IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Artificial Intelligence for Real-Time Cybersecurity Threat Detection and Prevention

Mrs. Meera Sawalkar[1], Vedanti Marne[2], Mansi Madgule[3], Aayush Zende[4]

Assistant Professor[1], Student[2], Student[3], Student[4]

Artificial Intelligence and Data Science

All India Shri Shivaji Memorial Society's - Institute of Information Technology, Pune, India

Abstract: Rapid digitization across industries has made cybersecurity a paramount concern for organizations worldwide. Modern computing environments generate enormous volumes of data including authentication logs, network flows, application events, and endpoint telemetry — volumes that far exceed human capacity for manual inspection. Artificial Intelligence (AI) has emerged as a transformative force in cyber defense, enabling systems to learn behavioral patterns, identify anomalies, classify threats, and facilitate faster incident response. This paper presents a comprehensive examination of AI-driven approaches for real-time threat detection and prevention. Core topics include intrusion detection, malware analysis, phishing identification, user behavior analytics, and automated security operations. Classical approaches such as Random Forest and Support Vector Machine are examined alongside deep learning architectures including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Autoencoders, and Transformer models. The paper additionally addresses persistent challenges including data quality issues, adversarial robustness, model drift, explainability demands, and infrastructure constraints. Findings indicate that while AI substantially enhances detection capabilities, its effective deployment depends on robust data pipelines, continuous monitoring, and collaborative human-AI workflows.

Key Words: Artificial Intelligence, Cybersecurity, Threat Detection, Intrusion Detection Systems, Malware Classification, Phishing Detection, Deep Learning, Anomaly Detection, Security Analytics, Behavior Analytics

1. INTRODUCTION

Contemporary organizations operate within an increasingly hostile digital landscape. Financial services, healthcare providers, government agencies, educational institutions, and critical infrastructure all depend on networked systems that are constantly exposed to adversarial activity. The threat spectrum includes ransomware campaigns, targeted phishing attacks, network intrusions, credential theft, insider misuse, advanced persistent threats (APTs), and distributed denial-of-service (DDoS) attacks. A defining characteristic of modern adversaries is adaptability — attackers continuously modify their techniques, obfuscate malicious code, and exploit newly discovered vulnerabilities to evade existing defenses.

Traditional security mechanisms based on static signatures and predefined rules remain useful for detecting known threats but are demonstrably insufficient against novel attack patterns, zero-day exploits, and behavior-based intrusions. The sheer volume of security events generated by enterprise environments — often reaching millions of log entries per day — makes manual triage operationally untenable. Security Operations Centers (SOCs) face persistent challenges including alert fatigue, delayed response times, and skill shortages.

AI-driven cybersecurity systems address these limitations by learning statistical patterns from historical data and applying predictive models to incoming event streams. Machine learning enables threat classifiers to generalize beyond seen examples, while deep learning facilitates end-to-end feature extraction from complex, high-dimensional data. The integration of AI into security workflows has progressed from experimental research to large-scale production deployment across multiple security domains.

This paper provides a structured technical survey of AI applications in real-time cybersecurity. Section 2 establishes foundational concepts. Section 3 reviews prior research. Section 4 covers major application domains. Section 5 describes a proposed detection architecture. Section 6 discusses datasets and metrics. Section 7 compares algorithms. Section 8 analyzes performance considerations. Section 9 identifies key challenges. Section 10 outlines future directions. Section 11 concludes the paper.

2. BACKGROUND AND RELATED CONCEPTS

2.1 The Need for Intelligent Defense

The evolving sophistication of cyber threats has exposed fundamental limitations in rule-based security paradigms. Sophisticated adversaries leverage living-off-the-land techniques, encrypted channels, and multi-stage attack chains that may unfold over extended periods. Detecting such behaviors requires correlating signals across heterogeneous data sources — network telemetry, endpoint logs, DNS queries, authentication events, and email metadata — a task that naturally benefits from machine learning-based correlation and pattern recognition.

