ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, July-December 2025, Submitted in July 2025

AI-Enabled Cyber Threat Detection and Prevention Framework for Smart Cities

Manjot Maan, Research Scholar, Department of Computer Science, Shri Khushal Das University, Hanumangarh (Raj)
Prof (Dr.) Kavita, Research Supervisor, Department of Computer Science, Shri Khushal Das University, Hanumangarh (Raj)

Abstract

The rapid adoption of Internet of Things (IoT) devices, cloud platforms, artificial intelligence (AI), intelligent transportation systems, smart healthcare, smart grids, surveillance infrastructure, and connected public services has transformed conventional urban environments into highly interconnected smart cities. Although these technologies improve efficiency, sustainability, and quality of life, they also introduce a large and heterogeneous cyber-attack surface. Traditional cybersecurity mechanisms are often inadequate for detecting sophisticated, distributed, and previously unseen attacks against smart-city infrastructure. This research proposes an AI-Enabled Cyber Threat Detection and Prevention Framework for Smart Cities (AI-CTDP-SC) that integrates machine learning, deep learning, edge computing, behavioral analysis, threat intelligence, and automated prevention mechanisms.

The proposed framework collects security-related information from IoT devices, edge nodes, communication networks, cloud infrastructure, and smart-city applications. AI-based detection models analyze network traffic and device behavior to identify anomalies, malware activities, denial-of-service attacks, intrusion attempts, credential attacks, botnet behavior, and other cyber threats. Edge-based intelligence enables near-real-time detection while reducing latency and unnecessary transmission of sensitive data to centralized servers. A risk-scoring mechanism prioritizes detected threats according to their severity, confidence, affected assets, and potential impact. The prevention layer automatically applies appropriate responses, including device isolation, traffic filtering, access-control modification, session termination, and security-policy updates.

The framework is designed to support scalability, privacy, adaptability, and explainability, which are essential requirements for smart-city cybersecurity. The proposed research also defines an experimental methodology using publicly available cybersecurity datasets and simulated smart-city environments to evaluate detection accuracy, precision, recall, F1-score, false-positive rate, detection latency, and computational overhead. The framework provides a foundation for developing intelligent, adaptive, and proactive cybersecurity systems capable of protecting next-generation smart-city infrastructures.

Keywords: Artificial Intelligence, Cybersecurity, Smart Cities, IoT Security, Machine Learning, Deep Learning, Edge Computing, Intrusion Detection, Threat Prevention, Anomaly Detection.

1. Introduction

Smart cities represent a major evolution in urban infrastructure in which digital technologies are integrated with transportation, energy, healthcare, education, public safety, water management, waste management, environmental monitoring, and administrative services. Sensors, cameras, connected vehicles, smart meters, mobile applications, cloud platforms, and edge-computing devices continuously generate and exchange information. This connectivity enables governments and service providers to make data-driven decisions and provide efficient public services.

However, extensive connectivity also creates significant cybersecurity risks. A vulnerability in a single IoT device can potentially become an entry point into a larger network. Attackers may exploit weak authentication, outdated firmware, insecure communication protocols, exposed interfaces, misconfigured cloud services, or compromised user credentials. Furthermore, smart-city systems are heterogeneous and may contain devices with different operating systems, hardware capabilities, communication protocols, and security requirements.

Conventional cybersecurity systems generally rely on predefined signatures, static rules, or manually configured security policies. These approaches can be effective against known attacks but may struggle with zero-day attacks, polymorphic malware, insider threats, and previously unseen behavioral patterns. Consequently, intelligent cybersecurity mechanisms capable of learning from data and identifying abnormal behavior are increasingly important.

Artificial intelligence provides opportunities to improve cyber-threat detection by automatically learning patterns from large-scale security data. Machine-learning algorithms can classify malicious and benign traffic, while deep-learning models can identify complex relationships in high-dimensional network data. Edge AI further enables security analysis closer to the data source, reducing response latency.

