

Research Paper

A Comprehensive approach to intelligent intrusion Detection in cyber security using machine learning

¹Thota Bhargav

¹M.Tech, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
tbargavnaidu28@gmail.com

²Dr. B. V. S. Varma

²Professor, Department of Computer Science and Engineering
DNR College of Engineering & Technology (Autonomous)
Bhimavaram – 534202, Andhra Pradesh, India
phdvarma@gmail.com

Abstract— Security Operations Centers (SOCs) play a critical role in defending organizations against evolving cyber threats, yet traditional SOC environments struggle with high alert volumes, excessive false positives, and fragmented data sources, which delay timely threat response. This study proposes a User-Centric Machine Learning Framework designed to enhance cyber defense by integrating advanced machine learning techniques with human-centered design principles. The proposed framework collects and preprocesses multi-source security data, including network traffic, endpoint logs, and threat intelligence feeds, to enable efficient and unified analysis. Machine learning algorithms such as decision trees, random forests, and anomaly detection models are employed to classify threats and identify suspicious behavior patterns. Natural Language Processing (NLP) techniques are applied to analyze unstructured security reports and phishing content, extracting actionable threat intelligence. Interactive visualization dashboards reduce analyst workload by presenting prioritized alerts and contextual threat information, while continuous feedback from analysts improves model accuracy through adaptive learning and retraining mechanisms. Predictive analytics and automated response systems further support proactive threat mitigation and rapid incident handling. Explainable AI and role-based access control are incorporated to ensure transparency, security, and trust in automated decisions. Overall, the proposed framework improves SOC efficiency, reduces response time, minimizes false positives, and strengthens organizational cybersecurity resilience.

Keywords— Intrusion Detection, Machine Learning, Security Operations Center, Anomaly Detection, Natural Language Processing, Explainable AI, Threat Intelligence, Alert Triage, Adaptive Learning, Cybersecurity Resilience.

I. INTRODUCTION

Cybersecurity has become a critical concern for organizations as the volume, sophistication, and frequency of cyberattacks continue to grow. Security Operations Centers (SOCs) serve as the frontline defense, monitoring network traffic, endpoint activity, and threat intelligence feeds to detect and respond to malicious behavior. Effective intrusion detection strongly depends on the ability to process large volumes of security data, distinguish genuine threats from benign anomalies, and respond within actionable timeframes. Recent advances in artificial intelligence and machine learning have created opportunities for intelligent, user-centric intrusion detection and SOC decision support [1]–[4].

Conventional intrusion detection approaches generally rely on signature-based detection, static rule sets, or manual analyst review of security alerts. Although these methods provide interpretable and immediate decisions, they have limited capability to detect novel or evolving attack patterns and often generate excessive false positives, leading to analyst fatigue [6],

[7]. Machine-learning techniques can learn relationships among historical network and endpoint features to identify suspicious behavior beyond fixed rules [1], [2], [4]. Natural Language Processing (NLP) techniques can further extract actionable insights from unstructured security reports and phishing content [9]–[12].

Threat detection accuracy alone is important but insufficient, because SOC analysts must also be able to trust, interpret, and act on system outputs efficiently. Explainable AI and human-centered visualization approaches have therefore been explored to reduce alert fatigue and improve analyst decision-making [6]–[8]. However, high alert volumes, fragmented multi-source data, limited labeled attack data, evolving attacker tactics, and the need for real-time processing remain important challenges [3], [5], [14], [15].

Detection alone cannot provide complete SOC intelligence. A practical AI-based system should integrate data acquisition, preprocessing, threat classification, anomaly detection, explainability, and analyst-facing decision support. This survey therefore aims to:

1. Review machine-learning and deep-learning approaches for intrusion detection and SOC alert analysis.
2. Examine classification techniques for network traffic, endpoint logs, and threat intelligence data.
3. Analyze NLP-based approaches for unstructured security report and phishing-content analysis.
4. Compare the strengths, limitations, and applicability of existing intrusion detection methods.
5. Examine explainable AI, visualization, and alert-prioritization techniques for reducing analyst workload.
6. Formulate an integrated, user-centric AI-based SOC framework.
7. Identify evaluation requirements, research challenges, and future directions for reliable and trustworthy intrusion detection systems.

