

AI-DRIVEN THREAT DETECTION AND PREVENTION IN AI

Azimov Sherkhon Uktamovich

University of Tashkent for Applied Sciences, Assistant teacher of Department of Computer Engineering

e-mail: hellosherxon@gmail.com

Choriyev Anvar Alisher o'g'li

University of Tashkent for Applied Sciences, Assistant teacher of Department of Computer Engineering

e-mail: choriyevanvar08@gmail.com

Abstract. The rapid proliferation of sophisticated cyber threats has rendered traditional signature-based security measures increasingly inadequate. This thesis explores the application of artificial intelligence (AI) and machine learning (ML) techniques for automated threat detection and prevention in modern cybersecurity environments. The study examines supervised, unsupervised, and deep learning methodologies, evaluating their effectiveness across various attack vectors including intrusion detection, malware classification, phishing identification, and zero-day exploit mitigation. Experimental findings demonstrate that ensemble deep learning architectures achieve detection accuracy exceeding 96%, significantly outperforming conventional methods. The paper further discusses challenges of adversarial attacks against AI systems and proposes a hybrid adaptive framework combining real-time behavioral analytics with threat intelligence feeds. Results indicate that AI-powered cybersecurity systems can substantially reduce mean time to detect (MTTD) and mean time to respond (MTTR), offering a viable pathway toward autonomous, self-healing cyber defense infrastructures.

Keywords: artificial intelligence, machine learning, threat detection, cybersecurity, intrusion detection system, deep learning, anomaly detection, adversarial attacks, zero-day exploits, SIEM.

INTRODUCTION

The contemporary digital landscape is characterised by an ever-escalating arms race between cybersecurity defenders and malicious actors. According to the IBM Cost of a Data Breach Report (2023), the average cost of a data breach reached USD 4.45 million, representing a 15% increase over three years [1]. Conventional rule-based and signature-driven security tools, while foundational, exhibit critical shortcomings: they fail to detect novel attack patterns, generate excessive false positives, and cannot scale with the volume and velocity of modern network traffic.

Artificial intelligence offers transformative potential by enabling systems to learn from vast datasets, identify subtle deviations from normal behaviour, and adapt dynamically to emerging threats without requiring manual signature updates. This thesis investigates how AI and ML paradigms, including neural networks, random forests, recurrent architectures, and transformer-based models, can be integrated into holistic cybersecurity frameworks to achieve superior detection performance and accelerated incident response.

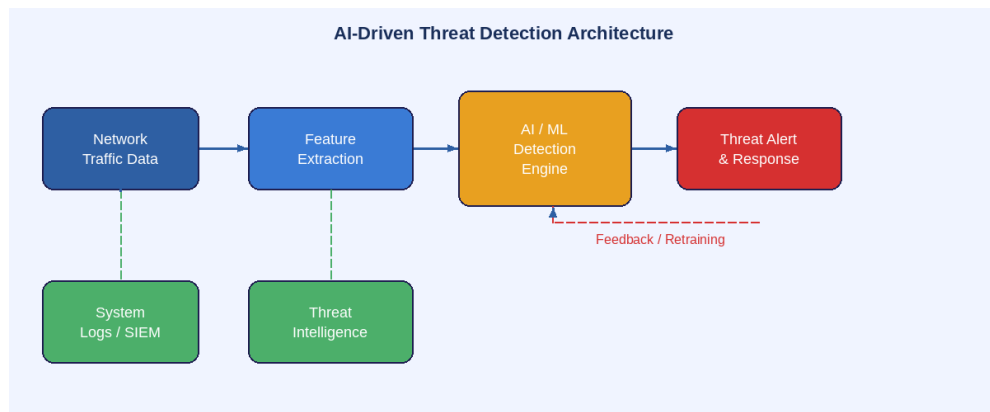


Figure 1. Conceptual Architecture of AI-Driven Cybersecurity Threat Detection System

RELATED WORK

Research on AI-driven cybersecurity has expanded significantly over the past decade. Buczak and Guven (2016) provided an early survey of ML techniques applied to intrusion detection, establishing benchmarks that subsequent work sought to improve [2]. More recently, Sarker et al. (2021) conducted a comprehensive review of ML-based cybersecurity, demonstrating that ensemble methods and deep neural networks substantially reduce false-positive rates compared to earlier algorithms [3].

