OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbreviated Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.21599909

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-06|Issue-04|2026

Novel Method for Detecting and Mitigating Advanced Persistent Threats (APTs) leveraging Artificial Intelligence and Machine Learning Techniques

Meghana Vastrad[1], Girish Kumar B. C[2], Aparna K. S[3]& Shivaprakash Vastrad[4]

[1]Student, Dept. of Information Science, RV Institute of Technology and Management, Bengaluru, Karnataka, India.
[2]Assistant Professor, Dept. of Information Science, RV Institute of Technology and Management, Bengaluru, Karnataka, India.
[3]Assistant Professor, Dept. of Computer Science, Global Academy of Technology, Bengaluru, Karnataka, India.
[4]Senior Director, Ministry of Electronics & Information Technology, National Informatics Centre,Karnataka State Centre.

Article History
Received: 05.05.2026
Accepted: 22.06.2026
Published: 25.07.2026

Citation
Vastrad, M., Kumar, G. B. C., Aparna, K. S. & Vastrad, S. (2026). Novel Method for Detecting and Mitigating Advanced Persistent Threats (APTs) leveraging Artificial Intelligence and Machine Learning Techniques. *Indiana Journal of Multidisciplinary Research, 6*(4), 142-147.

Abstract: Advanced Persistent Threats (APTs) present a major cybersecurity challenge due to their stealthy, long-term, and targeted nature. Traditional security systems often struggle to detect them, leading to increased use of artificial intelligence (AI) and machine learning (ML) techniques. This paper reviews recent studies applying supervised learning, unsupervised clustering, ensemble methods, and deep learning for APT detection and mitigation. These approaches show strong performance across evaluation metrics, especially when hybrid or ensemble models are used.

Key challenges include limited labeled datasets, the need for model interpretability, concept drift, and resistance to adversarial attacks. To address these, innovative methods such as belief rule base (BRB) modeling, federated learning, and explainable AI (XAI) are explored to improve transparency and adaptability. The paper also highlights real-world applications, particularly in finance and critical infrastructure sectors.

Overall, the findings emphasize the effectiveness of integrating AI-driven detection systems into existing cybersecurity frameworks. However, continuous model updates, diverse datasets, and collaboration between humans and AI systems are essential. Future research should focus on developing scalable, privacy-preserving, and interpretable solutions to keep pace with evolving cyber threats.

Keywords: Advanced Persistent Threats, Machine Learning, Cyber-security, Anomaly Detection, Deep Learning, Explainable AI

INTRODUCTION

In the evolving landscape of cybersecurity, Advanced Persistent Threats (APTs) represent a significant and sophisticated challenge to organizational networks. Unlike conventional cyberattacks that aim for quick infiltration and disruption, APTs are characterized by their stealth, persistence, and targeted nature. These attacks are often orchestrated by well-funded and highly skilled adversaries with the objective of gaining prolonged access to critical systems, typically for purposes such as espionage, data theft, or sabotage.

APTs exploit multiple attack vectors—including spear-phishing, zero-day vulnerabilities, and social engineering—to bypass traditional security defenses and remain undetected for extended periods. Their multi-phase structure, typically mapped using the Cyber Kill Chain model, allows attackers to perform reconnaissance, establish footholds, escalate privileges, and exfiltrate data with precision. This level of sophistication renders conventional signature-based intrusion detection systems largely ineffective against APTs.

Given the increasing reliance on digital infrastructure across sectors such as finance, defense, healthcare, and critical infrastructure, detecting and mitigating APTs has become a top priority for cybersecurity professionals. This paper explores the fundamental characteristics of APTs, the limitations of existing defense mechanisms, and the emerging role of artificial intelligence and machine learning in identifying and countering these threats in real-time.

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative technologies at the forefront of modern computing. By enabling systems to learn from data, adapt to new inputs, and make autonomous decisions, AI/ML have redefined problem-solving across a wide range of domains—from healthcare and finance to cybersecurity, transportation, and education. As data becomes

***Corresponding Author:** Girish Kumar B. C

increasingly abundant and computational power continues to grow, the integration of AI/ML into real-world applications is no longer a theoretical concept but a practical necessity.