The remainder of this paper discusses existing intrusion detection and SOC approaches, the proposed user-centric framework, data representation, explainability, adaptive learning, evaluation requirements, challenges, future research directions, and concluding observations.

II. EXISTING METHODS FOR AI-BASED INTRUSION DETECTION AND SOC MANAGEMENT

Existing intrusion detection and SOC management approaches can be broadly classified into signature-based detection, statistical/anomaly-based methods, conventional machine

learning, deep-learning approaches, NLP-based threat analysis, and explainable/human-centered alert triage frameworks.

A. Signature-Based and Rule-Based Detection

Signature-based systems detect known attack patterns using predefined rules and threat signatures. They provide fast and interpretable detection but cannot identify novel or zero-day attacks and require frequent manual signature updates [7], [15].

B. Statistical and Anomaly-Based Methods

Statistical approaches identify intrusions by detecting deviations from expected network or user behavior baselines. These methods can detect previously unseen attacks but often produce high false-positive rates in dynamic network environments [14], [15].

C. Conventional Machine Learning

Machine-learning approaches learn relationships between network and endpoint features from historical labeled data. Algorithms such as Decision Trees, Random Forests, and Support Vector Machines have been widely used for intrusion classification [1], [2], [4]. Their advantages include interpretability and efficient computation, though performance depends on feature quality, dataset characteristics, and class imbalance.

D. Deep Learning-Based Detection

Deep-learning models can capture complex nonlinear relationships among network traffic and behavioral features. Neural-network architectures have been investigated for identifying sophisticated and multi-stage attack patterns [3], [5]. These approaches provide strong representation power but require substantial training data and computational resources.

E. NLP-Based Threat and Phishing Analysis

NLP techniques analyze unstructured security data such as threat reports, log narratives, and phishing email content to extract actionable threat intelligence. Recent studies have applied text-classification and semantic-analysis techniques to detect phishing attempts and summarize security incidents [9]–[12]. Such approaches extend detection capability beyond structured numerical data.

F. Explainable and Human-Centered Alert Triage

Explainable AI and human-centered visualization frameworks aim to reduce alert fatigue by prioritizing, correlating, and contextualizing security alerts for analysts. Recent SOC-focused research has integrated SHAP- and LIME-based explanations with alert-ranking mechanisms to improve analyst trust and triage efficiency [6]–[8].

TABLE I: COMPARISON OF EXISTING INTRUSION DETECTION AND SOC METHODS

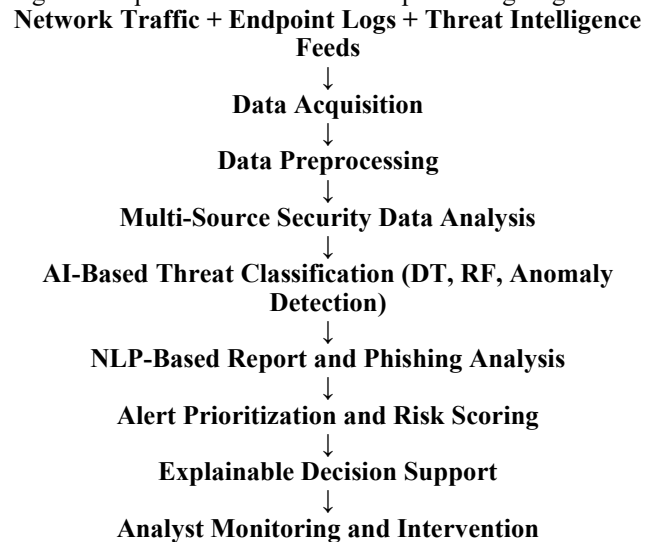
Method	Main Technique	Strength	Major Limitation
Signature-based	Rule matching	Fast, interpretable	Cannot detect novel attacks
Statistical/anomaly-based	Baseline deviation analysis	Detects unknown attacks	High false-positive rate
Machine learning	DT, RF, SVM	Automated classification	Data/feature dependent
Deep learning	Neural networks	Complex pattern learning	Requires sufficient data
NLP-based analysis	Text classification	Extracts unstructured threat intel	Language/domain dependent
Explainable alert triage	XAI + visualization	Reduces analyst workload	Requires careful design
Proposed framework	Integrated ML + NLP + XAI	Detection + explainability + human-	Requires comprehensive validation