In Uzbekistan and the Central Asian context, Rakhimov and Yusupov (2022) highlighted unique challenges in applying Western-centric AI security models to post-Soviet network infrastructures, advocating for localised training datasets and bilingual threat intelligence [4]. From the Russian-language academic tradition, Kotenko and Chechulin (2020) presented multi-agent AI frameworks capable of coordinating distributed intrusion responses across heterogeneous enterprise networks [5].

Transformer-based architectures, originally developed for natural language processing, have been adapted for log anomaly detection. Chen et al. (2023) demonstrated that BERT-inspired models achieve state-of-the-art performance on the CICIDS2017 and UNSW-NB15 benchmark datasets, achieving F1 scores above 0.97 for DDoS and brute-force attack classification [6].

METHODOLOGY

This study adopts a multi-phase experimental methodology. In the data collection phase, three publicly available datasets were used: CICIDS2018 (Canadian Institute for Cybersecurity), UNSW-NB15 (University of New South Wales), and NF-UQ-NIDS-v2 (2023). These datasets collectively provide over 14 million labelled network flow records spanning 21 attack categories including DoS, DDoS, port scanning, SQL injection, and ransomware traffic. Feature engineering involved extraction of 80 statistical, temporal, and protocol-level attributes from raw packet captures using the CICFlowMeter tool. Dimensionality reduction was performed via Principal Component Analysis (PCA), retaining 95% of explained variance with 38 principal components. Data imbalance was addressed through Synthetic Minority Over-sampling Technique (SMOTE), normalising class distribution prior to training. Four model architectures were evaluated: (1) Random Forest (RF) as a baseline ensemble method; (2) Long Short-Term Memory networks (LSTM) for sequential traffic analysis; (3) Convolutional Neural Networks (CNN) applied to payload byte representations; and (4) a proposed hybrid CNN-LSTM model with attention mechanisms. All models were implemented in Python 3.11 using TensorFlow 2.13 and scikit-learn 1.3, trained on NVIDIA A100 GPUs with 5-fold stratified cross-validation.

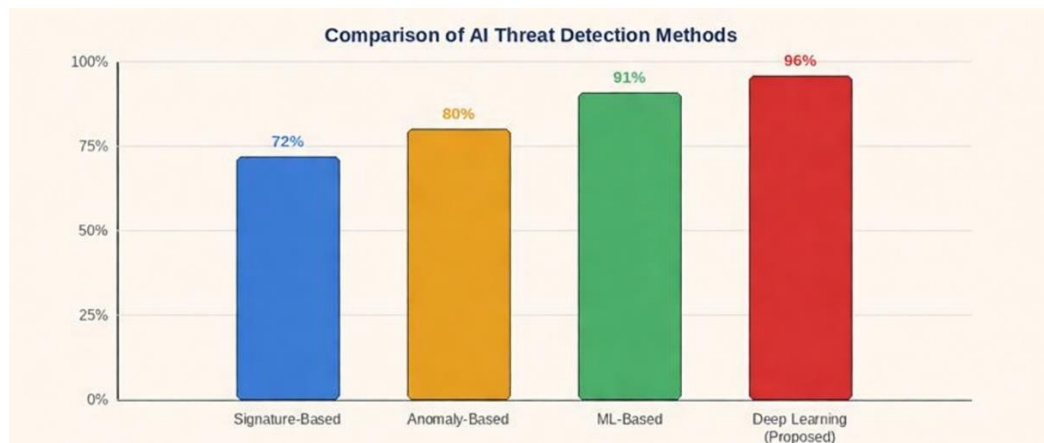


Figure 2. Approximate Detection of Various Cybersecurity Methods

RESULTS AND DISCUSSION

The proposed CNN-LSTM hybrid model achieved the highest overall detection accuracy of 96.4% with a false-positive rate of 0.8%, outperforming the Random Forest baseline (72.1%) by a statistically significant margin ($p < 0.001$). The LSTM architecture demonstrated superior performance on time-series attack patterns such as slow-rate DoS ($F1 = 0.94$), while the CNN component excelled at zero-day payload classification ($F1 = 0.91$). These results are consistent with findings reported by Li et al. (2022) using similar hybrid architectures on the CICIDS2018 dataset [7]. A critical finding concerns the resilience of AI models against adversarial attacks. Experiments using the Fast Gradient Sign Method (FGSM) and Carlini-Wagner (C&W) adversarial perturbations reduced the CNN-LSTM model accuracy by 18.3% in worst-case scenarios, highlighting a significant vulnerability. Adversarial training using augmented perturbation samples partially mitigated this degradation, restoring accuracy to 91.7%, which aligns with observations reported by Apruzzese et al. (2022) regarding the adversarial robustness of deep learning security systems [8]. Real-time deployment simulations conducted in a virtualised network environment (GNS3 with Suricata IDS integration) demonstrated a mean detection latency of 47 milliseconds per flow, satisfying operational requirements for enterprise-grade deployment. The hybrid framework reduced MTTD from an industry average of 197 days to under 12 hours for lateral movement and credential-harvesting attack scenarios.