This research therefore proposes an AI-enabled framework that combines AI-based threat detection, edge intelligence, behavioral analysis, risk assessment, threat intelligence, and automated prevention into a unified architecture for smart-city environments.

1.1 Problem Statement

Smart-city networks face several challenges:

1. Large numbers of heterogeneous IoT devices increase the attack surface.
2. Traditional signature-based systems cannot reliably detect unknown attacks.
3. Centralized security analysis can introduce communication and response delays.
4. Smart-city environments generate large volumes of security data.
5. False-positive alerts can overwhelm cybersecurity administrators.
6. Resource-constrained IoT devices cannot always execute complex security algorithms.
7. Privacy-sensitive urban data requires careful processing and protection.
8. Cyber threats continuously evolve, requiring adaptive security mechanisms.

Therefore, there is a need for an intelligent framework capable of detecting, analyzing, prioritizing, and preventing cyber threats in real time while maintaining scalability and privacy.

1.2 Research Objectives

The major objectives of this research are:

- To design an AI-enabled cybersecurity framework for smart-city infrastructures.
- To develop intelligent mechanisms for detecting known and unknown cyber threats.
- To integrate edge computing for low-latency threat detection and response.
- To develop an adaptive risk-scoring mechanism for prioritizing security incidents.
- To design automated prevention mechanisms for responding to detected attacks.
- To reduce false-positive rates while maintaining high detection performance.
- To evaluate the framework using appropriate cybersecurity datasets and performance metrics.
- To investigate the applicability of explainable AI for security decision-making.

1.3 Research Questions

RQ1: How can AI improve cyber-threat detection in heterogeneous smart-city networks?

RQ2: Can edge-based AI reduce threat-detection latency compared with centralized security analysis?

RQ3: How effectively can a hybrid AI model identify both known and previously unseen attacks?

RQ4: Can automated risk-based response reduce the impact of detected cyber threats?

RQ5: How can explainability improve the trustworthiness of AI-based cybersecurity decisions?

2. Related Work

Cybersecurity research has increasingly focused on machine learning and deep learning because of their ability to identify patterns in large security datasets. Traditional intrusion detection systems generally use either signature-based or anomaly-based techniques.

Signature-based systems compare network activities against previously identified attack patterns. Their major limitation is poor performance against unknown attacks.

Machine-learning techniques such as decision trees, random forests, support vector machines, k-nearest neighbors, and ensemble methods have been applied to network intrusion detection. These models can identify relationships between network features and attack classes. However, their performance depends strongly on feature engineering and dataset quality.

Deep-learning techniques provide an alternative by learning hierarchical representations from data. Convolutional neural networks can identify local patterns in transformed network features, while recurrent neural networks and long short-term memory networks can capture temporal relationships in sequential traffic. Autoencoders have also been applied to anomaly detection by learning representations of normal network behavior.

The development of IoT has further increased the importance of intelligent cybersecurity. IoT environments contain numerous devices that communicate continuously and may have limited computational resources. Edge computing provides an effective approach by moving selected computation from centralized cloud servers to edge nodes located closer to IoT devices.

Despite these developments, several gaps remain. Many existing solutions focus on individual attack types or specific network environments. Others evaluate only classification accuracy without considering detection latency, resource consumption, false positives, privacy, or automated response. Smart-city cybersecurity requires an integrated approach covering detection, risk analysis, prevention, and continuous adaptation.

3. Proposed AI-CTDP-SC Framework

The proposed AI-Enabled Cyber Threat Detection and Prevention Framework for Smart Cities (AI-CTDP-SC) consists of six major layers:

1. Smart-City Data Acquisition Layer
2. Edge Security Processing Layer
3. AI Threat Detection Layer
4. Threat Intelligence and Risk Assessment Layer
5. Automated Prevention and Response Layer
6. Cloud-Based Security Management Layer

3.1 Smart-City Data Acquisition Layer

The first layer collects security-related information from smart-city infrastructure. Potential sources include:

- IoT sensors
- Smart meters
- CCTV and surveillance systems
- Intelligent transportation systems
- Smart-grid devices
- Healthcare devices
- Environmental monitoring systems
- Public Wi-Fi infrastructure
- Edge gateways
- Cloud applications

The collected information may include packet statistics, source and destination addresses, protocol information, connection duration, packet size, authentication events, device behavior, system logs, and application-level events.

