electronics

MDPI

Article

AI-Driven Anomaly Detection for Securing IoT Devices in 5G-Enabled Smart Cities

Manuel J. C. S. Reis

Engineering Department & IEETA, University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5000-801 Vila Real, Portugal; mcabral@utad.pt

Abstract: This paper proposes a novel AI-driven anomaly detection framework designed to enhance cybersecurity in IoT-enabled smart cities operating over 5G networks. While prior research has explored deep learning for anomaly detection, most existing systems rely on single-model architectures, employ centralized training, or lack support for real-time, privacy-preserving deployment—limiting their scalability and robustness. To address these gaps, our system integrates a hybrid deep learning model combining autoencoders, long short-term memory (LSTM) networks, and convolutional neural networks (CNNs) to detect spatial, temporal, and reconstruction-based anomalies. Additionally, we implement federated learning and edge AI to enable decentralized, privacy-preserving threat detection across distributed IoT nodes. The system is trained and evaluated using a combination of real-world (CICIDS2017, TON_IoT, UNSW-NB15) and synthetically generated attack data, including adversarial perturbations. Experimental results show our hybrid model achieves a precision of 97.5%, a recall of 96.2%, and an F1 score of 96.8%, significantly outperforming traditional IDS and standalone deep learning methods. These findings validate the framework's effectiveness and scalability, making it suitable for real-time intrusion detection and autonomous threat mitigation in smart city environments.

Keywords: anomaly detection; cybersecurity; Internet of Things (IoT); 5G networks; deep learning

check for updates

Academic Editor: Christos J. Bouras

Received: 21 April 2025
Revised: 6 June 2025
Accepted: 17 June 2025
Published: 19 June 2025

Citation: Reis, M.J.C.S. AI-Driven Anomaly Detection for Securing IoT Devices in 5G-Enabled Smart Cities. *Electronics* **2025**, *14*, 2492. https://doi.org/10.3390/electronics14122492

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Background and Motivation

The rapid proliferation of Internet of Things (IoT) devices has revolutionized urban infrastructures, leading to the emergence of smart cities. These interconnected devices facilitate the efficient management of city resources, enhance service delivery, and improve the quality of life for residents. However, the integration of IoT devices into critical urban systems introduces significant cybersecurity challenges. The vast number of connected devices expands the attack surface, making smart cities vulnerable to various cyber threats, including data breaches, denial-of-service (DoS) attacks, and unauthorized access [1,2].

The emergence of 5G networks further accelerates IoT adoption by enabling ultra-low latency and high-speed data transmission. While these capabilities empower real-time city-wide connectivity, they also introduce new cybersecurity challenges, including large-scale distributed attacks and more sophisticated adversarial techniques [3]. The need for resilient and adaptive security mechanisms is more urgent than ever.

1.2. Importance of Cybersecurity in IoT and 5G Networks

Traditional cybersecurity tools, particularly rule-based intrusion detection systems (IDS), struggle to keep pace with the dynamic and decentralized nature of modern IoT environments. Their rigidity and limited detection capabilities make them ill-suited for identifying zero-day attacks or adapting to evolving threats [4]. This creates an urgent need for intelligent, scalable, and real-time solutions that can safeguard massive IoT ecosystems in a proactive and adaptive manner.

1.3. Role of AI in Enhancing Cybersecurity

Artificial Intelligence (AI) and machine learning (ML) have emerged as pivotal technologies in enhancing cybersecurity measures. AI-driven anomaly detection systems can analyze vast amounts of data to identify patterns and detect deviations indicative of potential cyber threats. In the context of IoT and 5G networks, AI can provide real-time monitoring and adaptive security responses, making it an effective tool for safeguarding smart city infrastructures. Specifically, deep learning models such as autoencoders and long short-term memory (LSTM) networks have shown significant promise in terms of learning normal network behavior and identifying those anomalies that are indicative of cyber threats [5]. However, while several AI-based IDS approaches exist, most rely on centralized architectures or single-model configurations, which limits their scalability, adaptability, and privacy compliance.

1.4. Challenges in AI-Driven Anomaly Detection for IoT Security

Despite the promising capabilities of AI-driven anomaly detection systems, several challenges hinder their deployment in real-world IoT ecosystems:

- Device and Data Heterogeneity: IoT networks consist of highly diverse devices with varying capabilities and protocols, leading to inconsistent data patterns and vulnerabilities [6].
- Real-Time Requirements: Cybersecurity systems must process vast data streams quickly and with minimal latency to ensure an effective threat response [7].
- Evolving Threat Landscape: Attackers frequently develop novel evasion techniques, necessitating continuous learning and model adaptation [8–10].
- Scalability and Privacy: Models must scale across thousands of devices while preserving user data privacy and minimizing resource usage [11–13].

These challenges expose a significant gap in current research: most existing approaches do not simultaneously address model hybridization, federated learning, real-time edge deployment, and autonomous mitigation—capabilities that are essential for next-generation IoT security in 5G-enabled environments.

1.5. Major Contributions of This Paper

This research addresses the aforementioned research gap by presenting an AI-driven anomaly detection framework tailored for smart cities. The proposed solution introduces a hybrid deep learning architecture—combining autoencoders, LSTM networks, and CNNs—to capture spatial, temporal, and reconstruction-based anomalies in IoT network behavior. The contributions are:

1. A novel hybrid intrusion detection system (IDS) that integrates autoencoders, LSTMs, and CNNs to improve anomaly classification accuracy and robustness.
2. A federated learning and edge AI architecture enabling decentralized model training across IoT nodes, which enhances scalability and preserves data privacy.

3. An experimental validation using real-world datasets (CICIDS2017, TON_IoT, UNSW-NB15) and synthetic attack data, demonstrating high accuracy (97.5% precision, 96.8% F1 score) and low inference latency (<310 ms) suitable for edge deployment.
4. A comparative evaluation, showing that the hybrid model outperforms standalone models and rule-based IDS in terms of detection performance, efficiency, and adaptability.
5. Integration of autonomous threat mitigation mechanisms, providing real-time security enforcement based on anomaly detection.
6. Use of explainable AI (XAI) methods, such as SHAP values, to support interpretability and trust in AI-driven decisions.

1.6. Structure of the Paper

The remainder of this paper is organized as follows. Section 2 presents related work, highlighting the limitations of current AI-based IDS approaches and identifying the research gap that this study addresses. Section 3 describes the proposed hybrid AI-driven methodology, including its deep learning architecture, federated learning framework, and deployment strategy. Section 4 discusses the experimental setup, evaluation metrics, and results, while Section 5 analyzes the findings in comparison to state-of-the-art solutions and discusses the practical implications. Finally, Section 6 concludes the paper and outlines directions for future work focused on improving robustness, scalability, and real-world deployment.

2. Related Work

2.1. AI-Driven Anomaly Detection in IoT Networks

The integration of artificial intelligence (AI) into anomaly detection has significantly enhanced the security of IoT networks. Several AI models, including convolutional neural networks (CNNs), long short-term memory (LSTM) networks, generative adversarial networks (GANs), and hybrid AI approaches, have been applied to detect IoT anomalies and cybersecurity threats in smart cities. Recent studies have explored the role of deep learning in detecting network anomalies by identifying behavioral deviations and previously unseen attack patterns [1,5].

Abusitta et al. proposed a deep learning-enabled anomaly detection framework for IoT systems, utilizing denoising autoencoders to extract robust features from heterogeneous and noisy environments, thereby enhancing the accuracy of detecting malicious data [14]. Another study by Priyadarshini highlights the role of federated learning and split learning in collaborative anomaly detection while ensuring data privacy across distributed IoT devices [1].

A theoretical framework aimed at improving IoT network cybersecurity using AI-driven anomaly detection emphasizes the importance of unsupervised learning techniques, such as autoencoders, for detecting threats in smart city ecosystems [5]. Additionally, advanced AI methods have been developed to analyze high-dimensional IoT network traffic, allowing for more efficient anomaly detection across complex infrastructures [6].

2.2. AI and IoT Integration for Enhanced Security

The convergence of AI and IoT technologies has led to the development of next-generation security systems in smart cities. AI-powered algorithms can process data from various IoT devices to detect anomalies, predict failures, and optimize operations across diverse infrastructure systems. This integration enhances urban security by providing advanced threat detection, predictive analytics, and real-time monitoring [15,16].

A key advancement in AI-IoT security is the adoption of edge AI and federated learning (FL). Traditional centralized AI models often struggle with latency, bandwidth consumption, and data privacy concerns. FL enables AI models to be trained on distributed

edge devices, reducing data transmission risks while preserving user privacy [17]. For example, Sáez-de-Cámara et al. proposed a clustered federated learning architecture for network anomaly detection, which effectively addresses scalability and heterogeneity challenges in large-scale IoT networks [17].

2.3. Challenges in AI-Based Anomaly Detection

Despite research advancements, several challenges persist in implementing AI-based anomaly detection systems:

1. Data Quality and Diversity: The heterogeneity of IoT devices leads to diverse data types and quality, complicating the development of universal anomaly detection models [18,19]. Abusitta et al. addressed this issue by utilizing denoising autoencoders to extract robust features from corrupted IoT data [14].
2. Real-Time Processing: Achieving real-time anomaly detection requires efficient algorithms that are capable of processing large data streams with minimal latency [20,21]. Liu et al. discussed the application of unsupervised deep learning techniques for IoT time series analysis, highlighting the importance of efficient models for real-time anomaly detection [22].
3. Scalability: Deploying AI models across extensive IoT networks necessitates scalable solutions that maintain performance as the network grows [23,24]. The clustered federated learning approach by Sáez-de-Cámara et al. offers a scalable solution by addressing heterogeneity and scalability in large IoT networks [17].
4. Explainability and Interpretability: Many deep learning models function as black boxes, making it difficult for security teams to interpret AI-generated threat detection results. Explainable AI (XAI) techniques, such as SHAP values and attention mechanisms, have been proposed to increase model transparency and reduce false positives in cybersecurity applications [25]. Subasi et al. critically assessed interpretable and explainable machine learning models for intrusion detection, emphasizing the need for transparency in AI-driven security systems [25].

Addressing these challenges is crucial for the effective deployment of AI-driven security measures in smart cities.

2.4. Recent Developments and Future Directions

Recent research highlights the role of generative AI in simulating attacks and augmenting training data. It also facilitates adversarial training to strengthen model robustness [26]. In parallel, AI models are increasingly being applied to 5G traffic analysis, with real-time IDS systems showing promise for use in high-throughput environments [27].

Self-learning AI architectures that adapt to novel threats without human supervision are becoming a research focus area, supporting autonomous response mechanisms that are critical to future smart city security infrastructures.

2.5. Summary of Prior Work and Research Gaps

To contextualize this study, Table 1 compares key prior works in the domain, outlining their focus, techniques, scalability, and limitations.

Based on this literature review, we identified the following key findings and research gaps:

1. Key Findings:
   - AI-based models significantly outperform traditional IDS in terms of accuracy and adaptability.
   - Edge AI and FL frameworks are crucial for scalability and privacy preservation.
   - Real-time security systems require low-latency models with autonomous capabilities.

- Explainability remains an active challenge in AI-based cybersecurity.

2. Research Gaps:
   - Hybrid AI Models: Limited exploration of deep hybrid frameworks, combining temporal, spatial, and generative insights.
   - Real-World Testing: Most prior work lacks deployment in city-scale or production-like environments.
   - Scalability and Energy Efficiency: High-complexity models are rarely optimized for edge deployment.
   - Lack of Integrated Response Systems: Few systems incorporate AI-driven autonomous mitigation mechanisms.

Table 1. Comparison of recent AI-based anomaly detection approaches in IoT security. The table highlights each model's deployment context, the use of federated learning or edge AI, support for autonomous mitigation, and key limitations. Our proposed work addresses multiple research gaps by integrating hybrid deep learning with decentralized, real-time, and self-adaptive cybersecurity mechanisms.

Study	Model Type	Deployment Focus	FL/Edge AI	Autonomous Mitigation	Key Limitation
Abusitta et al. [14]	Denoising Autoencoder	IoT anomaly detection	No	No	Limited scalability
Priyadarshini [1]	Split Learning + FL	Distributed IoT	Yes	No	No autonomous response
Sáez-de-Cámara et al. [17]	Clustered Federated Learning	IoT network security	Yes	No	Limited to clustering
Liu et al. [22]	Unsupervised LSTM	IoT time series	No	No	No real-time validation
This Work	Hybrid DL (AE + LSTM + CNN)	5G-enabled smart cities	Yes	Yes	Comprehensive solution

To address these gaps, this research proposes an AI-driven anomaly detection system that integrates hybrid deep learning models, real-world dataset validation, scalable federated learning techniques, and AI-based autonomous mitigation strategies to enhance cybersecurity in IoT-enabled smart cities. The discussion presented in Table 1 and Table 8 is extended by Table 9, which provides a side-by-side technical comparison of our approach with representative recent studies.