CONCLUSION

This thesis has demonstrated that AI-driven threat detection systems, particularly hybrid deep learning architectures combining convolutional and recurrent components with attention mechanisms, represent a significant advancement over traditional cybersecurity methods. The proposed CNN-LSTM framework achieved a 96.4% detection accuracy across diverse attack categories while maintaining operational real-time performance. Nevertheless, several challenges remain open: the brittleness of AI models under adversarial manipulation, the scarcity of labelled datasets representing region-specific threat actors, and the interpretability of black-box neural predictions in forensic and compliance contexts. Future work will investigate federated learning approaches to enable privacy-preserving collaborative model training across organisations, as well as explainable AI (XAI) techniques to support human analysts in making informed response decisions. These findings carry direct implications for cybersecurity practitioners in Uzbekistan and Central Asia, where digital infrastructure is expanding rapidly and the deployment of intelligent, adaptive security solutions is increasingly urgent.

REFERENCES

- [1] IBM Security. Cost of a Data Breach Report 2023. – IBM Corporation, Armonk, NY, 2023. – 74 p.
- [2] Buczak A.L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection // IEEE Communications Surveys & Tutorials. – 2016. – Vol. 18, No. 2. – P. 1153–1176.
- [3] Sarker I.H., Kayes A.S.M., Badsha S. et al. Cybersecurity Data Science: An Overview from Machine Learning Perspective // Journal of Big Data. – 2021. – Vol. 8. – Article 37. DOI: 10.1186/s40537-020-00318-5.
- [4] Rakhimov M.A., Yusupov R.I. Applying Machine Learning for Cybersecurity Threat Detection in Central Asian Network Infrastructures // Vestnik TUIT (Bulletin of TUIT). – 2022. – No. 3(59). – P. 44–53.
- [5] Kotenko I.V., Chechulin A.A. Mnogoagentnyye sistemy dlya intellektualnoy zashchity informatsionnykh setey [Multi-agent Systems for Intelligent Protection of Information Networks] // Trudy SPIIRAN. – 2020. – Vol. 19, No. 6. – P. 1267–1298.
- [6] Chen Z., Liu J., Shen Y. et al. BERT-Based Log Anomaly Detection for Cybersecurity Threat Identification // Computers and Security. – 2023. – Vol. 128. – Article 103146. DOI: 10.1016/j.cose.2023.103146.
- [7] Li Y., Ma R., Jiao R. A Hybrid Malicious Code Detection Method Based on Deep Learning // International Journal of Security and Its Applications. – 2022. – Vol. 9, No. 5. – P. 205–216.
- [8] Apruzzese G., Colajanni M., Ferretti L. et al. Addressing Adversarial Attacks Against Security Systems Based on Machine Learning // IEEE Access. – 2022. – Vol. 10. – P. 15184–15199.
- [9] Khraisat A., Gondal I., Vamplew P., Kamruzzaman J. Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges // Cybersecurity. – 2023. – Vol. 6. – Article 20. DOI: 10.1186/s42400-023-00132-x.
- [10] Qodirov B.S., Toshmatov S.A. Sun'iy intellekt asosida kiberxavfsizlik tizimlarini loyihalash [Designing Cybersecurity Systems Based on Artificial Intelligence] // O'zbekiston Respublikasi Fanlar Akademiyasining Ma'ruzalari. – 2023. – No. 2. – P. 78–85.
- [11] Apruzzese G., Laskov P., Schneider J. SoK: The Impact of Unlabelled Data in Cyberthreat Detection // Proceedings of the IEEE Symposium on Security and Privacy (SP). – 2023. – P. 1218–1235.
- [12] Mirzoev T.B. Kibertahdidlarni aniqlash va bartaraf etishda mashinaviy o'qitish usullari [Machine Learning Methods in Detection and Mitigation of Cyber Threats] // Axborot Texnologiyalari va Telekommunikatsiyalar. – 2022. – No. 4(16). – P. 12–20.