3.2 Edge Security Processing Layer

The edge layer processes security information close to the data source. Instead of transmitting every raw packet or event to a centralized cloud environment, edge nodes perform preliminary filtering, feature extraction, and threat analysis.

This architecture provides three major benefits:

- Reduced latency: Threats can be detected closer to the source.
- Reduced bandwidth consumption: Only relevant security information needs to be transmitted.
- Improved privacy: Sensitive raw data can remain within local infrastructure.

Lightweight AI models can be deployed on edge nodes, while computationally intensive models can execute within centralized cloud infrastructure.

3.3 AI Threat Detection Layer

The core of the proposed framework is a hybrid AI detection engine. The framework can combine multiple models rather than depending on a single algorithm.

A possible architecture includes:

Random Forest + CNN + LSTM + Autoencoder

Random Forest can provide robust classification for structured network features. CNN can identify complex feature patterns, while LSTM can analyze temporal relationships in network traffic. An autoencoder can identify deviations from learned normal behavior and assist in detecting unknown attacks.

A hybrid prediction can be represented as:

$P\_hybrid = w1\ P\_RF + w2\ P\_CNN + w3\ P\_LSTM + w4\ P\_AE$

where the individual terms represent predictions generated by individual models and the weights represent their corresponding contributions.

The final decision can be generated according to:

$$Attack = 1, \text{ if } P\_hybrid >= \tau$$
$$Attack = 0, \text{ if } P\_hybrid < \tau$$

where τ is the selected detection threshold.

3.4 Threat Intelligence Layer

The threat-intelligence layer enriches AI-generated alerts with contextual security information. Threat intelligence may include indicators of compromise, malicious IP addresses, suspicious domains, known malware patterns, attack techniques, and historical incident information.

Combining machine-generated predictions with threat intelligence can improve the contextual understanding of security events.

3.5 Risk Assessment Mechanism

Not every detected anomaly represents the same level of danger. Therefore, the proposed framework assigns a dynamic risk score.

$R = \alpha D + \beta C + \gamma I + \delta V$

where D is detection confidence, C is criticality of the affected asset, I is estimated impact, V is vulnerability or exposure level, and α, β, γ, δ are weighting parameters.

The risk score can be classified into Low risk, Moderate risk, High risk, and Critical risk. Critical events should receive immediate automated response, while low-risk events may be forwarded to administrators for further investigation.

4. Automated Prevention and Response

Detection alone is insufficient for modern smart-city cybersecurity. The proposed framework therefore includes automated response capabilities.

Depending on the threat severity, the system can:

- Block suspicious network traffic.
- Isolate compromised IoT devices.
- Disable suspicious sessions.
- Update firewall rules.
- Change access-control policies.
- Trigger endpoint security mechanisms.
- Restrict communication between compromised and trusted devices.
- Notify cybersecurity administrators.
- Initiate additional authentication.
- Collect forensic information for investigation.

A simplified response strategy is:

Response = Monitor, for low risk
Response = Alert, for moderate risk
Response = Restrict, for high risk
Response = Isolate, for critical risk

This approach supports proportional response rather than automatically applying the strongest action to every detected anomaly.

5. Explainable AI

AI-based cybersecurity decisions may influence critical urban infrastructure. Consequently, security administrators should understand why a model classified an event as malicious.

Explainable AI techniques can identify the most influential features responsible for a prediction. For example, a detection model may identify unusual packet rates, abnormal connection frequency, unexpected ports, or unusual authentication patterns as important indicators.