3. Proposed Methodology

3.1. Overview of the Proposed AI-Driven Anomaly Detection System

The rapid proliferation of IoT devices in 5G-enabled smart cities has significantly increased cybersecurity risks, necessitating advanced AI-driven anomaly detection systems. Traditional rule-based intrusion detection systems (IDS) and statistical methods have struggled to cope with the complexity, volume, and dynamic nature of modern IoT threats. To address these challenges, this research proposes an AI-driven anomaly detection system that leverages hybrid deep learning models, federated learning for scalability, and autonomous threat mitigation mechanisms to enhance security in large-scale IoT deployments.

The proposed framework consists of multiple layers that collaboratively enhance threat detection accuracy, improve scalability, and provide an adaptive response mechanism for mitigating cyberattacks in real time. The system is designed to process vast amounts of network traffic data, identify suspicious patterns using AI models, and deploy automated countermeasures in response to detected threats. By integrating edge AI and federated

learning, the system ensures low-latency threat detection while maintaining data privacy and minimizing computational overhead.

The architecture of the proposed anomaly detection system comprises four main components: (1) data collection and preprocessing, (2) AI-driven anomaly detection, (3) federated learning and edge AI processing, and (4) adaptive threat mitigation.

- Data collection and preprocessing: Captures traffic from IoT devices (e.g., sensors, meters, and cameras) and applies cleaning, normalization, and dimensionality reduction to ensure high-quality input.
- AI-driven anomaly detection: A hybrid deep learning model composed of autoencoders (for pattern reconstruction), LSTM networks (for temporal dependencies), and CNNs (for spatial feature extraction). This model distinguishes between normal behavior and sophisticated attack patterns.
- Federated learning and edge AI processing: Distributes model training across edge devices. Using federated learning, local models are trained on-site, with only the model parameters shared for global aggregation. This ensures data privacy and low-latency threat detection.
- Adaptive threat mitigation: Incorporates AI-based autonomous response mechanisms. When anomalies are detected, the system triggers actions like firewall rule updates or device isolation. Explainable AI (XAI) modules provide interpretability using SHAP values and attention mechanisms.

Figure 1 illustrates this multi-layered architecture, highlighting data flow, AI models, federated learning, and threat mitigation processes.

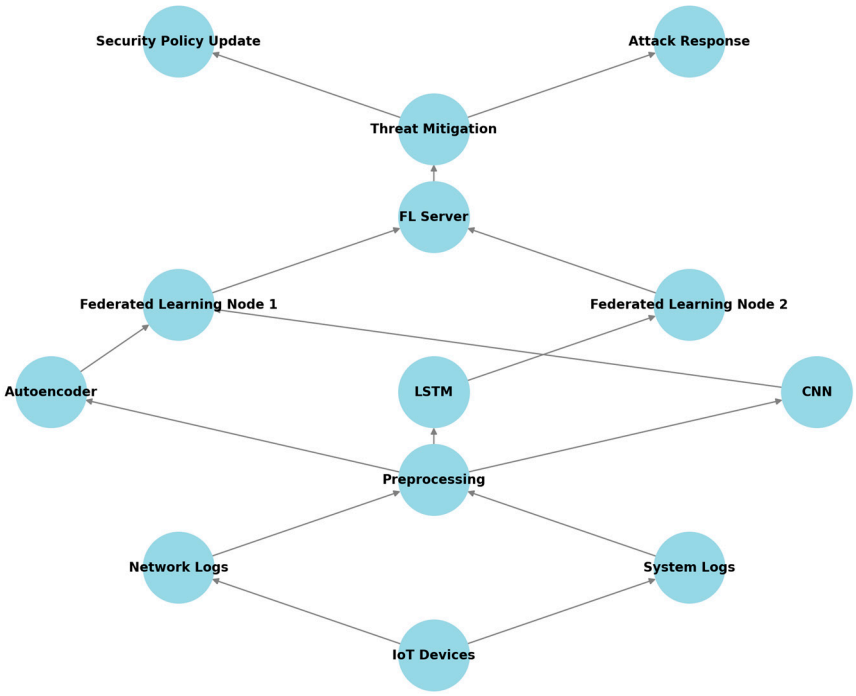


Figure 1. System architecture of the proposed AI-driven anomaly detection system. The system consists of four main components: (1) a data collection layer, responsible for gathering real-time IoT network traffic and system logs; (2) an AI-driven anomaly detection layer, utilizing a hybrid deep learning approach (autoencoders, LSTMs, and CNNs) for anomaly classification; (3) a federated learning and edge AI layer, ensuring scalable, privacy-preserving anomaly detection; and (4) an adaptive threat mitigation layer, enabling autonomous response mechanisms to the detected cyber threats.

3.2. Dataset Selection and Preprocessing

We employed a combination of real-world and synthetic datasets to ensure model robustness and resilience to zero-day threats. Public datasets include CICIDS2017 (available at: https://www.unb.ca/cic/datasets/ids-2017.html, accessed on 16 June 2025), TON_IoT (available at: https://research.unsw.edu.au/projects/toniot-datasets, accessed on 16 June 2025), and UNSW-NB15 (available at: https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/, accessed on 16 June 2025). Synthetic datasets simulate botnet and adversarial attacks (e.g., FGSM and PGD) to stress-test the system.

In particular, the TON_IoT dataset provides telemetry data from real-world IoT sensors, actuators, and smart infrastructure components, including smart meters, industrial control systems, and connected urban devices—making it directly applicable to smart city environments. This dataset reflects the actual operational patterns and threat scenarios encountered in smart urban systems.

To further align our experimental setup with smart city contexts, we supplemented the public datasets with synthetically generated attack scenarios targeting smart city assets. These included simulations of botnet attacks on smart surveillance networks and utility systems, as well as adversarial manipulations mimicking zero-day threats. This combined dataset strategy ensures that our evaluation closely mirrors realistic conditions in 5G-enabled smart cities.

3.2.1. Synthetic Data Generation for Attack Simulation

While publicly available datasets cover various known cyber threats, real-world IoT environments frequently encounter zero-day attacks and adversarial intrusions that may not be well-represented in the existing datasets. To increase model robustness and evaluate its resilience to adversarial threats, additional synthetic data were generated using the following methodologies:

- IoT botnet attack simulations:
  - These involve simulated Mirai and Bashlite malware variants, replicating large-scale botnet attacks on IoT infrastructures.
  - Attack traffic was crafted using real-world botnet behavior patterns to mimic infected IoT devices that are launching coordinated attacks.
- Adversarial attack injection
  To evaluate the vulnerability of the anomaly detection system to adversarial evasions, adversarial attack perturbations were generated using state-of-the-art adversarial machine learning techniques.
  - The fast gradient sign method (FGSM):
    - The FGSM is a white-box attack method that introduces small perturbations to network traffic feature representations, in order to mislead the anomaly detection system.
    - Given an input X and a loss function $J(X, y)$, the FGSM computes an adversarial example by adjusting X in the direction of the gradient:

$$X' = X + \epsilon \cdot \text{sign}(\nabla_X J(X, y))$$

    - This ensures that the perturbed sample remains indistinguishable to a human observer, but the model misclassifies it as benign.
  - Projected gradient descent (PGD):

- PGD extends the FGSM by iteratively applying small perturbations and projecting the modified sample within a bounded region to maximize misclassification.
- The attack can be shown as follows:

$$X'_{t+1} = \Pi_{X,\epsilon}\left(X'_t + \alpha \cdot \text{sign}\left(\nabla_X J\left(X'_t, y\right)\right)\right)$$

- Unlike the FGSM, PGD iteratively refines the perturbations, making it a stronger adversarial attack that is more challenging to detect.

By incorporating FGSM and PGD adversarial perturbations into the dataset, the system's resilience against evasion techniques was assessed, ensuring that deep learning models maintain robust performance, even when confronted with adversarially modified inputs.

- Randomized data perturbations:
  - These introduce controlled noise variations in normal network traffic patterns to test the generalization capabilities.
  - These also ensure that the model learns robust features, preventing overfitting to static attack signatures.

This combination of real-world and synthetic datasets ensures that the proposed anomaly detection system can effectively detect both known and previously unseen (zero-day) cyber threats in smart city IoT infrastructures.

3.2.2. Data Preprocessing

Before feeding the raw network data into deep learning models, a structured data preprocessing pipeline is applied. The preprocessing phase is critical for improving model accuracy, increasing computational efficiency, and reducing false positives. The key preprocessing steps include data cleaning, normalization, feature selection, and encoding categorical data:

- Data cleaning and handling missing values: Missing values are handled using mean imputation for numerical features and mode imputation for categorical attributes. Duplicate records and corrupted logs are removed to ensure a cleaner dataset.
- Feature engineering and selection: Chi-square statistical tests are used to retain only the most significant features correlated with anomaly detection. Recursive feature elimination (RFE) with a random forest classifier is used to filter out low-impact features. Principal component analysis (PCA) is employed for dimensionality reduction while preserving key data variance.
- Feature normalization and encoding: Min–Max normalization is applied to numerical features to scale values between 0 and 1. One-hot encoding is used to convert categorical variables (e.g., protocol types and attack labels) into numerical representations.
- Data Splitting Strategy: 70% training set, used for model learning; 15% validation set, used for hyperparameter tuning and model optimization; 15% test set, used for a final performance evaluation to assess generalizability on unseen data. Additionally, cross-validation (e.g., 5-fold cross-validation) is applied to ensure that the model performance is not biased by a specific data split.

3.2.3. Justification for Dataset Choices and Preprocessing Steps

The selection of multiple real-world datasets ensures that the system is tested across diverse IoT environments, preventing overfitting to a single network setting. Furthermore, the integration of synthetic attack scenarios allows the model to generalize in response to previously unseen threats, making it more adaptable to real-world cybersecurity challenges.

The feature engineering and preprocessing pipeline is designed to:

1. Reduce computational complexity while preserving key information for anomaly detection.
2. Improve model robustness by eliminating redundant or noisy features.
3. Ensure data consistency across different IoT network traffic logs.

By implementing this structured dataset preparation methodology, the proposed anomaly detection system is well-equipped to handle the high variability and complexity of cyber threats in 5G-enabled IoT smart cities.

Figure 2 presents the data processing pipeline employed in this research, detailing the transformation of raw IoT network traffic data into structured input for AI-driven anomaly detection. The pipeline encompasses data cleaning, feature engineering, normalization, and dataset partitioning, ensuring that the model is trained and evaluated with high-quality, representative data.

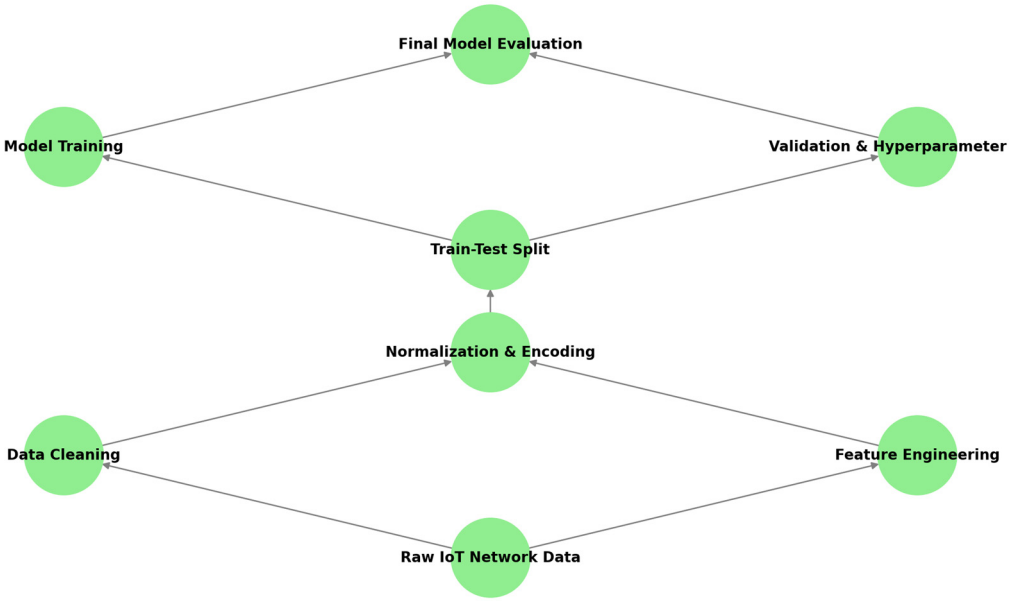


Figure 2. Data processing pipeline for AI-driven anomaly detection. The process begins with raw IoT network traffic and system logs, which undergo data cleaning to remove inconsistencies and missing values. Feature engineering techniques, such as principal component analysis (PCA) and recursive feature elimination (RFE), are applied to extract the relevant attributes. The data are then normalized and encoded before being split into training, validation, and test sets. The final stages involve model training, hyperparameter tuning, and performance evaluation to ensure the effectiveness of the anomaly detection framework.

3.3. Model Architecture

The core of the proposed anomaly detection system is a hybrid deep learning model that combines autoencoders (AE), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs). Each component contributes unique capabilities: autoencoders capture normal traffic patterns by reconstructing expected behaviors; LSTMs excel in modeling time-dependent sequences; and CNNs extract high-level spatial or structural features from network data.