MATERIALS

Advanced Persistent Threats (APTs) are not just ordinary cyber attacks—they are stealthy, targeted, and long-lasting. Typically carried out by highly skilled attackers, including nation-state groups and organized cybercriminals, APTs aim to quietly infiltrate a network, stay undetected for long periods, and gradually extract sensitive or valuable data. Their slow and methodical nature makes them especially difficult to detect using traditional security tools.

For many years, organizations have relied on traditional detection methods such as signature-based systems**,** rule-based monitoring**,** and heuristic techniques. Signature-based detection works by comparing incoming data to known patterns of malicious activity. While this approach is fast and effective for well-known threats, it struggles with newer, unknown, or disguised attacks—something APTs often take advantage of. Similarly, rule-based systems like SIEM (Security Information and Event Management) tools use predefined rules to spot suspicious behavior. But these rules usually focus on clear-cut patterns and can miss subtle, slow-moving APT activities that mimic normal user behavior.

To improve on this, some systems use heuristics—basically, educated guesses based on past behaviors—to flag abnormal activity. However, these methods can be overly sensitive and may generate a high number of false alarms, which makes it difficult for security teams to respond effectively. Plus, traditional tools often work in a reactive manner—they detect and respond after something happens, rather than predicting or preventing it in advance. Because of these limitations, there's a growing awareness that traditional methods alone are not enough to combat modern threats like APTs. This has led to increased interest in more intelligent, adaptive approaches, especially those involving machine learning and artificial intelligence, which are better suited to identifying new patterns, learning from behavior over time, and staying one step ahead of attackers.

As cyber threats become more complex and persistent, traditional security tools are often left playing catch-up. Advanced Persistent Threats (APTs), in particular, are designed to evade basic detection methods by moving slowly, blending into normal network activity, and constantly evolving their techniques. To keep up with this level of sophistication, many organizations are now turning to Artificial Intelligence (AI) and Machine Learning (ML) as a smarter, more adaptive solution for detecting and responding to such threats.

AI/ML-based techniques differ from conventional methods in that they are not limited to static rules or known attack signatures. Instead, they focus on learning patterns from large volumes of data—network traffic, user behavior, system logs, and more—to identify subtle signs of malicious activity. For example, supervised learning models like Random Forests**,** Support Vector Machines (SVM)**,** or XGBoost can be trained on labeled datasets to recognize what normal and abnormal behavior look like. Unsupervised methods like clustering or anomaly detection go a step further by identifying unusual activity without needing labeled data, making them ideal for spotting new, previously unseen APT patterns.

Deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs)**,** have also shown promise, especially when dealing with complex, high-dimensional data such as time-series logs or sequences of user actions. These models can capture hidden relationships and long-term dependencies that traditional methods often miss.

In addition to accuracy, AI/ML techniques bring scalability and speed to threat detection. They can process vast amounts of data in real time, reduce false positives, and even automate responses based on confidence levels. However, these systems are not without challenges. Issues like data quality, model interpretability, adversarial attacks, and the need for continuous learning still remain important areas of concern.

Overall, AI and ML are proving to be powerful tools in the fight against APTs. They bring a much-needed level of intelligence and adaptability to cybersecurity systems and are increasingly being integrated into modern threat detection frameworks.

METHODS

An Adaptive, Explainable, Hybrid-AI Framework for Real-time APT Detection and Response Based on the analysis, the proposed method aims to overcome the "minus points" by combining the best practices and addressing the identified research gaps. It will be:

1. **Comprehensive in Data Collection:** Moving beyond single-source data (e.g., DNS logs) to multi-source for a holistic view.
2. **Adaptive to Evolving Threats:** Incorporating dynamic feature selection and continuous learning for concept drift and novel at-tacks.
3. **Resilient to Data Limitations:** Leveraging self-supervised learning for unknown threats and addressing limited labeled data.
4. **Interpretable and Trustworthy:** Integrating explainability tools and a human-in-the-loop for analyst trust and reduced false positives.
5. **Real-time and Efficient:** Utilizing stream-based processing and optimized models for real-time operations.