Explainability can improve analyst trust, incident investigation, model validation, regulatory compliance, error identification, and security-policy development.

The framework can therefore integrate explainability into the analyst dashboard.

6. Privacy and Security Considerations

Smart-city data may contain sensitive information about residents, transportation patterns, public services, and infrastructure. The framework should therefore avoid unnecessary centralization of raw data.

Privacy-preserving mechanisms may include:

- Data minimization
- Encryption
- Secure communication
- Federated learning
- Access control
- Differential privacy where appropriate
- Local edge processing
- Secure model aggregation

Federated learning is particularly relevant because edge devices can train local models while sharing model updates instead of directly transmitting raw datasets.

7. Proposed System Architecture

The conceptual data flow of the framework is:

Smart-City Devices → Edge Gateway → Data Preprocessing → Feature Extraction → AI Detection → Threat Intelligence → Risk Assessment → Automated Response → Cloud Security Management

The edge gateway performs initial processing. Suspicious events are passed to the AI engine. The detection engine generates a probability score and attack classification. The risk engine combines AI confidence with asset criticality and threat intelligence. Finally, the prevention engine executes an appropriate response.

The cloud management layer maintains historical security information and supports model training, centralized monitoring, policy management, and long-term security analytics.

8. Research Methodology

The proposed framework can be evaluated through a controlled experimental environment.

8.1 Dataset Selection

Potential datasets include publicly available intrusion-detection and IoT-security datasets such as:

- CICIDS2017
- CSE-CIC-IDS2018
- UNSW-NB15
- Bot-IoT
- TON_IoT

The final dataset selection should depend on the specific smart-city attack scenarios being investigated.

8.2 Data Preprocessing

The preprocessing stage may include:

1. Removing duplicate records.
2. Handling missing values.
3. Encoding categorical features.
4. Normalizing numerical features.
5. Removing irrelevant attributes.
6. Addressing class imbalance.
7. Splitting data into training, validation, and testing sets.

Care must be taken to prevent data leakage between training and testing datasets.

8.3 Model Training

Individual AI models are trained independently before integration into the hybrid detection engine.

The experimental process can compare:

- Random Forest
- Support Vector Machine
- CNN
- LSTM
- Autoencoder
- Hybrid AI model

The hybrid model is expected to provide improved robustness by combining complementary learning capabilities.

9. Performance Evaluation

The framework should be evaluated using multiple metrics rather than accuracy alone.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

$$\text{False Positive Rate} = FP / (FP + TN)$$

Additional evaluation criteria should include detection latency, response latency, CPU utilization, memory consumption, network overhead, model size, energy consumption, scalability, and performance against previously unseen attacks.

The framework should report actual experimental results only after implementation and testing. Numerical performance values should not be fabricated.

10. Expected Results

The proposed framework is expected to provide several improvements over conventional cybersecurity mechanisms.

First, hybrid AI models are expected to improve the detection of complex attack patterns because different models capture different characteristics of network behavior. Second, edge-based processing should reduce detection latency by performing preliminary analysis near the source. Third, the risk-based response mechanism should help cybersecurity administrators prioritize critical incidents.

The anomaly-detection component is expected to improve the identification of previously unseen behavioral patterns. However, unknown-attack detection should be evaluated carefully because anomaly detection can produce false positives when legitimate smart-city behavior changes.

The framework is also expected to reduce unnecessary data transmission through edge processing and support privacy-preserving security analytics.

11. Advantages of the Proposed Framework

11.1 Real-Time Detection

Edge-based processing enables security events to be analyzed closer to their origin.

11.2 Adaptive Security

AI models can be periodically retrained using new security data and emerging attack patterns.

11.3 Multi-Layer Protection

The framework integrates device, network, edge, cloud, and application-level security.

11.4 Automated Response

High-risk incidents can trigger predefined containment mechanisms without waiting for manual intervention.

11.5 Scalability

Distributed edge intelligence can support large numbers of smart-city devices.