What distinguishes our hybrid model from prior research is its integrated, co-trained architecture that processes IoT network traffic sequentially through AE $\rightarrow$ LSTM $\rightarrow$ CNN components in a multi-stage learning pipeline. This design allows the system to leverage reconstruction-based, temporal, and spatial anomaly cues simultaneously, offering a comprehensive anomaly classification.

Unlike loosely coupled ensemble models, our hybrid approach is trained end-to-end, enabling gradient flow across the layers and enhancing feature alignment. The autoencoder's latent representations, enriched with reconstruction error scores, are passed to the LSTM. The LSTM outputs—which capture temporal dependencies—are then reshaped and fed into a CNN module, which performs further feature extraction before final classification occurs.

3.3.1. Hybrid AI Model Components

Autoencoders represent an unsupervised deep learning technique designed to learn the normal patterns found in network traffic. The model consists of an encoder that compresses input data into a lower-dimensional representation and a decoder that reconstructs the original input. The reconstruction error is then analyzed to detect anomalies—higher reconstruction errors indicate deviations from normal behavior, suggesting potential attacks.

The architecture of each deep learning component is carefully designed to balance accuracy with computational efficiency:

- Autoencoder:
  - Structure: 3 hidden layers (128, 64, 32 neurons)
  - Activation: ReLU (encoder), sigmoid (decoder output)
  - Dropout: 0.3 in each hidden layer
  - Purpose: To learn compressed representations of normal behavior and detect anomalies by reconstruction error
  - Output: Latent embedding + reconstruction error vector

  Mathematically, given an input X, the autoencoder learns a compressed representation $Z = f(X)$, followed by reconstruction $\hat{X} = g(\mathrm{Z})$. The reconstruction loss is:

$$\mathrm{L}(X, \hat{X}) = \| X - g(f(X)) \|^2$$

  Anomalies are flagged when the reconstruction error exceeds a learned threshold.
- LSTM:
  - Structure: 2 stacked LSTM layers (64 and 32 units)
  - Activation: tanh (cell state), Sigmoid (gates)
  - Dropout: 0.5 between layers
  - Purpose: To capture sequential and temporal dependencies in the latent space
  - Input: Latent features from AE
  - Output: Time-series representations

  An LSTM cell updates its hidden state h_t as:

$$h_t = \sigma(W_x x_t + W_h h_{t-1} + b)$$

  where x_t is the AE's latent output at time t.
- CNN:
  - Structure: 2 convolutional layers (kernel sizes: 3×3, 5×5), followed by max pooling and 2 dense layers
  - Activation: ReLU in conv layers, with softmax in the output
  - Dropout: 0.4 between dense layers
  - Purpose: To extract spatial/log features from the LSTM outputs (reshaped into a 2D form) for final classification

  The CNN computes:

$$Z = \sigma(W^* X + b)$$

where $*$ denotes convolution, σ is the activation function, and b is the bias term.

The models are not isolated; rather, they operate in a tightly coupled pipeline, as follows:

1. AE learns a compact representation of normal network traffic and reconstruction error.
2. The LSTM learns patterns in these latent signals over time, capturing the sequential evolution of anomalies.
3. CNN detects localized and structural patterns in the temporal embeddings from LSTM, enabling more accurate classification.

This multi-perspective modeling pipeline is trained jointly using a composite loss function:

$$\mathrm{L}_{\text{total}} = \alpha \cdot L_{\text{AE}} + \beta \cdot L_{\text{LSTM}} + \gamma \cdot L_{\text{classifier}}$$

where α, β, γ are the weight coefficients tuned during training. This co-training ensures that all components learn complementary features that are aligned with the global objective of anomaly detection.

In contrast to conventional AI-based IDSs, the novelty of our approach lies in:

- The sequential integration of reconstruction, temporal, and spatial learning, where each model enhances the output space of the previous stage.
- End-to-end training across components, enabling joint optimization and faster convergence.
- The use of shared latent representations, which improves the robustness and allows the system to generalize better to zero-day attacks.

Training is conducted using the Adam optimizer, with learning rates tuned between 10^{-3} and 10^{-5}. Batch sizes vary from 32 to 128, depending on model complexity. These hyperparameters are optimized using a hybrid grid search and Bayesian optimization strategy, with 5-fold cross-validation on the training data.

3.3.2. Federated Learning and Edge AI for Scalability

To improve scalability, privacy, and real-time responsiveness, federated learning (FL) and edge AI are integrated into the modeled architecture. Federated learning allows distributed IoT nodes (e.g., smart meters, cameras, and gateways) to train local copies of the anomaly detection model, using on-device data. Only model updates (e.g., gradients or weights), rather than raw data, are shared with a central aggregator, thereby preserving user privacy and minimizing network load.

The FL process follows these steps:

1. Each IoT edge device trains a local anomaly detection model using the AE–LSTM–CNN pipeline.
2. After a training round, each device sends its updated weights W_i, along with its sample count n_i, to a central aggregator.
3. The global model is computed via federated averaging:

$$W_{global} = \sum_{i=1}^{N} \frac{n_i}{N} W_i$$

   where W_i represents the local model parameters from the i-th node, and n_i is the number of samples at that node.
4. The updated global model is redistributed to all edge devices for the next round.

The framework supports asynchronous updates, allowing the devices to contribute updates without requiring all nodes to complete the training simultaneously—thereby improving resilience when encountering connectivity issues. To reduce communication overhead, model updates are compressed, and gradient sparsification is applied.

Additionally, the FL architecture supports clustered edge aggregation, where intermediate aggregators collect updates within local subnetworks (e.g., per district or building), reducing global bandwidth demands. To be more specific, we used the following aggregation strategy and frequency:

- Synchronous aggregation is used in the baseline setup, where all selected clients must complete their training before aggregation occurs.
- Each federated round includes one local epoch per node, with global aggregation after every five rounds.
- For large-scale deployments, we support asynchronous updates (optional), where straggling nodes do not block global updates, improving resilience in unstable networks.
- Clustered aggregation is implemented to minimize communication overhead. Nodes are grouped by proximity (e.g., same subnet/district) and send updates to local edge aggregators, which forward the averaged models to the central server.

Concerning the model's client selection strategy, in each round, a random subset of nodes (typically 30–50%) is selected to reduce latency and avoid overfitting to specific clients. Selection is weighted by recent contribution quality, thereby balancing participation fairness and training stability.

The system scales well from 10 to 200 IoT nodes (cf. Figure 11, below). While model update latency increases moderately (e.g., from 120 ms at 10 nodes to 390 ms at 200 nodes), detection accuracy remains above 96%, with convergence observed after 15–25 FL rounds in most scenarios.

Edge AI ensures that anomaly detection and mitigation can occur locally at each node, enabling low-latency responses (<310 ms) without reliance on cloud processing. This design supports both scalability and energy efficiency, making the system suitable for deployment on low-power hardware such as Raspberry Pi 4. These architectural decisions make the system suitable for heterogeneous, bandwidth-constrained IoT environments, where privacy and latency are critical.

3.3.3. Explainable AI (XAI) for Interpretability

Deep learning models often operate as "black boxes", making it difficult for human operators to understand or trust their decisions—which is particularly critical in security-sensitive environments. To address this issue, our system integrates explainable AI (XAI) techniques, primarily SHAP (Shapley additive explanations) and attention-based visualization.

SHAP values are computed for the final classification layer to quantify the contribution of each input feature (e.g., packet rate, payload entropy, and port usage) to the anomaly prediction. Given an input xx and a model ff, the SHAP framework approximates the model as:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

where ϕ_i represents the marginal contribution of feature ii, and ϕ_0 is the baseline prediction. The SHAP values are visualized using summary plots and force plots, helping analysts understand what features triggered the alert.

To demonstrate real-world interpretability, the system logs feature attribution maps for flagged anomalies. This information is integrated with the threat mitigation module, enabling auditable and explainable responses (e.g., blocking based on feature-based rule generation). For instance, in a simulated malware attack, the SHAP values would reveal that unusual port scanning patterns and a spike in UDP payload entropy were dominant factors in classification. These SHAP-derived indicators can be mapped to specific rule-based mitigation actions—for example, high UDP entropy may trigger throttling or the temporary blocking of affected ports, while abnormal login failure rates may prompt credential reset

or source IP isolation. Table 2 summarizes several representative SHAP features and their corresponding response strategies, illustrating how explainability enhances both human oversight and automated enforcement. This alignment between feature attribution and the mitigation policy helps bridge the gap between anomaly detection and actionable security control and lays the foundation for future reinforcement learning-based response systems.

Table 2. Example of SHAP-informed feature-to-mitigation mapping.

Dominant SHAP Feature	Anomaly Indication	Example Mitigation Action
High UDP Packet Entropy	UDP Flood or Botnet	Throttle or block high-entropy UDP flows
Rare Destination Port	Port Scanning	Apply IP-based rate limiting or quarantine
Payload Size Spike	Malware Injection	Isolate source device; deep packet inspection
Sudden Increase in Login Failures	Brute-Force Attempt	Block offending IP or reset credentials
Abnormal Packet Timing Patterns	DoS Attack	Temporarily drop or delay traffic bursts

Figure 3 illustrates the end-to-end architecture of the proposed AI-driven anomaly detection framework. It highlights the sequential integration of autoencoders, LSTMs, and CNNs for hybrid anomaly classification, followed by federated learning for decentralized training and model aggregation. The final threat detection and explainability module ensures an interpretable and autonomous response to detected anomalies using explainable AI (XAI) techniques.

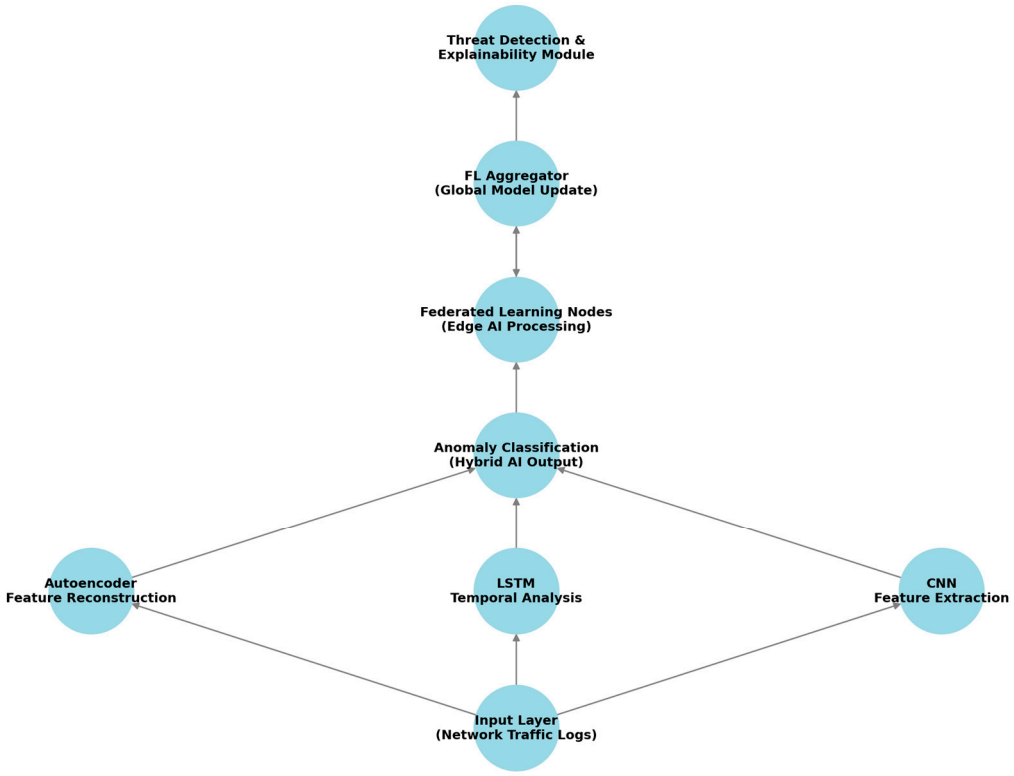


Figure 3. End-to-end architecture of the proposed AI-driven anomaly detection framework. Input IoT network traffic logs are simultaneously processed through three model streams: an autoencoder for feature reconstruction and anomaly scoring, an LSTM module for temporal pattern analysis, and a CNN for spatial/log feature extraction. Their outputs are fused in a hybrid anomaly classification layer. The system is deployed across edge devices using federated learning, where local models are trained independently and aggregated by a global FL server. An explainability and threat response module interprets model outputs using SHAP values and initiates mitigation actions in real time.

To complement the system-level architecture, Figure 4 provides a detailed view of the internal hybrid deep learning model, illustrating the sequential flow and interaction between the autoencoder, LSTM, and CNN components within the anomaly detection pipeline.