centered triage

The literature therefore indicates a transition from static, signature-based detection toward adaptive, predictive, and human-centered intrusion detection. However, many existing approaches focus on a single detection or analysis task in isolation. A comprehensive architecture should integrate multi-source data acquisition, preprocessing, threat classification, NLP-based analysis, explainability, and analyst decision support into a unified SOC pipeline.

III. PROPOSED USER-CENTRIC AI-POWERED SOC FRAMEWORK

The proposed system considers SOC threat management as an integrated sequence of interconnected processing stages:



The architecture consists of five major functional layers: data acquisition, preprocessing, intelligent threat analysis, alert prioritization, and analyst-facing decision support. The first layer collects security data from multiple sources, including network traffic, endpoint logs, system events, and external threat intelligence feeds. The preprocessing layer removes noise, handles missing or inconsistent records, and converts heterogeneous security data into suitable input features.

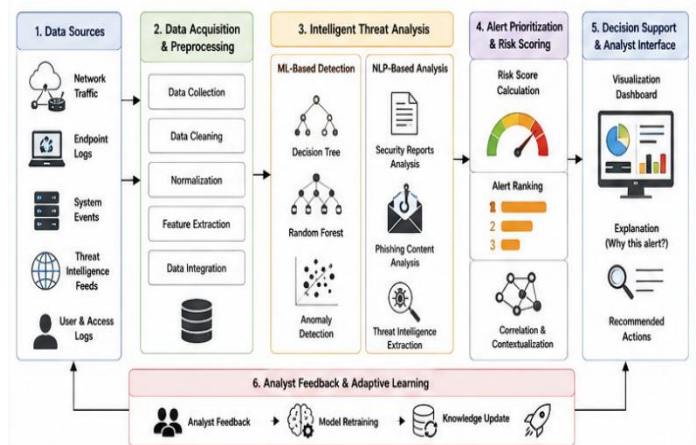


Fig. 1. Overall Architecture of the Proposed User-Centric AI-Powered SOC Framework

The intelligent analysis layer applies decision trees, random forests, and anomaly detection models to classify threats and

identify suspicious behavior patterns, while NLP components analyze unstructured security reports and phishing content in parallel. The alert-prioritization layer combines classification outputs and risk indicators to rank and contextualize alerts for analyst attention. Finally, the decision-support layer presents prioritized alerts, explanations, and recommended actions to analysts through an interactive visualization dashboard. The architecture is designed as a decision-support framework, assisting analysts in timely threat response rather than completely replacing human judgement. This integrated approach enables continuous monitoring, reduced alert fatigue, predictive threat detection, and more efficient SOC operations.

IV. SECURITY DATA AND THREAT REPRESENTATION

A. Security Data Acquisition and Processing

The proposed system collects security parameters such as network traffic features, endpoint log entries, system events, user behavior indicators, and threat intelligence attributes. Preprocessing handles missing, duplicate, inconsistent, and abnormal records, and normalizes heterogeneous data formats into a unified representation.

The structured security event is represented using key components: network-related features, endpoint-related observations, contextual/environmental indicators, and threat-intelligence-related indicators. These components are analyzed jointly to assess whether an event reflects normal activity or a potential intrusion. Normalization converts features with different scales and formats into comparable ranges, improving joint model learning and detection accuracy.