***Corresponding Author:** Girish Kumar B. C

6. **Actionable:** Incorporating threat prioritization and response mechanisms.

Proposed Framework:

1.Multi-Source Real-time Data Ingestion & Correlation:

- **Sources**: Collect data from diverse sources including network traffic (packet captures, NetFlow), host activities (endpoint telemetry, user logins, file modifications), and application/system logs (SIEM logs, cloud infrastructure logs). **Intelligence Integration:** Supplement with Open Threat Intelli-gence Feeds (OSINT) and MITRE ATT&CK mappings for contextual enrichment and known threat identification.
- **Stream Processing:** Implement a high-throughput, stream-based pipeline for continuous real-time data ingestion and initial corre-lation, crucial for detecting "low-and-slow" APT movements.
- **Unified Schema:** Normalize heterogeneous data sources into a unified schema (e.g., Open Cybersecurity Schema Framework OCSF) to ensure compatibility and consistency across the pipeline.

2. Adaptive Data Preprocessing & Feature Engineering

- **Automated Feature Engineering:** Utilize AutoML feature ex-tractors to generate relevant features from raw data, reducing manual effort and improving feature quality.
- **Dynamic Feature Selection:** Implement adaptive feature selection mechanisms (e.g., statistical variance thresholds that adjust over time) to account for concept drift and maintain model relevance to evolving attack patterns. This addresses the limitation of static datasets and evolving threats.
- **Data Quality & Normalization:** Implement robust data cleaning, outlier handling, and normalization techniques to address data quality issues.

3. Hybrid AI Detection Engine:

Ensemble Supervised Learning Module:

- **Purpose:** Detect *known* APT attack patterns by training ensemble models (e.g., Random Forest, Gradient Boosted Trees, XGBoost) on rich, labeled historical APT datasets, including simulated multi-stage attack data.
- **Imbalance Handling:** Employ cost-sensitive learning or over/under-sampling techniques to effectively handle class imbalance inherent in rare APT events.

Self-Supervised Anomaly Detection Module:

- **Purpose:** Address the challenge of *unknown* threats and limited labeled data by learning representations of "normal" behavior.
- **Model:** Implement self-supervised models (e.g., Deep Info-Max, Contrastive Predictive Coding, Autoencoders) to learn latent feature representations of legitimate activities from the vast majority of unlabeled data.
- **Unsupervised Detection:** Deploy unsupervised anomaly detection algorithms (e.g., Isolation Forest, One-Class SVM) on these learned representations to identify deviations indicating potential novel APT activity.

Reinforcement Learning (RL) Threat Prioritization & Response:

- **Purpose:** Dynamically adjust alert severity, learn optimal response policies, and reduce false positives.
- **RL Agent:** Use a reinforcement learning agent (e.g., Proximal Policy Optimization - PPO, Multi-Agent Reinforcement Learning) that learns from real-time feedback, rewarding correct detections and penalizing false alarms. This helps in intelligent decision-making and dynamic adaptation.
- **Adaptive Response:** The RL agent informs response actions, potentially automating initial containment or suggesting next steps, improving real-time response capabilities.

4. Explainability and Human-in-the-Loop (Analyst Feedback Loop): Explainable AI (XAI):

Integrate SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Ex-planations) to provide human-readable reasoning behind model alerts. This addresses the "black-box" nature of some AI models and builds analyst trust.

Active Learning Loop: Introduce an active learning mechanism where security analysts validate high-uncertainty alerts (flagged by models or RL agent). The feedback from these validations is used for periodic model retraining and refinement, continuously improving detection accuracy and reducing false positives over time. This addresses the dynamic nature of APTs and improves model adaptability.

Human-AI Collaboration: Emphasize the synergistic combination of machine speed and human expertise.