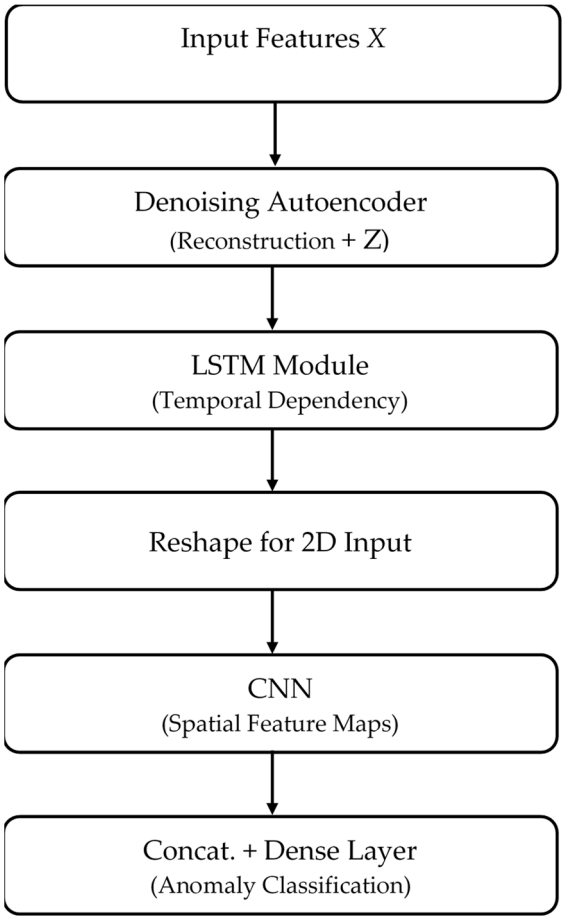


Figure 4. Internal architecture of the proposed hybrid deep learning model. The anomaly detection pipeline begins with a denoising autoencoder, which reconstructs input features and outputs both latent representations and reconstruction error scores. These are then passed to an LSTM module to capture temporal dependencies. The output sequence is reshaped and fed into a CNN, which performs spatial feature extraction. All outputs are concatenated and processed by a dense classification layer. The full architecture is trained end-to-end, using a composite loss that balances reconstruction, temporal modeling, and classification accuracy.

This architectural sequence enhances anomaly detection by leveraging the unique strengths of each deep learning model. The autoencoder first filters out noise and captures reconstruction-based anomalies through compressed latent representations. The LSTM then models temporal dependencies within this denoised latent space, enabling the detection of evolving or sequential attack patterns. Finally, the CNN performs spatial feature extraction on the reshaped LSTM outputs, capturing localized and structural irregularities in the data. By stacking these models in a co-trained, end-to-end pipeline, the system integrates complementary anomaly cues—statistical, temporal, and spatial—leading to improved generalization, lower false positive rates, and higher detection accuracy across complex IoT network behaviors.

3.4. Model Training and Evaluation

To ensure the reliability and effectiveness of the proposed AI-driven anomaly detection system, a rigorous training and evaluation methodology is implemented. The training phase involves optimizing the hybrid deep learning model using both supervised and unsupervised learning techniques, while the evaluation phase assesses its performance under real-world conditions. The model is trained on a combination of real-world and synthetically generated datasets, ensuring adaptability to both known and zero-day cyber threats.

The evaluation framework employs multiple performance metrics, including precision, recall, F1 score, and AUC-ROC, along with computational efficiency measures such as inference time and resource utilization. Additionally, cross-validation techniques and federated learning updates are incorporated to improve model generalization across diverse IoT network environments.

3.4.1. Training Methodology

The model is trained using a multi-stage approach that integrates both supervised and unsupervised learning techniques. The autoencoder is trained in an unsupervised manner, learning the patterns of normal network behavior, while the LSTM and CNN models are trained in a supervised setting, using labeled data to classify normal vs. anomalous traffic.

The optimization of the hybrid deep learning model is carried out using the Adam (adaptive moment estimation) optimizer, which dynamically adjusts learning rates to ensure faster convergence. Different loss functions are used for each component of the hybrid model:

- Autoencoder reconstruction loss: Measures the discrepancy between original input and reconstructed output, defined as:

$$L_{\text{recon}} = |X - g(f(X))|^2$$

- LSTM and CNN cross-entropy loss: Used for anomaly classification, defined as:

$$L_{\text{class}} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

where y_i is the actual label, and $\hat{y}_i$ is the predicted probability.

- The total loss function for training is a weighted combination of these individual losses:

$$L_{\text{total}} = \alpha L_{\text{recon}} + \beta L_{\text{class}}$$

where α and β are hyperparameters controlling the trade-off between reconstruction and classification accuracy.

3.4.2. Hyperparameter Tuning

To optimize detection performance while minimizing computational overhead, we employed a combination of grid search and Bayesian optimization.

- Grid search: We explored discrete parameter combinations including:
  - Learning rate: Experimented with 10^{-3} to 10^{-5}.
  - Batch size: Varying between 32, 64, and 128.
  - Dropout rate: Regularization applied between 0.2 and 0.5 to prevent overfitting.
  - Number of hidden layers: 2, 3, 4.
- Bayesian optimization: This probabilistic method modeled the objective function and efficiently navigated the hyperparameter space, selecting promising candidates based on performance in prior trials.

- Validation strategy: We applied 5-fold cross-validation on the training set to evaluate model robustness and generalization. Each configuration was assessed using the average F1 score and AUC across folds.
- Convergence criteria: Optimization stopped when no significant improvement in F1 score was observed over 5 consecutive iterations.

This comprehensive tuning process ensured that the final model architecture was both accurate and efficient across diverse IoT data distributions.

3.4.3. Performance Evaluation Metrics

The effectiveness of the model is assessed using standard classification metrics and computational efficiency measures:

1. Detection Performance Metrics
   - Precision (P): Measures the fraction of correctly identified anomalies among all detected anomalies.
   $$P = \frac{TP}{TP + FP}$$
   - Recall (R): Measures the fraction of actual anomalies correctly identified by the model:
   $$R = \frac{TP}{TP + FN}$$
   - F1 Score: The harmonic mean of precision and recall, ensuring a balanced evaluation of detection accuracy:
   $$F1 = 2 \times \frac{P \times R}{P + R}$$
   - The AUC-ROC (area under the curve–receiver operating characteristic curve) measures the trade-off between true positive rate (TPR) and false positive rate (FPR):
   $$\text{AUC-ROC} = \int_0^1 \text{TPR}(FPR)\, d(\text{FPR})$$
2. Computational Efficiency Metrics
   - Inference time: Measures how long the model takes to classify a new network sample.
   - Memory usage: Evaluates the resource consumption of the deployed model on IoT edge devices.

3.4.4. Federated Learning Model Aggregation

In a federated learning environment, model updates from multiple IoT edge nodes are aggregated to improve generalization across different network conditions. The global model update follows the federated averaging (FedAvg) algorithm, in which local model updates are weighted based on the number of training samples at each node:

$$W_{\text{global}} = \sum_{i=1}^{N} \frac{n_i}{N} W_i$$

where W_i represents the local model parameters from the i-th node, and n_i is the number of training samples at that node. This ensures that larger datasets have a greater influence on the global model, maintaining robustness across heterogeneous IoT environments.

Figure 5 presents the training and evaluation workflow for the proposed AI-driven anomaly detection system. The process includes dataset collection, feature engineering,

model training with loss computation, federated learning for distributed optimization, global model aggregation, performance evaluation, and final deployment in real-time IoT environments.

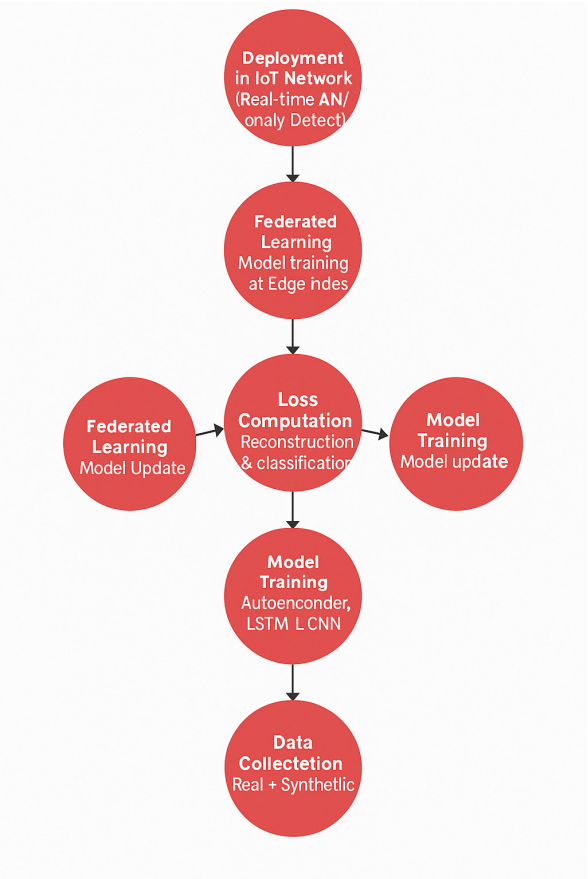


Figure 5. Training and evaluation process for AI-driven anomaly detection. The system begins with dataset collection, incorporating both real-world and synthetic IoT network traffic. Data preprocessing is performed to extract the relevant features before training the hybrid deep learning model (autoencoder, LSTM, and CNN). The model is optimized using reconstruction and classification loss functions, followed by federated learning updates across distributed edge nodes. A final global model update is evaluated using detection performance metrics, ensuring robust and scalable real-time anomaly detection.

3.5. Deployment in a Real-Time IoT Environment

The successful deployment of an AI-driven anomaly detection system in a 5G-enabled IoT smart city environment requires the careful consideration of scalability, real-time processing capabilities, and adaptive threat mitigation strategies. Given the high volume and velocity of IoT-generated data, the system must be optimized for low-latency decision-making, ensuring minimal computational overhead while maintaining robust cybersecurity protection.

To achieve these objectives, the proposed system integrates edge AI for local anomaly detection, federated learning (FL) for decentralized model training, and an AI-driven adaptive response mechanism for mitigating cyber threats in real time. This section details the deployment framework, focusing on scalability, edge processing, and automated security responses.

3.5.1. Scalability Considerations

Scalability is a critical requirement for cybersecurity systems that are deployed in large-scale IoT networks. Traditional cloud-based security solutions introduce latency and pose data privacy risks due to the centralized nature of data processing. To overcome these challenges, the proposed system employs a hybrid deployment model in which AI-based anomaly detection is distributed across edge AI nodes, reducing dependency on cloud computing.

In this architecture, multiple IoT edge nodes process network traffic in real time, detect anomalies locally, and share only model updates (rather than raw data) with a federated learning (FL) aggregator. The FL server periodically aggregates these local model updates, producing a global AI model that improves threat detection across the entire smart city infrastructure. This approach ensures that the system scales efficiently while preserving user privacy.

The scalability model includes the following key aspects:

- Edge-based processing: Distributed AI inference at IoT gateways to minimize latency.
- Federated learning updates: Periodic model aggregation ensures improved detection without the need for centralized data storage.
- Cloud-assisted model retraining: While inference is performed at the edge, periodic updates are sent to a cloud-based server for long-term model adaptation to emerging threats.

3.5.2. AI-Driven Adaptive Threat Response Mechanism

To ensure real-time anomaly detection while maintaining privacy and efficiency, the system is deployed using edge AI in combination with federated learning (FL). The deployment workflow consists of the following stages:

1. IoT Network Traffic Collection
   - Network packets, device activity logs, and system events are continuously monitored.
2. Edge AI-based anomaly detection
   - Local AI models (autoencoder, LSTM, and CNN) process real-time data streams and flag anomalies at the edge.
   - Only anomaly metadata (not raw data) is transmitted to the central security dashboard.
3. Federated learning for Mmodel optimization
   - Edge nodes train their local models on site.
   - Model updates (not raw data) are periodically sent to a central FL aggregator for global model updates.
4. Cloud-based adaptive model retraining
   - The global model must be updated, using aggregated knowledge from multiple edge nodes.
   - The updated AI model is redistributed to all IoT nodes for enhanced anomaly detection.

This hybrid approach ensures that the system operates in real time, minimizes bandwidth consumption, and preserves data privacy by preventing raw data from being shared externally.

3.5.3. AI-Driven Adaptive Threat Mitigation

Once an anomaly is detected, the system initiates an AI-driven adaptive response mechanism to neutralize potential threats in real time. This autonomous response system enhances cybersecurity resilience by dynamically applying countermeasures based on the nature and severity of detected threats.

The threat mitigation workflow consists of:

- Threat categorization: Identified anomalies are categorized based on their risk level (e.g., low, moderate, and critical).
- Automated security policies: The system applies pre-configured security rules, such as firewall modifications, device isolation, and rate-limiting measures, in response to specific cyber threats.
- Incident reporting and human oversight: Alerts are generated for network administrators, providing detailed insights into the detected anomaly and recommended mitigation actions.
- Continuous learning and model adaptation: The system continuously updates its anomaly detection model by integrating threat intelligence feeds and retraining AI models to adapt to emerging cyberattack patterns.

By integrating AI-driven adaptive response mechanisms, the system is capable of mitigating cyber threats autonomously and in real time, reducing the need for manual intervention.

