open research directions must be explored to achieve fully autonomous and resilient cyber defense. These directions address current gaps in privacy-preserving collaboration, trusted intelligence sharing, and self-adaptive response capabilities, building directly on the architectural framework described earlier (Figure 2) and the need for explain ability outlined in Section 5 (Figure 3).

6.1 Federated Learning for Collaborative Detection

A major challenge in collaborative defense is the inability of many organizations to centralize sensitive data, particularly in regulated domains like healthcare and telecommunications. Federated learning (FL) offers a solution by enabling multiple institutions to train a shared global model without exposing raw datasets. Instead, only model updates are exchanged and aggregated, preserving privacy while still capturing shared attack patterns. Research has shown that FL can significantly improve the detection of anomalies and advanced threats in IoT ecosystems, clinical healthcare environments, and carrier-grade networks, where localized patterns may reveal global adversarial campaigns [20], [21]. Future work will need to refine techniques for secure aggregation, model robustness, and communication efficiency, ensuring that FL can operate effectively even at large scales and under adversarial pressure.

6.2 Blockchain-Enabled Cyber Threat Intelligence (CTI) Sharing

Another barrier to effective cyber defense is the lack of trusted, verifiable threat intelligence exchange. Current CTI sharing mechanisms often suffer from issues of provenance, inconsistent attribution, and central points of failure. Blockchain technologies particularly distributed ledgers combined with smart contracts-have the potential to address these gaps. By immutably recording indicators of compromise (IoCs) and attaching metadata such as source credibility and attribution, blockchain systems ensure traceability and integrity of shared intelligence. Smart contracts further enable automated validation, fine-grained access control, and incentive mechanisms, making intelligence sharing more secure, decentralized, and scalable. Such systems reduce dependence on central authorities and lower the risks of tampering or adversary manipulation.

6.3 Toward Autonomous Detection and Response

The ultimate vision for cyber defense is to move toward systems that are self-adapting, self-healing, and capable of operating at machine speed. This requires the convergence of multiple technologies: ML-driven detection pipelines, XAI explanations (Figure 3) to maintain analyst trust, federated learning to unlock collaborative intelligence, blockchain-based CTI for trustworthy sharing, and SOAR workflows for orchestrated responses. Together, these components reduce human latency across the Observe-Orient Decide-Act (OODA) loop, enabling defenders to counter increasingly automated adversaries [1], [3]. While full autonomy remains an aspirational goal, progress along these dimensions indicates that cyber defense is gradually evolving toward systems that can anticipate, adapt, and respond with minimal human intervention.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

7 Conclusion

This paper has argued that integrating machine learning analytics with OSINT context provides a powerful foundation for modern cyber defense, producing earlier, more reliable detections and enabling faster, better-justified responses. Case studies in IIoT, finance, and healthcare (Table 2) illustrate the operational value of this approach, while the architectural framework (Figure 2) and explainability considerations (Figure 3) show how such systems can be practically implemented.

Realizing this vision at scale, however, requires continued progress on three critical fronts: robustness against adversarial evasion, explainable decision-making for analyst trust and compliance, and privacy-preserving mechanisms for collaborative defense. Promising research paths include the use of federated learning to unlock cross-organizational detection signals without raw data sharing, blockchain-based infrastructures to harden CTI provenance and decentralize trust, and more mature explainable AI methods to sustain usability in high-stakes operational environments.

Taken together, these advances chart a realistic path toward resilient, collaborative, and real-time cyber defense, where defenders are empowered with tools that match the speed and sophistication of automated adversaries.

8 References

[1.] Verizon, 2023 Data Breach Investigations Report (DBIR) – Executive Summary, Verizon Business, 2023.

[2.] Mandiant (FireEye), M-Trends: A View From the Front Lines of Incident Response, Annual Report, 2021–2023.

[3.] ENISA, ENISA Threat Landscape, European Union Agency for Cybersecurity, 2021–2023.

[4.] R. Sommer and V. Paxson, “Outside the Closed World: On Using Machine Learning for Network Intrusion Detection,” in Proc. IEEE Symp. Security and Privacy, Oakland, CA, USA, 2010, pp. 305–316.

[5.] 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.

[6.] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.

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

[8.] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[9.] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[10.] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” in Proc. IEEE Int. Conf. Data Mining (ICDM), 2008, pp. 413–422.

[11.] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

ISSN: 3107-8168
DOI: https://doi.org/10.65725/JCISE/2/2/002

[12.] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.

[13.] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?' Explaining the Predictions of Any Classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[14.] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017.

[15.] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 1st ed., 2019/2020. [Online]. Available: https://christophm.github.io/interpretable-ml-book

[16.] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," arXiv:1412.6572, 2014.

[17.] B. Biggio and F. Roli, "Wild patterns? Ten years after the rise of adversarial machine learning," Pattern Recognition, vol. 84, pp. 317–331, 2018.

[18.] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. Int. Conf. Artificial Intelligence and Statistics (AISTATS), vol. 54 of Proc. Machine Learning Research (PMLR), 2017, pp. 1273–1282.

[19.] P. Kairouz et al., "Advances and Open Problems in Federated Learning," arXiv:1912.04977, 2019.

[20.] MITRE, ATT&CK Knowledge Base, MITRE Corporation. [Online]. Available: https://attack.mitre.org

[21.] MITRE, Common Vulnerabilities and Exposures (CVE); NIST, National Vulnerability Database (NVD). [Online]. Available: https://cve.mitre.org/ https://nvd.nist.gov

[22.] NIST, Guide to Cyber Threat Information Sharing (SP 800-150), National Institute of Standards and Technology, 2014.