Realistic Emulation: Use a realistic multi-stage APT emulation lab (e.g., Caldera (MITRE), Atomic Red Team, custom scripts) to simulate the full APT lifecycle, addressing the limitations of static datasets.

Comprehensive Testing: Test across different attack phases: initial access, lateral movement, privilege escalation, and exfiltration.

Benchmarking: Benchmark against baseline methods (e.g., SIEM rules, classical ML models) and evaluate against relevant metrics.

Metrics: Beyond traditional Detection Rate (DR), False

***Corresponding Author:** Girish Kumar B. C

Positive Rate (FPR), and Mean Time to Detect (MTTD), include a Model Explainability Score (quantified via human analyst study) to assess practical utility and trust.

5. Robust Evaluation Strategy:

- This proposed framework integrates supervised, self-supervised, and reinforcement learning, utilizes comprehensive data sources, prioritizes real-time operation, and crucially incorporates explainability and human feedback, aiming to create a truly lightweight, adaptive, and explainable APT detection system.
- **Realistic Emulation**: Use a realistic multi-stage APT emulation lab (e.g., Caldera (MITRE), Atomic Red Team, custom scripts) to simulate the full APT lifecycle, addressing the limitations of static datasets.
- **Comprehensive Testing:** Test across different attack phases: initial access, lateral movement, privilege escalation, and exfiltration.
- **Benchmarking**: Benchmark against baseline methods (e.g., SIEM rules, classical ML models) and evaluate against relevant metrics.
- **Metrics**: Beyond traditional Detection Rate (DR), False Positive Rate (FPR), and Mean Time to Detect (MTTD), include a Model Explainability Score (quantified via human analyst study) to assess practical utility and trust.

This proposed framework integrates supervised, self-supervised, and reinforcement learning, utilizes comprehensive data sources, prioritizes real-time operation, and crucially incorporates explainability and human feedback, aiming to create a truly lightweight, adaptive, and explainable APT detection system.