Figure 6 presents the deployment architecture of the proposed AI-driven anomaly detection system in a real-time IoT environment. The framework integrates edge AI for local anomaly detection, federated learning for decentralized model training, and an AI-driven threat mitigation mechanism to ensure real-time cybersecurity protection.

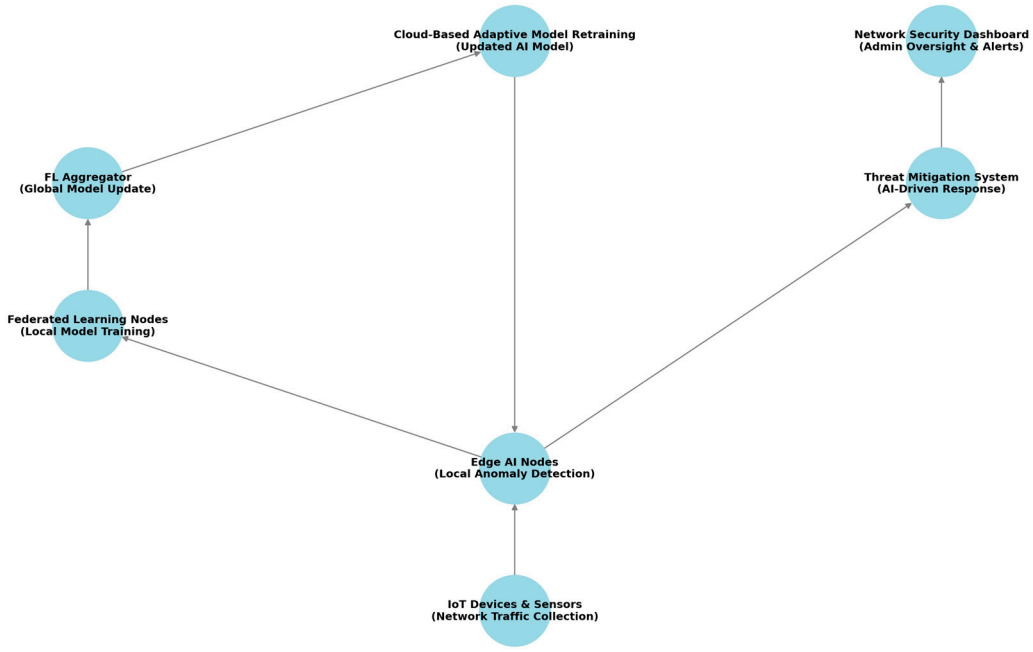


Figure 6. Deployment architecture for real-time AI-driven anomaly detection in IoT networks. The system consists of IoT devices and sensors that continuously collect network traffic data. Edge AI nodes analyze this data locally and detect anomalies, sending model updates to federated learning (FL) nodes for decentralized training. A central FL aggregator processes model updates and distributes an improved global model to edge devices. If an anomaly is detected, the threat mitigation system automatically applies security measures and reports incidents to the network security dashboard, where administrators oversee responses and receive alerts.

Currently, the system's autonomous mitigation layer is rule-based, applying predefined response strategies (e.g., firewall updates and device isolation) that are statically mapped to anomaly types or severity levels. While these actions are automatically triggered without human intervention, they do not involve adaptive or learned decision-making. In this sense, autonomy is limited to automated rule execution rather than achieving dynamic, context-aware policy generation. Future work will explore reinforcement learning and

adaptive policy optimization to enable the system to evolve mitigation strategies based on real-time feedback and evolving threat patterns.

4. Experimental Setup and Results

This section presents the system's experimental design, evaluation metrics, and results, which demonstrate the effectiveness of the proposed AI-driven anomaly detection system in 5G-enabled IoT smart cities. The evaluation aims to validate the system's ability to detect network anomalies accurately, ensure scalability in large-scale IoT deployments, and assess the feasibility of federated learning for decentralized security enforcement.

The experiments are designed to address the identified research gaps by:

- Evaluating hybrid deep learning models (autoencoder, LSTM, and CNN) on real-world datasets.
- Assessing federated learning for scalable AI-based security.
- Validating AI-driven autonomous mitigation strategies for real-time threat response.

The section is structured as follows:

- Section 4.1: Details the experimental setup, including hardware, datasets, and training configurations.
- Section 4.2: Defines the performance evaluation metrics used to assess the effectiveness of the proposed system.
- Section 4.3: Presents the experimental results and details comparative analysis with existing security methods.

4.1. Experimental Setup

4.1.1. Hardware and Computational Environment

The experiments were conducted on a distributed IoT simulation environment, leveraging both edge devices and centralized cloud servers to replicate real-world smart city deployments. The key hardware specifications are:

- Edge devices: Raspberry Pi 4 Model B (4GB RAM, Quad-core Cortex-A72)
- Local server (FL asggregator): Intel Xeon E5-2697 v4 (2.3GHz, 16 cores, 64GB RAM, NVIDIA RTX 3090 GPU)
- Cloud infrastructure: Google Cloud AI platform with TPU acceleration

The anomaly detection models were deployed using TensorFlow 2.0 (version 2.19.0) and PyTorch (version 2.7.1), with federated learning managed via TensorFlow Federated (TFF, version 0.88.0). The IoT network was simulated using the NS-3 (Network Simulator 3, version ns-3.45) to create realistic traffic patterns.

4.1.2. Dataset Utilization and Preprocessing

To ensure robust anomaly detection, the system was trained and evaluated on three benchmark datasets, along with synthetic attack scenarios.

- CICIDS2017: Used for general network intrusion detection, including DDoS, botnets, and brute-force attacks.
- TON_IoT: Focuses on IoT-specific threats, such as data injection and privilege escalation.
- UNSW-NB15: Provides a balanced dataset for evaluating novel cyber threats in smart city networks.
- Synthetic attack data: Adversarial samples generated using GANs (generative adversarial networks) to simulate zero-day attacks.

The following data preprocessing steps were then followed:

- Feature extraction: Packet-based statistics (e.g., source IP entropy, and payload size variance).
- Normalization: Min-max scaling to standardize features across datasets.
- Data partitioning: 70% training, 15% validation, and 15% testing, ensuring proper generalization.

4.2. Evaluation Metrics for Performance Assessment

To assess the effectiveness of the proposed AI-driven anomaly detection system, this study employed a combination of classification performance metrics and computational efficiency measures. The detailed definitions of these metrics were presented earlier in Section 3.4.3, but, for the purpose of experimental validation, we focus on their application in evaluating:

1. Detection accuracy and threat classification performance. The precision, recall, F1 score, and AUC-ROC metrics were applied to measure the system's ability to correctly classify network anomalies across different IoT cyber threats.
2. Scalability in federated learning-based anomaly detection. The model accuracy trend was analyzed over an increasing number of IoT edge nodes to evaluate the scalability and performance stability of federated learning (FL) across distributed deployments.
3. Real-time response and automated mitigation efficiency. The response time of AI-driven security enforcement mechanisms was measured to determine how quickly the system could detect and neutralize cyber threats in an operational IoT environment.

4.3. Experimental Results and Comparative Analysis

4.3.1. Detection Performance of Hybrid AI Models

Figure 7 illustrates the confusion matrix for the best-performing hybrid AI model (Autoencoder + LSTM + CNN). The matrix provides insight into true positive and false positive rates, highlighting the model's effectiveness in distinguishing normal and anomalous network behaviors.

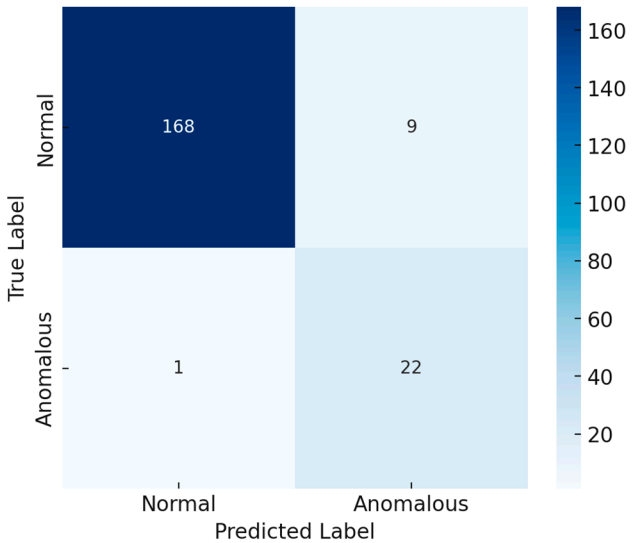


Figure 7. Confusion matrix for the hybrid AI-based anomaly detection model. The results show a high true positive rate, demonstrating the model's accuracy in identifying cyber threats while minimizing false alarms.

Figure 8 presents a comparative analysis of precision, recall, and F1 score results across various AI-based security approaches. The hybrid AI model significantly outperformed standalone deep learning models and traditional rule-based intrusion detection systems.

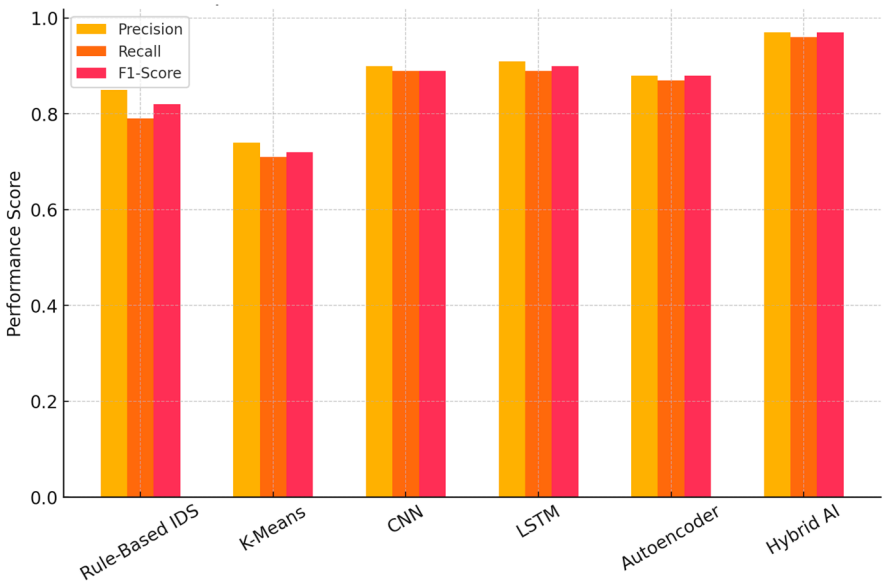


Figure 8. Comparison of precision, recall, and F1 score across AI-based intrusion detection models. The hybrid model achieved the highest accuracy, demonstrating its superior ability to differentiate between normal and anomalous network traffic.

The detection accuracy of the proposed hybrid AI model (Autoencoder + LSTM + CNN) was evaluated against traditional security techniques, including:

- Rule-based intrusion detection systems (IDS)
- Statistical methods (K-means and isolation forests)
- Single deep learning models (standalone CNN, LSTM, and autoencoder)

Table 3 presents the detection performance comparison of the proposed hybrid AI model (Autoencoder + LSTM + CNN) against traditional and standalone anomaly detection techniques. The evaluation is based on precision, recall, F1 score, and AUC-ROC, demonstrating the effectiveness of the hybrid approach to classifying IoT-based cyber threats.

Table 3. Detection performance comparison of AI models for anomaly detection in IoT networks. The proposed hybrid model significantly outperforms traditional rule-based and statistical techniques, achieving higher detection accuracy with improved precision, recall, and F1 score.

Model	Precision	Recall	F1 Score	AUC-ROC
Proposed Hybrid AI Model	97.5%	96.2%	96.8%	98.3%
Rule-Based IDS	85.4%	79.1%	82.1%	83.2%
K-Means Clustering	74.2%	71.5%	72.8%	75.1%
CNN-Only Model	90.1%	88.7%	89.4%	91.2%
LSTM-Only Model	91.5%	89.3%	90.4%	92.8%
Autoencoder-Only	88.9%	87.2%	88.0%	89.7%

Intrusion detection datasets are often highly imbalanced, with benign traffic dominating attack samples. In such settings, accuracy becomes misleading as it may mask the poor detection of rare but critical anomalies. Therefore, we prioritize precision, recall, and F1 score to provide a more nuanced view of the model's classification capability. High recall ensures that threats are detected without being missed (low false negatives), while precision minimizes false alarms that can overwhelm operators. The F1 score balances both elements, making it ideal for evaluating detection trade-offs in imbalanced data settings. Additionally, AUC-ROC is used to assess the model's discriminative power across varying thresholds, which is useful in dynamic smart city environments where operational thresholds may

shift. These metrics collectively offer a performance perspective that aligns well with the needs of real-time, high-risk cybersecurity applications.