2.2 Machine Learning Fundamentals

Machine learning constructs predictive models by optimizing parameters over labeled or unlabeled training data. Supervised learning is applicable when labeled datasets of normal and malicious activity are available. Unsupervised learning discovers latent structure in unlabeled data and supports anomaly detection without requiring exhaustive threat catalogs. Semi-supervised approaches leverage small labeled sets alongside abundant unlabeled data, well-suited to cybersecurity contexts where labeling is costly. Reinforcement learning, while less common in detection tasks, has applications in adaptive security policy optimization.

2.3 Deep Learning Architectures

Deep neural networks extend classical machine learning through multi-layered hierarchical representations. Convolutional Neural Networks (CNNs) extract spatially local features and are applied to malware visualization and packet-level analysis. Recurrent architectures, particularly Long Short-Term Memory (LSTM) networks, model temporal dependencies in sequential event streams such as API call traces and user activity logs. Transformer architectures, pre-trained on large corpora, deliver state-of-the-art performance on text-centric security tasks including phishing detection and security document classification. Autoencoders learn compact normal representations and identify anomalies through elevated reconstruction error.

2.4 Threat Detection vs. Prevention

Threat detection involves identifying potentially malicious activity within observed data. Threat prevention extends detection by coupling model inference to automated countermeasures such as network isolation, access revocation, or traffic blocking. Modern security platforms increasingly combine both functions, enabling low-latency automated responses that limit attacker dwell time and reduce blast radius before human analysts can intervene.

3. LITERATURE REVIEW

Research applying machine learning to cybersecurity has evolved considerably over the past two decades. Early studies evaluated classical supervised algorithms — Decision Trees, Naive Bayes, Support Vector Machines, and Random Forest — on benchmark datasets such as KDD Cup 1999 and NSL-KDD. These investigations established that learned classifiers outperform signature-based baselines on traffic containing novel attack variations, though dataset realism was a recognized limitation.

Subsequent work adopted more representative benchmarks including UNSW-NB15 and CICIDS2017, which incorporate diverse contemporary attack categories and realistic enterprise traffic characteristics.

Studies employing gradient-boosted ensembles and LSTM-based sequence models on these datasets reported strong multi-class classification performance, while also highlighting that no single architecture dominates across all threat categories.

Malware analysis research progressed from hash-based detection toward behavioral and structural analysis. A notable development involved encoding binary executables as grayscale images and applying CNN classifiers, enabling recognition of malware families based on visual structural similarity even after code-level obfuscation. Dynamic analysis methods tracking system call sequences with RNN models further enhanced detection of evasive samples.

Phishing detection evolved from simple blacklist lookups to feature-engineered machine learning pipelines incorporating URL lexical analysis, domain registration metadata, and email header attributes. The adoption of Transformer-based language models, particularly BERT and its derivatives, markedly improved detection of contextually sophisticated spear-phishing content that evades keyword-based filters.

Behavioral analytics research established frameworks for modeling user and entity activity baselines using Autoencoders and Isolation Forests, enabling detection of insider threats and compromised accounts that lack conventional malware signatures. A parallel research stream has examined adversarial robustness, demonstrating that cybersecurity models are vulnerable to carefully crafted adversarial inputs, data poisoning attacks, and model extraction attempts — considerations critical for production security deployments.

4. APPLICATIONS OF AI IN CYBERSECURITY

4.1 Intrusion Detection and Prevention

Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS) represent the most extensively studied application of AI in network security. Contemporary AI-powered IDS analyze bidirectional network flows, protocol metadata, and packet payloads to classify traffic as benign or malicious across categories including DDoS, port scanning, brute-force authentication attacks, command-and-control communication, and lateral movement. Random Forest and gradient-boosted ensembles deliver competitive performance on structured flow features, while LSTM models capture temporal dependencies in event sequences. Production deployments couple detection outputs to automated firewall rules, rate limiting policies, and endpoint isolation actions.