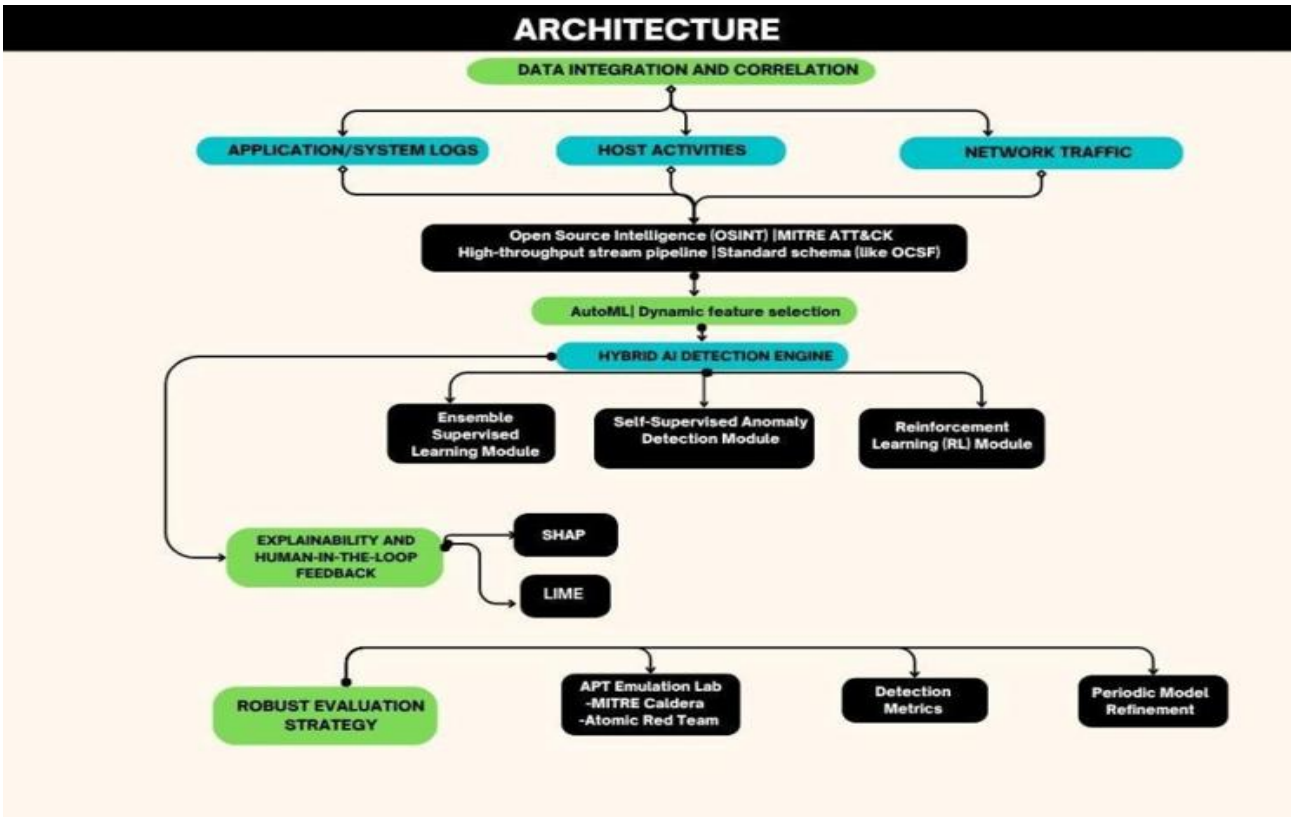

Figure 1: Architecture of Proposed Framework – Gives the flow of the proposed method.

This framework is a multi-layered, AI-powered system designed to detect Advanced Persistent Threats (APTs) in real-time. It focuses on combining supervised learning (for known threats), self-supervised anomaly detection (for unknown threats), and reinforcement learning (to prioritize and adapt responses). It also emphasizes explainability and human-in-the-loop feedback for continuous improvement and analyst trust.

1. Multi-Source Real-Time Data Ingestion & Correlation

What it does: Continuously gathers and correlates data from:

- **Network traffic** (like NetFlow, packets),
- **Host activities** (logins, file changes),
- **Application/System logs** (e.g., SIEM, cloud logs).

Key additions:

- Enrich data using Open Source Intelligence (OSINT) and MITRE ATT&CK tactics.

Use a high-throughput stream pipeline for live analysis. Normalize everything into a standard schema (like OCSF) to ensure consistent processing across tools.

***Corresponding Author:** Girish Kumar B. C

2. Adaptive Data Preprocessing & Feature Engineering
What it does: Cleans, transforms, and selects relevant features from incoming data. Key mechanisms: Uses AutoML tools for automated feature generation.

- Applies dynamic feature selection to adapt to evolving attacks (concept drift).
- Cleans outliers, missing values, and normalizes features to maintain high data quality.

3. Hybrid AI Detection Engine
This is the core engine, consisting of three submodules:

Ensemble Supervised Learning Module
- **Purpose:** Detect known APT attacks from historical data.
- **How:** Trains ensemble models like Random Forest, XGBoost on labeled APT datasets.
- **Challenge addressed:** Deals with rare attack data using imbalance handling techniques.

Self-Supervised Anomaly Detection Module
Purpose: Identify unknown or zero-day threats. How: Uses models like Deep InfoMax or Autoencoders to learn patterns of normall behavior from unlabeled data.

Flags anomalies using unsupervised models like Isolation Forest or One-Class SVM.

Reinforcement Learning (RL) Module
Purpose: Improve response effectiveness and alert prioritization. How: Uses RL agents like PPO to: Learn which alerts are valid. Adjust severity scores.

Suggest or even trigger automated responses (e.g., containment, blocking).

4. Explainability and Human-in-the-Loop Feedback
Explainability: Use SHAP and LIME to explain alerts in hu-man-readable form. Analyst Feedback Loop: Human analysts validate high-uncertainty alerts.