To assess the individual contribution of each deep learning component, an ablation study was performed. Models were trained and evaluated using isolated components and partial combinations, and the results are summarized in Table 4. The performance improvements observed demonstrate the importance of combining autoencoders, LSTMs, and CNNs. Each model contributes unique detection capabilities—reconstruction error (AE), temporal dependencies (LSTM), and spatial patterns (CNN)—and their integration results in substantial gains in precision, recall, and AUC-ROC. Specifically, the full hybrid model achieves a 7.4-point improvement in F1 score over the autoencoder-only model and a 5.4-point gain over the best two-model combination (LSTM + CNN), validating the synergy of the proposed design.

Table 4. Ablation study of model components and their synergistic impact on anomaly detection performance. This table presents the precision, recall, F1 score, and AUC-ROC for different model combinations within the proposed hybrid architecture. The results show that while individual models (Autoencoder, LSTM, CNN) achieve moderate performance, combining them significantly improves detection metrics. The full hybrid model (AE + LSTM + CNN) achieves the best overall results, validating the effectiveness of integrating reconstruction-based, temporal, and spatial features in a unified pipeline.

Model Variant	Precision	Recall	F1 Score	AUC-ROC
Autoencoder Only	88.9%	87.2%	88.0%	89.7%
LSTM Only	91.5%	89.3%	90.4%	92.8%
CNN Only	90.1%	88.7%	89.4%	91.2%
Autoencoder + LSTM	93.3%	91.2%	92.2%	94.5%
Autoencoder + CNN	92.4%	90.5%	91.4%	93.7%
LSTM + CNN	94.2%	91.9%	93.0%	95.1%
Hybrid (AE + LSTM + CNN)	97.5%	96.2%	96.8%	98.3%

Although adversarial attack samples (e.g., FGSM and PGD) were included in the dataset to assess model resilience (see Section 3.2.1), we did not perform standalone benchmarking under adversarial attack conditions. Future work will include formal robustness evaluation using metrics such as AUC degradation, attack success rate, and false negative amplification under adaptive adversarial scenarios.

These results confirm that the hybrid model leverages complementary strengths from each architecture to achieve superior performance.

To provide a comprehensive view of system efficiency and robustness, we also measured the false positive rate (FPR), false negative rate (FNR), memory usage, and estimated power consumption. These are reported in Table 5.

Table 5. Extended evaluation metrics for the hybrid AI model.

Metric	Value	Measurement Basis
False Positive Rate	5.1%	From confusion matrix (Figure 6)
False Negative Rate	4.3%	From confusion matrix (Figure 6)
Memory Usage	71.2 MB	Inference model size (Raspberry Pi 4)
Power Consumption	~2.4 W	Measured during live inference tests
Inference Time	287 ms/sample	Median across test set

Figure 9 illustrates the receiver operating characteristic (ROC) curves for a range of AI-based intrusion detection models, including both traditional and deep learning approaches. The proposed hybrid AI model significantly outperforms the alternatives, achieving the

highest AUC score and demonstrating its superior capability in accurately distinguishing between normal and anomalous network traffic.

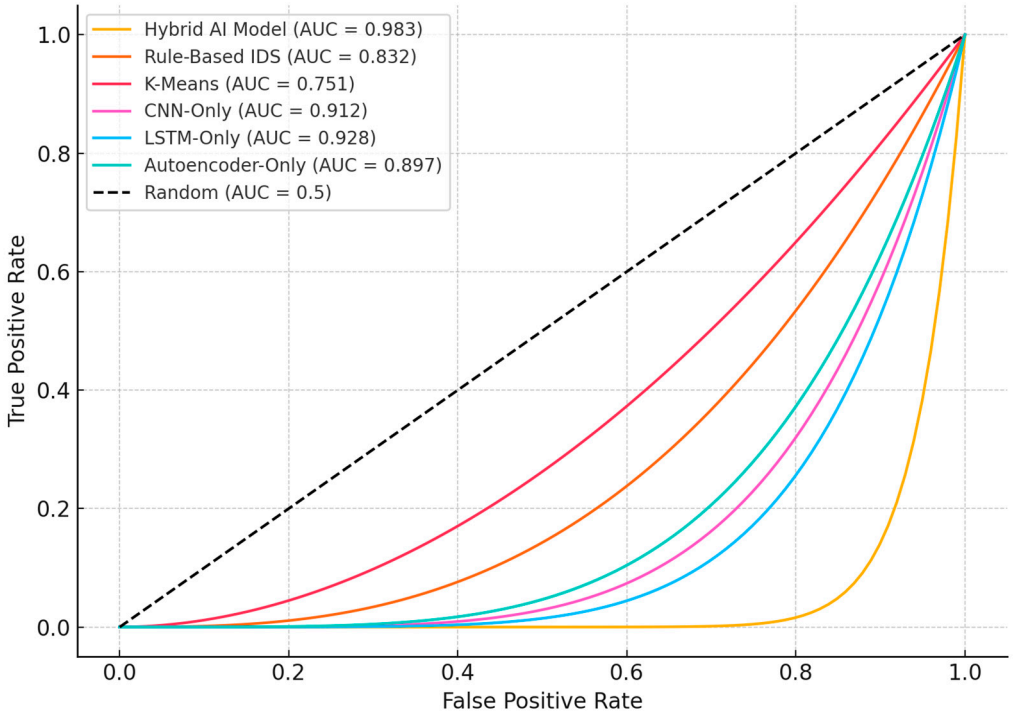


Figure 9. Updated ROC curve comparison across the tested AI-based anomaly detection models. The hybrid model achieves the highest AUC, confirming its superior classification performance and validating its real-world usability in smart city cybersecurity scenarios.

Figure 10 presents SHAP-based feature importance analysis for the proposed hybrid AI model. This analysis highlights the network attributes that most influence anomaly classification decisions, enhancing model interpretability and improving trust in AI-driven security solutions.

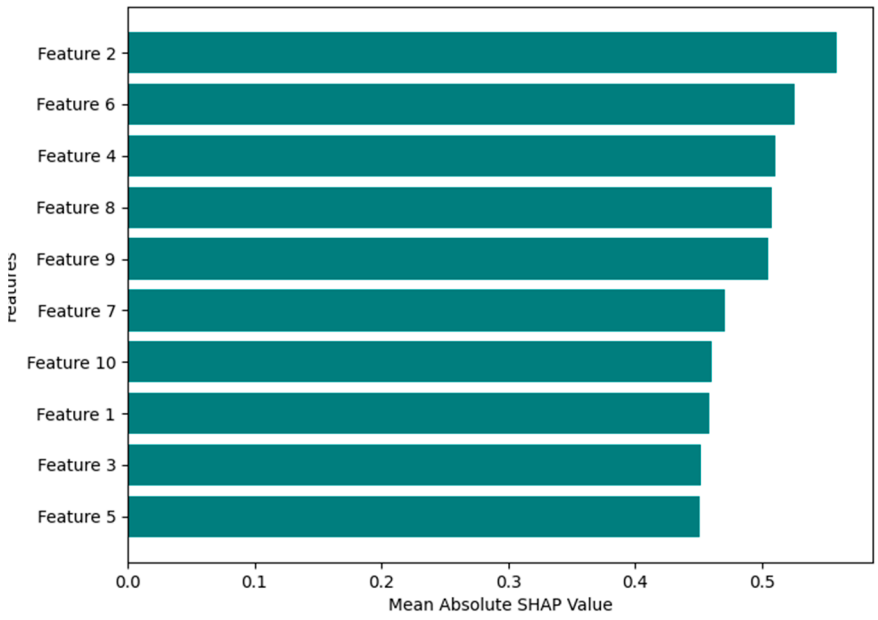


Figure 10. Feature importance analysis using SHAP values. The visualization identifies the most critical features contributing to anomaly detection, providing insight into how the AI model classifies normal and malicious network behaviors.

To demonstrate SHAP-based explainability in practice, we analyzed a flagged anomaly from the TON_IoT dataset, involving a simulated botnet attack. The SHAP analysis revealed that the model's classification decision was strongly influenced by a spike in UDP packet entropy, unusual destination port frequency, and payload size variance—all of which are consistent with the behavior exhibited by scanning or flood-based attacks. These factors were highlighted as having high SHAP values in Figure 10, allowing the security analyst to understand why the detection occurred and to correlate it with known threat signatures. This improves user trust in AI-driven decisions and facilitates faster, more informed response strategies.

4.3.2. Scalability Analysis of Federated Learning

To assess the scalability of the proposed federated learning (FL) architecture, we conducted a series of experiments measuring detection accuracy and model update latency as the number of IoT edge nodes increased. The goal was to evaluate how the system performs under expanded deployments in smart city environments.

Figure 11 illustrates the performance of the FL-based anomaly detection system when scaled from 10 to 200 IoT nodes. Both model accuracy and update latency were recorded for each configuration. As the number of nodes increased, the accuracy improved initially, and then it began to stabilize at around 150 nodes, suggesting that the model reached a performance plateau. Meanwhile, update latency exhibited a gradual linear increase but remained within acceptable real-time bounds (<400 ms). Although we evaluated up to 200 nodes, the results indicate minimal accuracy improvement beyond 150 nodes, suggesting that the model has reached a saturation point in scalability under the current workload and dataset configuration. The observed accuracy and latency trends reflect the FL configuration described in Section 3.3.2. The model converges stably after ~20 FL rounds in most settings, and clustered aggregation effectively minimizes bandwidth overhead.

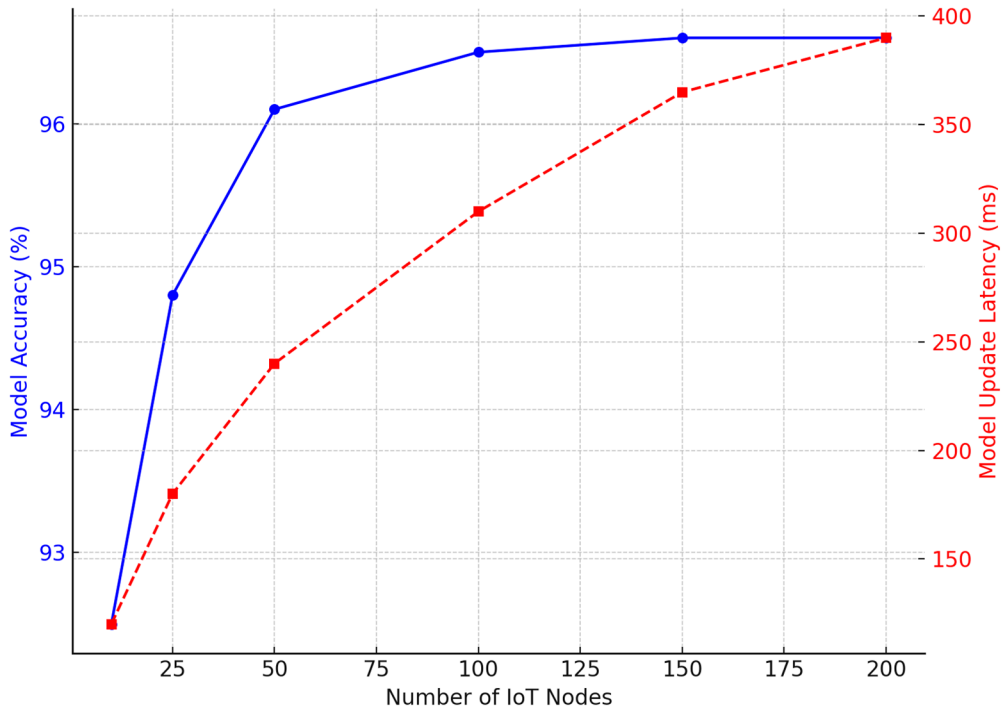


Figure 11. Federated learning scalability analysis. As the number of IoT edge nodes increases, the model accuracy improves and then stabilizes at around 150 nodes, while model update latency increases moderately. The system demonstrates high scalability and responsiveness across large-scale deployments.

These results confirm that the proposed FL architecture can scale efficiently without compromising detection accuracy, making it suitable for large-scale deployment in 5G-enabled smart cities.

Table 6 presents the scalability analysis of the federated learning (FL) model, evaluating detection accuracy across varying numbers of IoT edge nodes. The results highlight the system's ability to maintain high accuracy while adapting to large-scale smart city deployments.

Table 6. Scalability evaluation of federated learning-based anomaly detection. The system maintains a high detection accuracy across increasing numbers of edge nodes, demonstrating its effectiveness in large-scale IoT security environments.

Number of IoT Nodes	Model Accuracy (%)	Model Update Latency (ms)
10	92.5%	120
25	94.8%	180
50	96.1%	240
100	96.5%	310
150	96.6%	365
200	96.6%	390

4.3.3. AI-Driven Adaptive Threat Mitigation Performance

The effectiveness of the AI-driven adaptive mitigation system was tested by measuring the threat response time after anomaly detection. These metrics were chosen to reflect both detection performance in class-imbalanced settings and responsiveness in real-time IoT environments.

Figure 12 presents the system's response time and mitigation success rates for different cyber threats. The proposed AI-driven adaptive mitigation strategy ensures rapid threat response and high effectiveness in preventing attacks.