11.6 Privacy Preservation

Local processing and federated-learning mechanisms can reduce unnecessary transmission of sensitive information.

11.7 Explainability

AI explanations can help security analysts understand important factors behind automated decisions.

12. Challenges and Limitations

Despite its potential, the framework has several limitations.

AI models require representative and high-quality training data. Cybersecurity datasets may not fully represent real smart-city environments. Changes in device behavior can also cause concept drift, reducing model performance over time.

Edge devices have limited computational and memory resources, making deployment of large deep-learning models difficult. Model compression, quantization, pruning, or lightweight architectures may therefore be required.

Automated response also creates risks. An incorrect classification could cause a legitimate device to be isolated, potentially disrupting critical services. Consequently, response actions should consider confidence, asset criticality, and operational context.

Adversarial machine-learning attacks represent another challenge. Attackers may deliberately manipulate input data to evade AI detection. Robust training and continuous model validation are therefore important.

13. Future Research Directions

Future research can extend the framework in several directions.

First, federated learning can be integrated with edge AI to enable collaborative model training without centralized collection of sensitive data. Second, graph neural networks can model relationships between IoT devices and identify coordinated attacks.

Third, reinforcement learning could be investigated for adaptive cyber-response policies. Instead of using only fixed response rules, reinforcement-learning agents could learn response strategies based on security outcomes.

Fourth, digital twins of smart-city infrastructures could provide realistic environments for testing cybersecurity strategies without disrupting real-world services.

Future studies should also investigate adversarially robust AI, continual learning, zero-day detection, explainable deep learning, and energy-efficient security models.

14. Conclusion

This research proposes an AI-Enabled Cyber Threat Detection and Prevention Framework for Smart Cities that integrates artificial intelligence, edge computing, anomaly detection, threat intelligence, risk assessment, explainable AI, and automated response. The framework addresses important limitations of conventional cybersecurity approaches by providing an adaptive and distributed security architecture.

The combination of machine learning and deep learning can support detection of both known and anomalous cyber behaviors, while edge intelligence can reduce detection latency and communication overhead. The proposed risk-scoring mechanism enables security events to be prioritized according to their potential impact, and automated response mechanisms can help contain high-risk incidents.

The framework provides a conceptual foundation for developing intelligent cybersecurity systems suitable for complex smart-city environments. Its effectiveness should be validated experimentally using representative IoT and network-security datasets and, where possible, realistic smart-city testbeds. Future integration of federated learning, graph neural networks, continual learning, adversarial robustness, and digital twins can further strengthen the proposed architecture.

Overall, AI-enabled cybersecurity can play a critical role in protecting the digital infrastructure of future smart cities. However, successful deployment requires not only high detection

accuracy but also privacy, explainability, low latency, scalability, robustness, and safe automated response.

References

1. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176.
2. Doshi, R., Apthorpe, N., & Feamster, N. (2018). Machine learning DDoS detection for consumer Internet of Things devices. IEEE Security and Privacy Workshops.
3. Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50.
4. Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. Military Communications and Information Systems Conference.
5. Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. International Conference on Information Systems Security and Privacy.
6. Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. (2018). N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders. IEEE Pervasive Computing, 17(3), 12–22.
7. Abomhara, M., & Køien, G. M. (2014). Cyber security and the Internet of Things: Vulnerabilities, threats, intruders and attacks. Journal of Cyber Security and Mobility, 4(1), 65–88.
8. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.
9. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of AISTATS.
10. Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. IEEE Symposium on Security and Privacy.
11. Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. Cybersecurity, 2, 20.
12. NIST. Cybersecurity Framework (CSF) 2.0. National Institute of Standards and Technology.
13. ENISA. ENISA Threat Landscape. European Union Agency for Cybersecurity.
14. ISO/IEC 27001. Information security, cybersecurity and privacy protection — Information security management systems — Requirements.
15. ISO/IEC 27005. Information security, cybersecurity and privacy protection — Guidance on managing information security risks.