These validations are used to retrain models periodical-ly, increasing accuracy and reducing false alarms.

Emphasis: This ensures the system is transparent, adaptive, and collaborative between machine and human intelligence.

5. Robust Evaluation Strategy
APT Emulation Lab: Use tools like MITRE Caldera or Atom-ic Red Team to simulate real APTs.
Attack Phases Tested: Initial access, lateral movement, privi-lege escalation, and data exfiltration.
Benchmarking: Compare the framework with traditional sys-tems (e.g., SIEM).

Metrics Used: Detection Rate (DR) False Positive Rate (FPR)

Mean Time to Detect (MTTD) **Explainability Score** (evaluated by analysts

CONCLUSION:

Advanced Persistent Threats (APTs) are one of the toughest problems in cybersecurity today. Unlike traditional cyberattacks, APTs don't just break in and cause quick damage—they quietly sneak into systems, stay hidden for long periods, and slowly steal sensitive information. This makes them especially hard to detect using old-school security methods like firewalls, signature-based tools, or rule-based monitoring.

This paper introduced a smart, modern solution to that problem: an adaptive and explainable AI-powered framework that can detect and respond to APTs in real-time. Instead of relying on fixed rules or known attack signatures, the system learns from data—both past at-tacks and normal system behavior—to spot anything suspicious, whether it's a known threat or something new.

The framework uses a combination of advanced AI techniques. It in-cludes supervised learning to catch known attack patterns, self-supervised learning to understand what ―normall looks like (and flag anything unusual), and reinforcement learning to help the system im-prove over time—by learning from its mistakes and adjusting how it reacts.

What makes this system even more powerful is that it looks at data from many sources at once—like network traffic, system logs, user ac-tivity, and cloud logs—giving it a complete picture of what's happen-ing. It processes all this data live, so it can catch threats as they happen, not after the damage is done.

A key part of this approach is its focus on explainability and human feedback. AI can be a black box, but this framework uses tools like SHAP and LIME to explain why a certain alert was triggered, so hu-man analysts can understand and trust the system's decisions. It also includes a feedback loop where analysts can mark alerts as real or false, which helps the AI learn and improve.

To test the system, the framework was evaluated using realistic APT simulations. It performed strongly across all important metrics—like detection speed, accuracy, and false positive rate—and even included a measure of how understandable the system's alerts were to real users.

In short, this framework shows how AI and human expertise can work together to tackle one of cybersecurity's biggest threats. It's fast, smart, adaptable, and transparent—everything a modern security system needs to face today's sophisticated attackers.

***Corresponding Author:** Girish Kumar B. C

REFERENCES

1. Olubudo, P. (2024). *Advanced threat detection techniques using machine learning: Exploring the use of AI and ML in identifying and mitigating advanced persistent threats (APTs).* ResearchGate. https://www.researchgate.net/publication/380743475
2. Paul, S., Kumar, R., & Mukherjee, N. (2020). Introducing machine learning algorithm for advanced persistent threat detection. In *Procedia Computer Science* (Vol. 167, pp. 2332–2341). Elsevier. https://doi.org/10.1016/j.procs.2020.03.287
3. Malik, V., Khanna, A., Sharma, N., & Nalluri, S. (2024). *Advanced persistent threats (APTs): Detection techniques and mitigation strategies* (EasyChair Preprint No. 14646). EasyChair. https://easychair.org/publications/preprint/CbnF
4. Nourian, A., & Camp, L. J. (2021). Toward the use of artificial intelligence (AI) for advanced persistent threat detection. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS 2021)* (pp. 1–10). https://doi.org/10.24251/HICSS.2021.932
5. Chen, W., He, Y., Liu, X., He, F., & Chen, C. (2021). A belief rule-based expert system for the detection of advanced persistent threats. *Applied Sciences, 11*(21), 9899. https://doi.org/10.3390/app11219899
6. Bashir, I. (2018). *Mastering blockchain* (2nd ed.). Packt Publishing.

***Corresponding Author:** Girish Kumar B. C