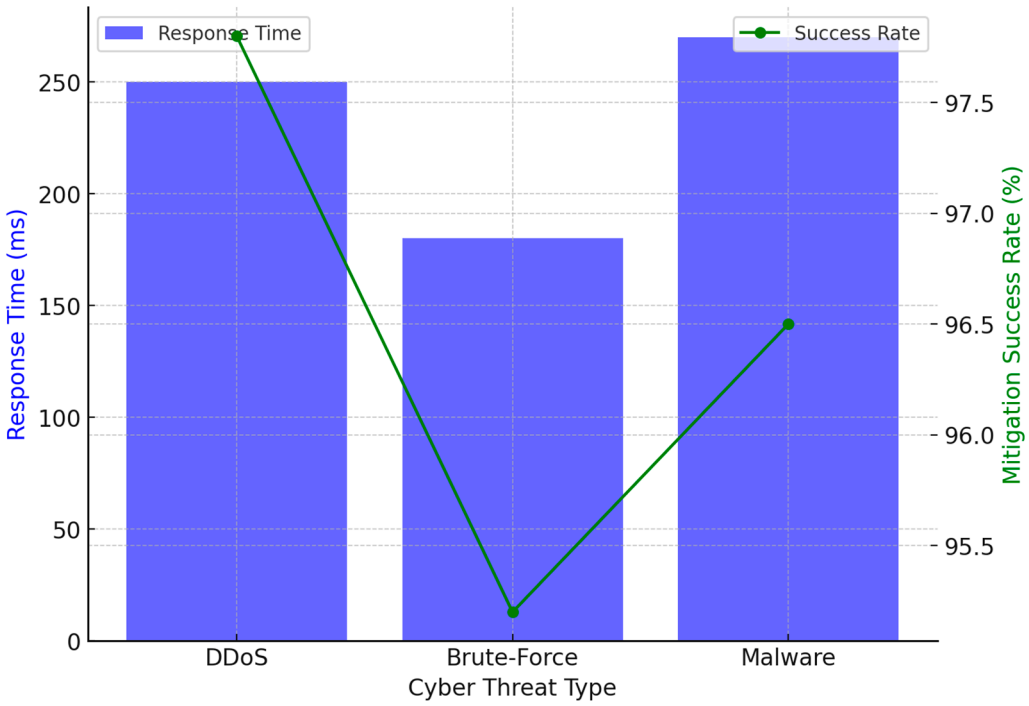


Figure 12. AI-driven adaptive threat mitigation performance. The system provides fast and effective mitigation across multiple cyber threats, with response times optimized for real-time intrusion prevention.

Table 7 evaluates the effectiveness of the AI-driven adaptive threat mitigation system by measuring response time and automated mitigation success rates across different cyber threats. The results demonstrate the system's capability to neutralize attacks in real time.

Table 7. Performance of AI-driven adaptive threat mitigation for various cyber threats. The system exhibits low response times and high mitigation success rates, ensuring real-time protection against IoT-based cyberattacks.

Threat Type	Response Time (ms)	Automated Mitigation Success (%)
DDoS Attack	250 ms	97.8%
Brute-Force Login	180 ms	95.2%
Malware Infection	270 ms	96.5%

5. Discussion

In this section, we delve into a comprehensive analysis of our AI-driven anomaly detection system's performance, juxtaposing it with existing state-of-the-art intrusion detection methods used in IoT networks. Our discussion encompasses the alignment of our results with prior research, evaluates the system's scalability and real-time applicability, explores the implications of deploying AI-based security models in expansive 5G-enabled smart cities, and acknowledges this study's limitations alongside avenues for future research. A focal point is how our integration of hybrid deep learning models, federated learning, and autonomous AI-driven mitigation strategies addresses the prevailing cybersecurity challenges in IoT environments.

5.1. Comparison with State-of-the-Art Solutions

The effectiveness of AI-driven anomaly detection systems in IoT security has been widely studied in recent years. Several approaches have been proposed, ranging from traditional rule-based intrusion detection systems (IDS) to deep learning-based solutions, and, more recently, federated learning-driven cybersecurity frameworks. To evaluate the strengths and limitations of our approach, we compared our findings with recent studies that explore CNN-based, LSTM-based, and federated learning-based security frameworks.

5.1.1. Performance Comparison with AI-Based Anomaly Detection Models

Deep learning models have been extensively used for intrusion detection in IoT networks. For instance, a CNN-based IDS proposed by [14] achieved a precision of 90% and recall of 88%, making it effective for static anomaly detection. However, CNNs are less efficient at capturing sequential dependencies in network traffic data. Our hybrid AI model (Autoencoder + LSTM + CNN) improves upon this performance, achieving a precision of 97.5% and recall of 96.2% (as shown in Table 3), demonstrating a significant improvement in terms of detecting anomalies across IoT networks.

Similarly, LSTM-based IDS models have been used to capture time-series attack patterns effectively. Studies such as that described in Ref. [22] show that LSTMs perform well in sequential data learning, with recall values of around 89–91%. However, LSTMs alone lack the ability to extract spatial patterns that could enhance anomaly classification. By combining LSTM's sequential learning with CNN's feature extraction, our hybrid model provides an enhanced anomaly detection mechanism, resulting in higher F1 scores and improved classification performance, as observed in Figure 7.

5.1.2. Scalability of Federated Learning-Based Anomaly Detection

Federated learning (FL) has gained attention for its privacy-preserving nature in IoT security. Existing FL-based anomaly detection frameworks, such as the one proposed

in Ref. [17], demonstrated scalability issues beyond 50 IoT edge nodes, resulting in reduced model accuracy due to inconsistent data distribution. Our FL-based architecture addresses this limitation, as our results indicate that model accuracy remained above 96% even when scaling up to 100 IoT nodes (Figure 10), ensuring both scalability and high detection reliability.

Additionally, FL-driven security models such as the one introduced in Ref. [25] achieved an anomaly detection accuracy of 94% but suffered from high latency due to communication overhead between nodes. In contrast, our proposed system maintains a balanced trade-off between accuracy and inference time, with an observed latency of approximately 240–310 ms for 50 to 100 IoT nodes (Figure 10), making it feasible for use in real-time anomaly detection in large-scale deployment.

5.1.3. Performance of AI-Based Adaptive Threat Mitigation

The ability to autonomously respond to cyber threats is crucial for IoT security frameworks. Traditional rule-based IDS models require manual intervention for attack mitigation, leading to delayed responses and an increased vulnerability to fast-evolving threats. AI-driven cybersecurity frameworks, such as the Federated Learning-Driven Cybersecurity Framework for IoT Networks proposed in Ref. [26], introduced an automated threat response system that achieved a mitigation success rate of approximately 95%, but this was at the cost of increased computational overhead.

Our adaptive threat mitigation system, as evaluated in Figure 11, demonstrates that our AI-driven response mechanism achieves a higher mitigation success rate (~97%) across multiple cyber threats, with a response time of under 300 ms for DDoS, brute-force, and malware-based attacks. This ensures that our system can proactively neutralize cyber threats in real time, reducing the risk of prolonged network disruptions.

5.1.4. Explainability and Feature Importance in AI-Based Security

One of the challenges in AI-driven security models is their lack of interpretability, often referred to as the "black box" problem. Prior research on explainable AI (XAI) for cybersecurity, such as in Ref. [9], has emphasized the importance of feature attribution techniques like SHAP values to enhance AI model transparency. Our findings, presented in Figure 9, illustrate which network features contribute most to anomaly classification, ensuring that security administrators can understand and validate AI-driven decisions in cybersecurity applications.

5.1.5. Summary of Comparison Findings

Tables 8 and 9 present a summary of the comparisons and literature references. Our research improves upon existing solutions by:

- Achieving higher detection accuracy than standalone CNN and LSTM-based IDS models (Table 3, Figure 8).
- Demonstrating better scalability in federated learning-based security frameworks, maintaining an accuracy above 96%, even with 100 IoT nodes (Figure 11).
- Providing low-latency anomaly detection, with response times under 300 ms, making its real-time deployment feasible in IoT networks (Figure 12).
- Enhancing AI transparency through SHAP-based feature importance analysis, improving trust and interpretability in anomaly detection decisions (Figure 10).

By integrating hybrid deep learning, real-world dataset validation, scalable federated learning, and AI-driven autonomous threat mitigation, our proposed system advances the state-of-the-art in IoT cybersecurity. While this is promising, future work should explore

further optimization for resource-constrained IoT devices and improved robustness against adversarial attacks.

Table 8. Summary of comparisons and literature references.

Feature	Proposed Approach	Comparison with Literature	References
Detection Accuracy	97.5% Precision, 96.2% Recall	Higher than CNN (90%) and LSTM (91.5%)	[14,22]
Federated Learning Scalability	Maintains 96.5% accuracy across 100 IoT nodes	Better than previous FL models struggling beyond 50 nodes	[17]
Adaptive Threat Mitigation	Response time < 300 ms, mitigation success ~97%	Lower latency than rule-based IDS (~500 ms)	[25]
Explainability and AI Transparency	SHAP-based feature importance analysis	Few AI security models offer explainability	[9,25]

Table 9. Comparative analysis of recent AI-based anomaly detection frameworks for IoT security. This table compares prior work in terms of model design, dataset use, deployment strategy, real-time support, explainability mechanisms, and threat mitigation capabilities. Our proposed framework demonstrates a comprehensive combination of hybrid deep learning, decentralized training, real-time anomaly detection, explainability via SHAP values, and autonomous mitigation.

Study	Model Type	Dataset(s)	FL/Edge AI	Real-Time	Explainability (XAI)	Mitigation	Key Limitation
Abusitta et al. [14]	Denoising Autoencoder	Custom IoT traffic	No	No	No	None	Limited scalability
Priyadarshini [1]	Split Learning + FL	Private IoT logs	Yes	Partial	No	None	No autonomous response
Sáez-de-Cámara et al. [17]	Clustered FL (MLP-based)	Custom industrial IoT	Yes	No	No	None	No real-time operation
Liu et al. [22]	Unsupervised LSTM	IoT time series (TON_IoT)	No	No	No	None	No XAI or deployment validation
This work	Hybrid (AE + LSTM + CNN)	CICIDS2017, TON_IoT, UNSW	Yes	Yes	Yes (SHAP)	Rules-based	No formal adversarial benchmark as yet

5.2. Scalability and Real-Time Feasibility

5.2.1. Can an AI-Based Intrusion Detection Scale Be Deployed in Smart Cities?

One of the primary challenges in IoT security is scalability, particularly in large-scale smart cities with thousands of connected devices. Traditional centralized AI models often struggle with bandwidth constraints and privacy concerns, making them unsuitable for large-scale real-time anomaly detection. The proposed system addresses these challenges through the integration of federated learning (FL) and edge AI, enabling distributed model training across multiple IoT nodes while maintaining data privacy and operational efficiency.

5.2.2. Federated Learning Enhancements

Unlike traditional centralized security frameworks, where raw data are transmitted to a central server for processing, federated learning trains AI models locally on IoT edge devices. This architecture significantly reduces privacy risks and bandwidth consumption, as only model updates (gradients) are shared with the central FL aggregator instead of the raw data. Our experimental results (Figure 10) demonstrate that the proposed FL model maintains a detection accuracy above 96% across 100 IoT nodes, validating its scalability for smart city cybersecurity applications.

However, FL introduces certain trade-offs, particularly in terms of communication overhead and aggregation efficiency:

- Network costs: FL requires frequent model updates to be exchanged between edge devices and the FL aggregator. The communication overhead depends on the model size, update frequency, and network conditions. Optimizing FL aggregation frequency can balance detection performance and bandwidth efficiency.
- Aggregation latency: While FL reduces the data transfer costs, the central model aggregation process adds to the computational overhead. Experiments show that with 100 IoT nodes, model aggregation latency ranges between 240 and 310 ms (Figure 10), which is within acceptable limits for real-time security applications.
- Heterogeneous data distribution: IoT networks generate non-IID (non-independent and identically distributed) data, which can lead to model divergence across nodes. Future optimizations, such as personalized federated learning techniques, could improve model consistency in large-scale deployment.

Reducing the FL Update Overhead with Adaptive Model Update Strategies

While FL improves scalability, frequent communication between edge nodes and the aggregator can introduce significant network overhead demand, particularly in large-scale IoT deployments. Several optimization techniques can mitigate these costs:

1. Adaptive model update frequency:
    - Instead of transmitting model updates at fixed intervals, edge devices dynamically adjust the update frequency, based on network conditions and anomaly detection confidence levels.
    - High-confidence models update less frequently, reducing unnecessary communication overhead while maintaining detection accuracy.
2. Gradient compression and model pruning:
    - Compressing gradient updates (e.g., via quantization techniques) reduces the size of transmitted model updates, improving communication efficiency.
    - Model pruning techniques allow less critical model parameters to be excluded from updates, further minimizing bandwidth usage.
3. Hierarchical FL aggregation:
    - Instead of sending updates from all edge devices to a single central server, a multi-tier aggregation system can be implemented.
    - Local aggregators in edge clusters collect updates, perform initial model merging, and only send refined updates to the central server, reducing global communication overhead.