4.2 Malware Detection and Classification

AI-based malware analysis encompasses static examination of binary structure and dynamic profiling of runtime behavior. Static methods extract features from imported libraries, byte n-gram distributions, section entropy, and control flow graphs without executing the sample — providing rapid classification at low computational cost. Dynamic methods monitor system call sequences, registry modifications, network connections, and memory patterns during controlled execution, providing richer behavioral signatures at greater cost. Hybrid pipelines combining static pre-filtering with dynamic analysis of borderline samples optimize the coverage-efficiency tradeoff.

4.3 Phishing Email and URL Detection

Phishing attacks exploit social engineering to elicit credential disclosure or malware installation. AI enhances detection by analyzing lexical URL features, domain age and reputation signals, email authentication headers, sender behavioral history, and full message content. Transformer models pre-trained on large text corpora distinguish manipulation intent and contextual incongruity in message content with greater accuracy than token-frequency approaches, particularly against sophisticated targeted attacks that employ plausible cover stories and domain spoofing.

4.4 User and Entity Behavior Analytics

User and Entity Behavior Analytics (UEBA) systems establish statistical baselines for individual users, service accounts, and infrastructure components, then generate risk scores for observed deviations. Relevant features include access time distributions, geographic login patterns, data transfer volumes, privilege usage, and application interaction sequences. UEBA is particularly effective for detecting compromised accounts and malicious insiders whose activities lack conventional malware indicators, as threat actors using legitimate credentials may nonetheless exhibit statistically anomalous behavioral patterns.

4.5 Security Operations Center Automation

SOC teams contend with thousands of daily alerts from heterogeneous security controls. AI-assisted triage applies multi-class classifiers to prioritize alerts by predicted severity, correlates related events into coherent incident timelines, suppresses known false positive patterns, and recommends investigative actions based on historical analyst decisions. Natural language models support automated enrichment by extracting threat intelligence from unstructured reports and linking indicators to active incidents.

4.6 Fraud and Identity Abuse Detection

Fraud detection shares methodological foundations with cybersecurity threat detection, applying supervised and anomaly-based models to transactional data, device fingerprints, login patterns, and geographic signals. Identity abuse detection in enterprise environments monitors for credential sharing, privilege escalation attempts, and dormant account activation — indicators that may precede data exfiltration or sabotage.

5. METHODOLOGY AND PROPOSED ARCHITECTURE

The proposed framework implements a six-stage layered pipeline for real-time AI-driven threat detection, designed to accommodate diverse data sources while remaining adaptable to organizational deployment constraints.

5.1 Stage 1: Data Collection

The pipeline ingests telemetry from network flow exporters, host-based agents, email gateways, authentication systems, DNS resolvers, cloud service APIs, and web proxy logs. Multi-source collection is essential because sophisticated attack chains generate correlated signals across multiple infrastructure layers that are individually insufficient for reliable classification.

5.2 Stage 2: Preprocessing

Raw security telemetry requires normalization to support consistent model inference. The preprocessing pipeline performs deduplication, missing value imputation, timestamp normalization, categorical variable encoding, and class rebalancing using SMOTE or random undersampling. Text tokenization prepares email and log content for NLP-based models, while sliding window aggregation constructs fixed-length feature vectors for sequential models.

5.3 Stage 3: Feature Engineering

Structured data domains benefit from domain-informed features. Network intrusion detection leverages connection duration, protocol distribution, inter-packet timing, byte asymmetry, and error flag counts. Phishing detection incorporates URL entropy, domain age, redirect depth, and linguistic urgency scores. Deep learning components perform representation learning directly from preprocessed inputs, reducing reliance on manual feature design for unstructured data types.

5.4 Stage 4: Model Training

Model selection follows task requirements as summarized in Table 1. Training employs stratified temporal splits to prevent data leakage from future events into training sets. Hyperparameter optimization uses cross-validated grid or Bayesian search. Ensemble methods combine complementary base learners to improve robustness across attack category distributions.