Fig. 2. Security Data Processing and Threat Classification Pipeline

B. Threat Condition Representation

The condition of a security event is represented using relevant network, endpoint, and contextual parameters. The system extracts:

- Source and destination IP/port information
- Traffic volume and frequency
- Protocol and packet-level characteristics
- Endpoint process and file activity
- User authentication and access behavior
- Threat intelligence indicators (known malicious signatures, IOCs)
- Textual content from security reports or phishing emails
- Historical alert and incident context

Representing the collected security parameters using a unified structure enables the proposed system to analyze the relationship between network behavior, endpoint activity, and

threat intelligence context. The resulting representation is subsequently provided to the decision tree, random forest, and anomaly detection models for threat classification, and to the NLP components for unstructured content analysis, depending on the objective of the proposed system.

V. THREAT-RESPONSE MATCHING AND ALERT PRIORITIZATION SCORING

A central component of the proposed SOC framework is threat-response matching. Unlike simple threshold-based alerting, the proposed approach identifies relationships between observed security events and appropriate response actions. The system considers parameters such as traffic anomaly severity, endpoint behavior, threat intelligence relevance, and historical incident context to determine the priority of each alert.

The classification and anomaly detection models generate a risk representation of each security event that captures behavioral patterns beyond simple rule-based comparison. This representation is compared against known threat profiles and organizational baselines to assess how closely the current event aligns with malicious activity.

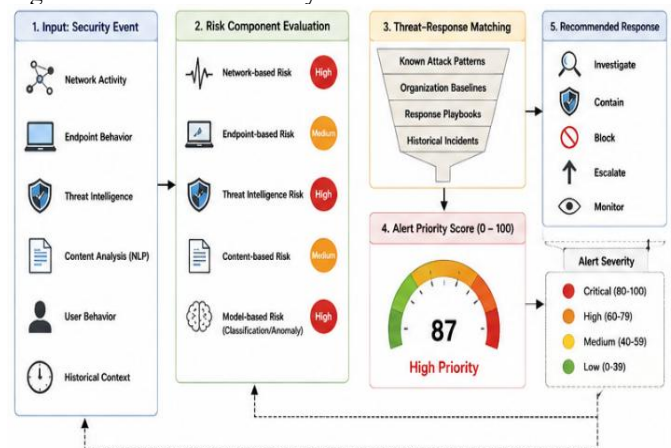


Fig. 3. Threat-Response Matching and Alert Priority Scoring

Model output alone, however, should not determine the final alert priority. An event with a moderate anomaly score may still contain a critical indicator, such as a known malicious IP or a credential-access pattern, that independently signals a high-priority threat. Therefore, the proposed system combines model-based risk with explicit threat-specific indicators.

The overall alert priority score is derived from a weighted combination of individual risk components, including:

- Network-based risk — how unusual the traffic pattern is relative to baseline behavior,
- Endpoint-based risk — suspicious process, file, or access activity on the host,
- Threat-intelligence-based risk — matches against known indicators of compromise,
- Content-based risk — NLP-derived risk from phishing or report content,
- Model-based risk — the combined classification and anomaly detection confidence.

This weighted formulation allows the SOC framework to assign different importance to different security indicators. For example, a network-monitoring application may assign greater importance to traffic-based risk, whereas a phishing-defense application may emphasize content-based risk.

Security events are subsequently evaluated according to their final priority scores to generate an alert-severity status and corresponding response recommendation, enabling timely intervention and reduced analyst workload.

VI. EXPLAINABLE AI AND DECISION-AWARE SOC MANAGEMENT



Fig. 4. Explainable AI Framework for Threat Prediction and Analyst Decision Support

AI-based SOC systems should provide more than a classification label or numerical score. Analysts need to understand why the system flagged a particular event as suspicious. Explainable AI (XAI) techniques such as SHAP and LIME can provide feature-level explanations of model predictions.

For each alert, the proposed system can generate an explanation containing:

1. Network traffic contribution
2. Endpoint behavior influence
3. Threat intelligence match influence
4. Textual/phishing content contribution
5. User behavior indicators
6. Major factors contributing to the final classification
7. Recommended SOC response action

For example:

Security Alert — 92% Malicious

- Network traffic: significant deviation from baseline volume.
- Endpoint behavior: unauthorized process execution detected.
- Threat intelligence: source IP matches known malicious indicator.
- Content analysis: associated email flagged as phishing.
- Major influencing factor: combination of threat intelligence match and endpoint anomaly.
- Recommendation: Escalate to Tier-2 analyst for immediate investigation and containment.