Table 1: Model Selection by Cybersecurity Task

Task	Typical Models	Data Type
Intrusion Detection	RF, XGBoost, LSTM	Tabular / Sequential
Malware Classification	CNN, RF, Autoencoder	Binary / Image / Behavior

Phishing Detection	BERT, Transformer, LR	Text / URL
Behavior Analytics	Autoencoder, Isolation Forest	Temporal / Anomaly
Alert Prioritization	Gradient Boosting, Rankers	Mixed Features

5.5 Stage 5: Real-Time Inference

Deployed models process streaming event data through the same preprocessing and feature pipeline used during training. Model outputs are post-processed by a threshold layer that combines probabilistic scores with deterministic rule conditions, reducing false positive rates for high-confidence benign patterns while maintaining sensitivity to novel threat signatures.

5.6 Stage 6: Response and Feedback

High-confidence threat detections trigger automated response actions including endpoint isolation, firewall rule insertion, email quarantine, and session termination. Analyst verdicts on escalated alerts are logged as feedback, enabling periodic supervised retraining that incorporates emerging threat patterns and corrects systematic classification errors. The complete pipeline is summarized as: Data Ingestion → Preprocessing → Feature Extraction → AI Inference → Risk Scoring → Automated Response → Analyst Feedback → Model Retraining.

6. DATASETS AND EVALUATION METRICS

6.1 Benchmark Datasets

Standardized datasets enable reproducible comparison across research efforts. Table 2 summarizes commonly used cybersecurity benchmarks. NSL-KDD improved upon the original KDD Cup 1999 dataset by removing duplicate records, though its traffic characteristics no longer reflect contemporary network environments. UNSW-NB15 introduced greater attack category diversity and more realistic feature distributions. CICIDS2017 simulates enterprise network traffic with labeled attack scenarios spanning multiple days. EMBER provides a large-scale corpus for static malware classification. Email corpora and URL datasets support phishing and web threat research.

Table 2: Commonly Used Cybersecurity Datasets

Dataset	Domain	Primary Application
NSL-KDD	Network intrusion	IDS benchmarking
UNSW-NB15	Network traffic	Multi-class attack classification
CICIDS2017	Enterprise traffic	Realistic intrusion studies
EMBER	PE malware files	Static malware detection
Email corpora	Email text/metadata	Phishing research

6.2 Evaluation Metrics

Accuracy alone is an unreliable indicator in cybersecurity contexts due to severe class imbalance between benign and malicious events. A classifier that predicts all events as benign may achieve high accuracy while providing no operational value. The standard evaluation suite includes:

- Precision = TP / (TP + FP): proportion of flagged events that are genuinely malicious

- Recall = TP / (TP + FN): proportion of actual attacks that are detected
- F1-Score = 2 × (Precision × Recall) / (Precision + Recall): harmonic mean balancing precision and recall
- AUC-ROC: discrimination ability across detection thresholds
- False Positive Rate and False Negative Rate: operational cost indicators
- Detection Latency: time from event occurrence to alert generation

6.3 Operational Evaluation

Research metric performance must be complemented by operational evaluation covering alert volume per analyst shift, mean time to detect, mean time to respond, retraining frequency, and system stability under peak load. These criteria determine whether laboratory performance translates to practical security improvement.

7. ALGORITHMIC COMPARISON

The selection of an appropriate AI algorithm involves balancing detection performance, computational requirements, interpretability, and operational constraints. Table 3 provides a structured comparison of major approaches.

7.1 Random Forest

Random Forest aggregates predictions from an ensemble of decision trees trained on random feature subsets, producing robust classifiers with inherent feature importance estimates. Its ensemble structure provides natural resistance to overfitting and delivers competitive performance on structured network flow features. Interpretability through feature importance rankings supports analyst understanding of model decisions, an important property in regulated environments.

7.2 Support Vector Machine

SVMs identify maximum-margin hyperplanes that separate classes in high-dimensional feature spaces. They are effective on well-defined classification problems with moderate dataset sizes and carefully engineered features. Computational complexity scales poorly to very large datasets, and kernel selection significantly influences performance. SVMs have been applied to network intrusion classification and malware detection with competitive results on established benchmarks.

7.3 CNN and LSTM Networks

CNNs extract local spatial features through learned convolutional filters and have been successfully applied to malware visualization, packet payload analysis, and network traffic classification. LSTM networks address the vanishing gradient problem in standard RNNs through gating mechanisms that selectively retain relevant historical context, making them effective for modeling temporal dependencies in event logs, API call sequences, and user activity timelines.

7.4 Autoencoder

Autoencoders learn compressed latent representations of normal data distributions through unsupervised training. During inference, samples that cannot be accurately reconstructed are flagged as anomalous. This approach is particularly valuable in cybersecurity contexts where labeled attack data is scarce or unavailable, enabling zero-shot detection of novel threat patterns based solely on their distributional distance from the learned normal baseline.

7.5 Transformer Models

Transformer architectures leverage self-attention mechanisms to model long-range contextual dependencies in sequential data. Pre-trained models such as BERT achieve state-of-the-art performance on phishing detection, security document classification, and threat intelligence extraction tasks. Their primary limitation for real-time deployment is computational cost; large models require GPU inference infrastructure and may introduce latency incompatible with streaming detection requirements.

Table 3: Comparison of AI Approaches in Cybersecurity

Approach	Key Advantages	Key Limitations

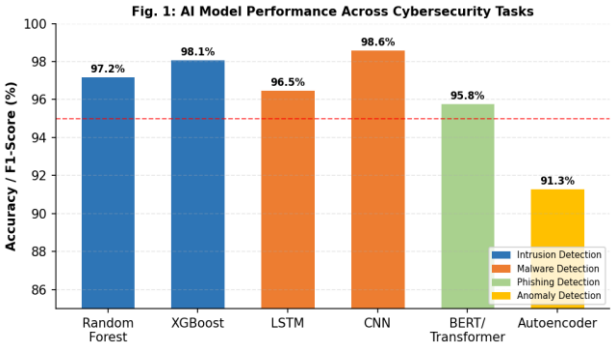
Rule-Based Systems	Fast, deterministic, explainable	Cannot detect novel attacks
Random Forest	Robust, interpretable, fast training	Limited on sequential/text data
SVM	Strong on clean structured data	Slow on large datasets
CNN	Automatic spatial feature learning	Requires substantial training data
LSTM	Models temporal dependencies	Slow training, complex tuning
Autoencoder	Effective with minimal labels	Threshold sensitivity, false alarms
Transformer	Excellent contextual understanding	High compute cost
Hybrid Systems	Balances complementary strengths	Complex design and maintenance

8. RESULTS, ANALYSIS AND DISCUSSION

Reported benchmark performance across the AI cybersecurity literature is generally strong. CNN-based malware classifiers consistently achieve accuracy above 98% on well-curated static analysis datasets. Ensemble and LSTM-based intrusion detection models report accuracy in the 94-99% range on NSL-KDD and CICIDS2017 benchmarks. Transformer-based phishing detection models achieve F1-scores exceeding 95% on held-out test sets. Table 4 summarizes representative performance trends.

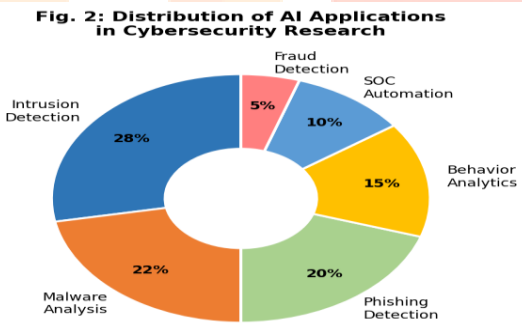
Table 4: Representative AI Cybersecurity Performance