Such explanations help analysts understand whether the AI prediction is supported by meaningful security evidence rather than relying only on the final classification value.

A. Explainability and Decision Monitoring

SOC prediction models can be influenced by multiple network, endpoint, and contextual parameters. Therefore, the system should continuously monitor the contribution of each input feature to the prediction. Explainability helps identify which parameters have the greatest influence on threat classification,

false-positive reduction, or other target outcomes considered by the proposed system.

The proposed framework therefore integrates explainability as a decision-support component. The system should report both: Performance = How accurately does the system detect genuine security threats?

Explainability = Can the analyst understand which security parameters caused the prediction?

This dual objective is important because a highly accurate model that provides no understandable reasoning may be difficult to trust or use in practical SOC environments.

VII. ADAPTIVE SOC INTELLIGENCE AND HUMAN-IN-THE-LOOP DECISION SUPPORT

Security threats are not static. Attackers continuously develop new techniques, tools, and evasion strategies to bypass existing defenses. A threat detection model trained using fixed historical attack patterns may therefore become less effective as new threats emerge.

Model Update

Adaptive mechanisms can update detection thresholds, feature importance, threat intelligence knowledge, and response recommendations while maintaining model version control.

The system can periodically incorporate newly collected security data and analyst feedback to improve its understanding of evolving threats. However, adaptive AI should not automatically modify critical containment or blocking decisions without validation. Model updates should first be evaluated using historical and newly collected security datasets before being deployed.

Human-in-the-Loop Workflow

The final SOC decision-support workflow is:

Data Collection → AI Prediction → Threat Assessment → Explanation → Analyst Verification → Response Recommendation → Final Analyst Decision

The SOC analyst remains responsible for the final decision, particularly when security events are unusual or when additional contextual information is required that may not be available in the collected data.

This human-in-the-loop design allows analysts to combine AI-based predictions with practical security expertise and real-time contextual knowledge. Analyst feedback can also be captured and used for future model retraining, while preventing the AI system from making completely autonomous containment decisions without human supervision.

VIII. EVALUATION FRAMEWORK

A comprehensive evaluation framework is adopted to assess the proposed SOC framework across classification performance, alert prioritization, explainability, fairness, efficiency, and robustness. Rather than relying on a single performance measure, the evaluation considers multiple complementary dimensions to determine the practical effectiveness of the proposed framework.

A. Classification Performance

The classification module is evaluated using accuracy, precision, recall, and F1-score. Accuracy measures overall classification correctness, while precision indicates the proportion of flagged events that are genuinely malicious. Recall measures the ability to identify actual intrusions, and F1-score provides a balanced assessment of precision and recall. These metrics are important because both false alarms and

missed intrusions can significantly affect organizational security.

B. Alert Prioritization Performance

Since prioritizing high-risk alerts is a primary objective, the system is evaluated using ROC-AUC, Precision-Recall AUC, and reduction in false-positive rate. These metrics determine whether genuinely malicious events receive appropriately high priority scores and whether legitimate activity is not unnecessarily escalated.

C. NLP and Phishing Detection Performance

The NLP-based threat and phishing analysis components are evaluated separately using text-classification accuracy, precision, recall, and F1-score, assessing their effectiveness in extracting actionable insights from unstructured security content.

D. Explainability

The generated threat explanations are evaluated based on faithfulness, consistency, completeness, and human understandability. The explanation should accurately represent the factors influencing the alert classification and enable analysts to verify network, endpoint, threat-intelligence, and content-based indicators.

E. Fairness

Fairness evaluation examines potential differences in alert outcomes across controlled organizational segments (e.g., departments, user groups). Controlled testing can determine whether modifying irrelevant user attributes produces unjustified changes in threat classification. Thus, both predictive performance and responsible decision-making are considered.

F. Efficiency and Robustness

Computational evaluation considers alert processing time, detection latency, memory consumption, throughput, and scalability. Robustness is evaluated through testing on imbalanced datasets, varying attack types, and diverse network environments. This ensures that the proposed system maintains reliable performance under practical real-time SOC conditions. TABLE II EVALUATION DIMENSIONS OF THE PROPOSED SOC FRAMEWORK

Dimension	Evaluation Metrics / Criteria	Purpose
Classification	Accuracy, Precision, Recall, F1-Score	Evaluate intrusion identification performance
Alert Prioritization	ROC-AUC, PR-AUC, False-Positive Reduction	Evaluate alert triage quality
NLP/Phishing Analysis	Text classification Accuracy, Precision, Recall, F1-Score	Assess unstructured threat-intel extraction
Explainability	Faithfulness, Consistency, Completeness, Human Understandability	Evaluate decision transparency
Fairness	Group disparity, Controlled segment evaluation	Identify potential detection bias
Efficiency	Processing Time, Detection Latency, Memory, Throughput	Assess computational efficiency
Robustness	Imbalanced-data testing, Attack-type variation, Network variation	Evaluate generalization capability

IX. RESEARCH CHALLENGES AND FUTURE ROADMAP

Although AI-powered SOC frameworks can improve threat detection and alert triage, several challenges remain in developing a reliable and practical intrusion detection system.

A. Dataset Availability and Class Imbalance

Intrusion detection datasets typically contain a very small proportion of malicious events relative to benign traffic. Developing sufficiently diverse and representative datasets, along with effective resampling techniques, is therefore important for reliable threat classification.

B. Evolving Attack Patterns

Attackers continuously adapt their strategies to bypass detection systems. The proposed system must therefore handle variations in attack typologies, evasion techniques, and multi-stage intrusion sequences to achieve reliable detection.

C. Detection Reliability

Threat classification should consider multiple security-relevant factors rather than depending on a single anomaly score. Future improvements should focus on maintaining consistent detection performance across different network environments and organizational contexts.

D. Explainability

AI-generated security alerts should provide understandable reasons for recommendations. Future systems should improve the transparency of network-based, endpoint-based, threat-intelligence-based, and content-based risk contributions.

E. Fairness

SOC systems should minimize inappropriate influence from irrelevant user or organizational attributes. Fairness-aware evaluation is therefore required to ensure that threat flagging primarily reflects security-relevant risk indicators.

F. Real-Time Processing

Effective intrusion detection requires near-instantaneous decisions. Future SOC frameworks should therefore support low-latency inference without reducing detection accuracy, particularly for high-throughput network environments.

G. Human-AI Interaction

AI recommendations should assist SOC analysts rather than completely replace human judgement. Future systems should provide effective analyst feedback mechanisms and decision-support interfaces for reviewing and validating flagged alerts.

H. Privacy and Security

Security data contains sensitive network, endpoint, and user information. Secure data management, controlled access, anonymization, and privacy-preserving processing are therefore important for practical SOC deployment.

The research roadmap can therefore be organized into three stages:

Short Term: Improve multi-source data integration, decision tree/random forest tuning, class-imbalance handling, and explainability evaluation.

Medium Term: Develop adaptive detection-threshold updating, improved real-time inference pipelines, diverse intrusion datasets, and robust cross-environment evaluation.

Long Term: Develop reliable, human-centered SOC intelligence that combines adaptive learning, explainable threat scoring, fairness-aware evaluation, and human-in-the-loop decision support across organizations.

X. CONCLUSION

This paper presented a user-centric AI-powered framework for intelligent intrusion detection in Security Operations Centers. The framework addresses the limitations of conventional