Application	Model Type	Reported Performance
Malware Detection	CNN / Transfer Learning	Accuracy > 98%
Intrusion Detection	RF / XGBoost / LSTM	Accuracy 94-99%
Phishing Detection	BERT / NLP Models	F1-Score > 95%
Behavior Analytics	Autoencoder / LSTM	Strong recall; threshold-dependent precision
Alert Prioritization	Ensemble / Ranking	Reduced manual triage burden



These benchmark figures must be interpreted in context. Several systematic factors cause production performance to differ from controlled experimental results: (1) Distribution shift occurs as network traffic characteristics, user behavior, and attacker techniques evolve after model training. (2) Class imbalance in live environments is typically more extreme than in curated datasets, inflating false positive rates at fixed thresholds. (3) Adversarial adaptation means that once detection logic becomes known, sophisticated adversaries can modify their techniques to evade specific model weaknesses. (4) Infrastructure constraints may necessitate model compression or quantization that degrades accuracy relative to full-precision research models.

Human-AI collaboration is a critical operational factor. Models that provide explanatory signals — flagging the specific features that contributed to a high-risk score — enable analysts to make faster, more confident triage decisions. Explainability also supports regulatory compliance in environments subject to data protection or critical infrastructure regulations. The practical value of a security AI system should therefore be evaluated not only on detection rates but on its ability to support effective analyst workflows.



9. CHALLENGES AND LIMITATIONS

9.1 Data Quality and Label Scarcity

Security training data frequently contains incomplete records, incorrect labels, and class distributions that inadequately represent real-world event frequencies. Ground truth labels for malicious activity require skilled analyst effort to produce and are often unavailable for novel threat categories. Models trained on synthetic or outdated data may generalize poorly to production environments.

9.2 Adversarial Robustness

Cybersecurity uniquely involves adversarial dynamics absent from most ML application domains. Attackers with knowledge of or access to a deployed model can craft adversarial inputs designed to bypass detection, poison training data to degrade model performance, or probe detection boundaries through targeted experimentation. Robust model design, input validation, and ensemble diversity are necessary mitigations.

9.3 Concept Drift

Statistical properties of both normal and malicious traffic evolve continuously due to software updates, infrastructure changes, organizational behavioral shifts, and attacker adaptation. Models trained on historical data experience performance degradation over time without periodic retraining. Drift detection mechanisms and automated retraining pipelines are essential components of production security AI systems.

9.4 Explainability Requirements

High-stakes security decisions — suspending user accounts, blocking critical services, or isolating network segments — require justification that deep learning models often cannot provide in human-interpretable form. Post-hoc explanation methods such as SHAP and LIME partially address this limitation but introduce additional computational overhead and may not faithfully represent model decision logic.

9.5 Privacy and Regulatory Compliance

Security monitoring necessarily processes personal data including communication content, authentication records, and device identifiers. Compliance with data protection regulations requires minimizing data collection scope, implementing retention limits, applying appropriate access controls, and documenting lawful processing bases. These requirements constrain dataset construction, model training, and real-time inference pipelines.

9.6 Infrastructure and Integration Cost

Large-scale AI security systems require substantial computational infrastructure, including GPU-accelerated inference servers, distributed log aggregation pipelines, low-latency data streaming platforms, and MLOps tooling for model lifecycle management. Integration with existing SIEM platforms, incident response workflows, and analyst toolchains introduces additional engineering complexity. These costs create adoption barriers for resource-constrained organizations.

10. FUTURE SCOPE

10.1 Federated Learning

Federated learning enables collaborative model training across organizational boundaries without centralizing raw data. Each participating organization trains on local data and contributes model updates to a shared global model. This architecture is particularly valuable in cybersecurity because it allows broader threat intelligence sharing while respecting data sensitivity and regulatory constraints.