signature-based and rule-based detection systems by incorporating multi-source data acquisition, preprocessing, machine-learning-based threat classification, NLP-based analysis of unstructured security content, alert prioritization, explainability, and human-in-the-loop decision support. The proposed approach classifies threats using decision trees, random forests, and anomaly detection models, while NLP techniques extract actionable insights from unstructured security reports and phishing content. Interactive visualization dashboards further reduce analyst workload by presenting prioritized alerts and contextual threat information, supported by continuous analyst feedback and adaptive retraining. Explainability and fairness are incorporated to improve the transparency and reliability of SOC recommendations. Analysts can examine the major factors contributing to threat classifications, while fairness evaluation helps identify potentially inappropriate influences on detection outcomes. The human-in-the-loop design, combined with role-based access control, ensures that AI-generated recommendations support analyst decision-making rather than completely replacing human judgement, while maintaining transparency, security, and trust in automated decisions. The proposed framework demonstrates that AI-powered, user-centric intrusion detection can significantly improve SOC efficiency, reduce response time, minimize false positives, and strengthen organizational cybersecurity resilience. Future research should focus on diverse and imbalanced intrusion datasets, adaptive threat-threshold learning, robust real-time evaluation, faithful explainability, fairness-aware learning, privacy-preserving processing, and reliable AI-based SOC decision support.

REFERENCES

- [1] M. A. M. Hasan, M. Nasser, S. Ahmad, and K. I. Molla, "Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique," *Cybersecurity*, vol. 5, art. no. 2, 2022, doi: 10.1186/s42400-021-00103-8. [Open access — free PDF]
- [2] T. Tulasi Bhavani, M. Kameswara Rao, and A. Manohar Reddy, "Network intrusion detection system using random forest and decision tree machine learning techniques," in *Advances in Intelligent Systems and Computing*, Springer, 2020, doi: 10.1007/978-981-15-0029-9_50.
- [3] "A deep learning/machine learning approach for anomaly based network intrusion detection," *Frontiers in Artificial Intelligence*, 2025, doi: 10.3389/frai.2025.1625891. [Open access — free PDF]
- [4] J. L. Resende and A. C. Drummond, "A survey of random forest based methods for intrusion detection systems," *ACM Computing Surveys*, vol. 51, no. 3, art. no. 48, 2018, doi: 10.1145/3178582.
- [5] N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 dataset," *Information Security Journal*, 2016.
- [6] S. Ndichu, T. Ban, S. Ozawa, T. Takahashi, and D. Inoue, "AI-driven security alert screening and alert fatigue mitigation in security operations centers: A survey," *ACM Computing Surveys*, 2026. [Free preprint: arxiv.org/abs/2605.08316]
- [7] S. Alahmed, "Alert fatigue in security operations centres: Research challenges and opportunities," *ACM Computing Surveys*, 2025, doi: 10.1145/3723158. [Open access]
- [8] "An explainable AI based alert triage framework for security operations centres: Architecture, evaluation, and analyst centred design (ExTriage)," *InfoTech Spectrum: Iraqi Journal of Data Science*, 2026. [Open access]
- [9] V. Bountakas and C. Xenakis, "HELPHED: Hybrid ensemble learning phishing email detection," *Journal of Network and Computer Applications*, 2023.
- [10] P. An, R. Shafi, T. Mughogho, and O. A. Onyango, "Multilingual email phishing attacks detection using OSINT and machine learning," *arXiv preprint arXiv:2501.08723*, 2025. [Free PDF]
- [11] A. Chakravorty, M. Price, N. Elsayed, and Z. ElSayed, "Context-aware phishing email detection using machine learning and NLP," *arXiv preprint arXiv:2603.27326*, 2026. [Free PDF]
- [12] I. Altan, A. Bachir, Y. Parbhulkar, A. M. Rizvi, and M. Farazi, "Dual-path phishing detection: Integrating transformer-based NLP with structural URL analysis," *arXiv preprint arXiv:2509.20972*, 2025. [Free PDF]
- [13] A. Karim, M. Shahroz, K. Mustofa, S. B. Belhaouari, and S. R. K. Joga, "Phishing detection system through hybrid machine learning based on URL," *IEEE Access*, vol. 11, pp. 36805–36822, 2023. [Open access]
- [14] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.
- [15] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Computers & Security*, vol. 28, no. 1–2, pp. 18–28, 2009.
- [16] A. Eriksson and M. Grov, "Explainable AI for SOC alert analysis: A user study on trust and transparency using SHAP and LIME," presented at a Security Operations research venue, 2024–2025