10.2 Graph Neural Networks

Many sophisticated attack campaigns involve coordinated activity across multiple entities — users, devices, processes, and network nodes — connected by relationships that graph-structured representations can naturally encode. Graph Neural Networks applied to attack graph analysis and entity relationship modeling can detect lateral movement, privilege escalation chains, and coordinated threat actor activity that point-in-time classifiers may miss.

10.3 Explainable AI Integration

Advancing explainability research toward natively interpretable security models — rather than relying on post-hoc approximations — will increase analyst adoption, support regulatory compliance, and enable more principled model debugging. Future security AI systems will likely combine high-performance detection models with structured explanation outputs that align with analyst mental models.

10.4 Edge AI for Distributed Security

The proliferation of IoT devices, operational technology, and distributed computing environments creates security monitoring challenges that centralized cloud architectures cannot efficiently address. Lightweight quantized models deployable on edge hardware support local threat detection with minimal latency, reduced bandwidth consumption, and continued operation during connectivity interruptions.

10.5 AI-Augmented SOC Operations

Next-generation SOC platforms will leverage large language models and multimodal AI systems to accelerate incident investigation through automated timeline reconstruction, natural language querying of security logs, intelligent playbook recommendation, and autonomous execution of pre-approved response actions. Human analysts will focus increasingly on high-judgment decisions while AI systems handle structured analysis tasks.

11. CONCLUSION

This paper has examined the role of Artificial Intelligence in addressing the scale and adaptability challenges that define contemporary cybersecurity operations. The survey covered key application domains — intrusion detection, malware analysis,

phishing identification, user behavior analytics, and SOC automation — and evaluated the algorithmic landscape spanning classical ensemble methods, deep learning architectures, and large language models.

The evidence indicates that AI substantially enhances threat detection coverage, response speed, and analyst productivity relative to static rule-based approaches. However, sustained operational effectiveness depends on robust data pipelines, continuous monitoring for
concept drift, adversarial robustness measures, and transparent human-AI collaboration frameworks. High benchmark accuracy is a necessary but insufficient criterion for successful production deployment.

Future progress will be driven by federated learning for privacy-preserving intelligence sharing, graph-based threat modeling, natively interpretable architectures, and AI-augmented analyst workflows. As both AI capabilities and attacker sophistication continue to advance, the design of responsible, maintainable, and explainable security AI systems will remain a critical research and engineering priority.

ACKNOWLEDGEMENT

The authors express sincere gratitude to the Department of Artificial Intelligence and Data Science, AISSMS Institute of Information Technology, Pune, for providing the academic environment and research support that facilitated this work. The authors also acknowledge the broader cybersecurity research community for establishing the open datasets and evaluation benchmarks referenced throughout this paper.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
[2] R. Vinayakumar et al., "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525-41550, 2019.
[3] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in Proc. ICLR, 2015.
[4] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in Proc. MilCIS, 2015.
[5] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset," in Proc. ICISSP, 2018.
[6] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A detailed analysis of the KDD CUP 99 dataset," in Proc. IEEE CISDA, 2009.
[7] H. S. Anderson and P. Roth, "EMBER: An open dataset for training static PE malware machine learning models," arXiv:1804.04637, 2018.
[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019.
[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, no. 5786, pp. 504-507, 2006.
[10] A. Tuor et al., "Deep learning for unsupervised insider threat detection in structured cybersecurity data streams," in Proc. AAAI Workshops, 2017.
[11] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, pp. 1-19, 2019.
[12] D. Gunning and D. Aha, "DARPA's explainable artificial intelligence program," AI Magazine, vol. 40, no. 2, pp. 44-58, 2019.
[13] S. Marchal, K. Saari, N. Singh, and N. Asokan, "Know your phish: Novel techniques for detecting phishing sites," in Proc. IEEE ICDCS, 2016.
[14] S. Garcia, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," Computers & Security, vol. 45, pp. 100-123, 2014.
[15] M. Abdel-Basset et al., "A cyber security framework to identify malicious edge device in fog computing and IoT," Computer Networks, vol. 167, p. 106970, 2020.