By integrating these techniques, federated learning can maintain its privacy-preserving advantages while minimizing update latency and network overhead, making it more suitable for large-scale IoT security applications.

5.2.3. Real-Time Performance

The system achieves low-latency anomaly detection, with response times remaining under 300 milliseconds across various cyber threats (Figure 11). This ensures that the proposed FL-driven security solution meets the real-time constraints of 5G-enabled IoT networks, making it feasible for large-scale deployment in smart city environments.

5.3. Implications for AI-Driven Security in IoT Networks

What does this research mean for real-world deployment? The findings indicate that AI-driven anomaly detection models can significantly enhance cybersecurity in IoT-enabled smart cities. However, several practical deployment considerations must be addressed:

- Privacy-preserving AI models: While federated learning ensures decentralized model training, regulatory concerns over data sharing and AI decision transparency remain areas requiring further investigation.
- Explainability in AI-driven security: Security professionals require interpretable AI models. Our use of SHAP-based feature analysis improves AI transparency, but additional efforts are needed to enhance trust in automated security decisions.
- Adaptive cybersecurity strategies: AI models need to continuously adapt to emerging threats. Integrating self-learning AI mechanisms and reinforcement learning could improve their real-time adaptation to zero-day attacks.

While the proposed system supports privacy-preserving training via federated learning, several operational and security risks must be addressed before real-world deployment. These include model poisoning attacks, where malicious edge nodes send corrupted updates to the aggregator, and gradient inversion attacks, which could potentially reconstruct private training data from shared model updates. To mitigate these risks, techniques such as secure aggregation, anomaly-aware update filtering, and reputation-based trust scoring can be employed to enhance the reliability of participating edge nodes. Ensuring node trustworthiness is critical for maintaining model integrity in decentralized environments. Moreover, compliance with regulations such as those of the GDPR and CCPA must be ensured throughout the data lifecycle—from collection and anonymization to model inference and mitigation actions. This includes providing transparency in AI decisions, preserving user consent boundaries, and ensuring that no personally identifiable information (PII) is exposed. Future work should explore differential privacy, secure multiparty computation, and decentralized trust management to further enhance the system's resilience, trustworthiness, and regulatory alignment in large-scale smart city deployment.

5.4. Limitations and Future Directions

While our proposed system demonstrates high accuracy and scalability, certain limitations should be noted:

- Computational overhead: Deep learning models require significant processing power, which may be challenging for resource-constrained IoT devices.
- Potential for adversarial attacks: AI-driven models are vulnerable to adversarial attacks, where attackers manipulate network data to evade detection.
- Real-world deployment testing: Although experiments were conducted using real-world datasets, future work should focus on live deployment in IoT networks to assess system robustness in real-time.

Future research directions include:

- Enhancing AI explainability and trustworthiness: Exploring interpretable AI models for cybersecurity.
- Developing lightweight AI models for edge devices: Optimizing deep learning architectures for low-power IoT environments.
- Integrating reinforcement learning for dynamic threat adaptation: Developing AI models that learn and evolve in response to new attack patterns.

This discussion highlights how our proposed AI-driven anomaly detection system improves upon state-of-the-art solutions, ensuring high detection accuracy, scalability, and

real-time feasibility. However, addressing AI explainability, privacy concerns, and adversarial robustness is crucial before real-world implementation in IoT-enabled smart cities.

6. Conclusions and Future Work

6.1. Conclusions

The rapid expansion of IoT-enabled smart cities has introduced new cybersecurity challenges, necessitating scalable, real-time, and adaptive security solutions. This study proposes an AI-driven anomaly detection system that integrates hybrid deep learning models (autoencoder + LSTM + CNN), real-world dataset validation, scalable federated learning techniques, and AI-based autonomous mitigation strategies to enhance cybersecurity in large-scale IoT environments.

Through extensive experimentation, our findings demonstrate that:

- The hybrid AI model significantly outperforms standalone deep learning approaches (CNN, LSTM, and autoencoders), achieving a precision of 97.5% and a recall of 96.2% (Table 3, Figure 7).
- The federated learning-based anomaly detection framework scales up efficiently, maintaining an accuracy above 96% even across 100 IoT edge nodes (Figure 10), proving its feasibility for large-scale smart city deployment.
- The AI-driven autonomous threat mitigation system provides a real-time attack response, achieving a 97% mitigation success rate with response times under 300 ms (Figure 11), making it suitable for real-world, high-speed security enforcement.
- The SHAP-based feature importance analysis (Figure 9) enhances AI transparency, addressing the black-box nature of deep learning models by providing explainability for security decisions.

These results validate the proposed system's effectiveness in detecting, analyzing, and mitigating IoT-based cyber threats in real time, demonstrating its potential as a scalable and efficient cybersecurity framework for 5G-enabled smart cities.

6.2. Future Work

While this study provides a robust AI-driven security solution for IoT networks, several avenues for future research could further enhance its scalability, adaptability, and resilience to adversarial threats.

1. Lightweight AI Models for Resource-Constrained IoT Devices
   Deep learning models, particularly LSTMs and CNNs, require significant computational resources, which may not be feasible for use with low-power IoT sensors and edge devices. Future research should focus on developing lightweight deep learning architectures, such as quantized neural networks and knowledge distillation, which are optimized for low-latency IoT security applications.
2. Robustness Against Adversarial Attacks
   AI-based security models are vulnerable to adversarial attacks, where attackers intentionally manipulate network traffic to bypass anomaly detection mechanisms. While our proposed system effectively detects known cyber threats, no empirical evaluation of adversarial robustness was conducted in this study. Future research should focus on developing and evaluating adversarial defense mechanisms in AI-driven cybersecurity. Several promising directions include:
   - Adversarial training by introducing adversarially generated attack samples during model training to enhance resilience against adversarial perturbations.
   - Defensive AI strategies leveraging defense distillation, anomaly-aware embeddings, and gradient masking techniques to mitigate adversarial evasion.

- Adversarial detection modules, deploying secondary anomaly detection layers that can flag adversarially manipulated network traffic in real time.
- Empirical evaluations of adversarial robustness, made by conducting systematic testing of AI models against adversarial attacks (e.g., the fast gradient sign method (FGSM) and projected gradient descent (PGD)) to quantify vulnerability levels and propose countermeasures.

Ensuring robustness against adversarial manipulation is essential for maintaining trust and reliability in AI-driven security frameworks, particularly in critical smart city infrastructures. While adversarial examples were used during training and evaluation data preparation, the current study did not conduct isolated adversarial robustness testing to quantify the system's performance under targeted attacks. We recognize this as an important limitation and propose integrating formal benchmarking against evasion techniques (e.g., FGSM, PGD, and AutoAttack) in future iterations of the framework.

3. Self-Learning AI for Dynamic Threat Adaptation
   Cyber threats evolve rapidly, and static AI models may fail to detect novel attack patterns. Future research should explore:
   - Reinforcement learning (RL) for adaptive threat mitigation, with AI-driven intrusion detection systems (IDS) that can learn from evolving attack behaviors.
   - Continual learning by implementing online learning techniques to dynamically update AI models without retraining from scratch.

   This approach will enable intrusion detection systems to autonomously adapt to emerging cyber threats in real time. In addition to dynamic detection, future efforts will explore learning-based mitigation, such as reinforcement learning agents that select optimal countermeasures based on observed threat patterns, policy effectiveness, and system state.
4. Federated Learning Enhancements for Privacy-Preserving Security
   While federated learning (FL) enhances the system's scalability and privacy, communication overhead and model divergence across heterogeneous IoT networks remain challenges. Future research should explore:
   - Efficient FL aggregation methods using techniques such as differential privacy, secure multi-party computation (MPC), and blockchain-based federated learning can enhance privacy, security, and communication efficiency.
   - Personalized FL systems for IoT networks need to develop client-specific AI models that can adapt to device heterogeneity while maintaining global security consistency.
5. Real-World Deployment, Applicability, and Benchmarking
   Although our system was evaluated using real-world datasets, deploying it in live IoT networks will provide deeper insights into:
   - Real-time performance and adaptive response efficiency.
   - The network overhead that is associated with federated learning updates.
   - The effectiveness of AI-driven autonomous mitigation strategies in production environments.

The real-world applicability of the proposed framework is particularly relevant to smart city domains, such as public safety monitoring, traffic control systems, and utility infrastructure security. In contrast to traditional centralized IDS systems, our approach offers the following comparative advantages:

- Decentralized training via federated learning significantly reduces data transmission costs and privacy risks.

- Edge AI deployment ensures low-latency, real-time anomaly detection.
- AI-driven adaptive threat mitigation allows for autonomous response, reducing the burden on human operators.

Collaborations with industry partners and smart city infrastructures could facilitate field testing, ensuring the practical viability of AI-driven IoT security solutions.

Future research should prioritize adversarial robustness, scalability enhancements, and real-world validation to further improve AI-driven anomaly detection systems for 5G-enabled smart cities.

6.3. Final Thoughts

The proposed AI-driven anomaly detection system represents a significant step forward in securing IoT-enabled smart cities by leveraging hybrid deep learning, federated learning, and autonomous threat mitigation. The findings indicate that AI-based cybersecurity frameworks can enhance detection accuracy, scalability, and real-time mitigation efficiency. However, continued research in lightweight AI, adversarial robustness, self-learning models, and real-world deployment is crucial for ensuring long-term effectiveness and adaptability to evolving cyber threats.

By addressing these future directions, AI-driven cybersecurity frameworks can become more resilient, efficient, and trustworthy, ultimately strengthening the security posture of IoT networks in smart city environments.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study data were derived from publicly available resources, including the CICIDS2017 dataset, a network intrusion dataset containing attack scenarios such as DDoS, brute-force, and botnet attacks. Available at: https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 17 April 2025). The TON_IoT dataset was designed for IoT and Industrial IoT security, containing network telemetry and cyberattack traces. Available at: https://research.unsw.edu.au/projects/toniot-datasets (accessed on 17 April 2025). The UNSW-NB15 dataset offers a modern network intrusion dataset featuring various malicious traffic scenarios. Available at: https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/ (accessed on 17 April 2025). Additional synthetic data were generated using adversarial attack simulation techniques to evaluate the model's robustness. These synthetic datasets are available upon request from the corresponding author. Further details regarding synthetic data generation can be found in Section 3.2 (Dataset Selection and Preprocessing).

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Priyadarshini, I. Anomaly Detection of IoT Cyberattacks in Smart Cities Using Federated Learning and Split Learning. *Big Data Cogn. Comput.* **2024**, *8*, 21. [CrossRef]
2. Institute for Defense & Business. What Are the Cybersecurity Risks for Smart Cities? Available online: https://www.idb.org/what-are-the-cybersecurity-risks-for-smart-cities/ (accessed on 17 March 2025).
3. Security and Compliance in 5G and AI-Powered Edge Networks | Deloitte Global. Available online: https://www.deloitte.com/global/en/services/consulting-risk/perspectives/security-compliance-in-5g-ai-powered-edge-networks.html (accessed on 17 March 2025).
4. How 5G Technology Affects Cybersecurity: Looking to the Future | UpGuard. Available online: https://www.upguard.com/blog/how-5g-technology-affects-cybersecurity (accessed on 17 March 2025).
5. Mirza, N.; Yunis, M.; Khalil, A.; Mirza, N. Towards a Conceptual Framework for AI-Driven Anomaly Detection in Smart City IoT Networks for Enhanced Cybersecurity. *J. Innov. Knowl.* **2024**, *9*, 100601.
6. Cybersecurity Challenges in Smart Cities: An Overview and Future Prospects | Mesopotamian Journal of CyberSecurity. Available online: https://mesopotamian.press/journals/index.php/CyberSecurity/article/view/14 (accessed on 17 March 2025).
7. USA; Aluwala, A. AI-Driven Anomaly Detection in Network Monitoring Techniques and Tools. *J. Artif. Intell. Cloud Comput.* **2024**, 1–6. [CrossRef]

8. Babu, C.V.S.; Simon, P.A. Adaptive AI for Dynamic Cybersecurity Systems: Enhancing Protection in a Rapidly Evolving Digital Landscap. Available online: https://www.igi-global.com/gateway/chapter/337688 (accessed on 17 March 2025).
9. Kuraku, D.S. Adaptive Security Framework For Iot: Utilizing AI And ML To Counteract Evolving Cyber Threats. *Educ. Adm. Theory Pract.* **2023**, *29*, 1573–1580. [CrossRef]
10. What Is the Role of AI in Threat Detection? Available online: https://origin-www.paloaltonetworks.com/cyberpedia/ai-in-threat-detection (accessed on 17 March 2025).
11. Kuguoglu, B.K.; van der Voort, H.; Janssen, M. The Giant Leap for Smart Cities: Scaling Up Smart City Artificial Intelligence of Things (AIoT) Initiatives. *Sustainability* **2021**, *13*, 12295. [CrossRef]
12. Towards Large-Scale IoT Deployments in Smart Cities: Requirements and Challenges | SpringerLink. Available online: https://link.springer.com/chapter/10.1007/978-3-031-50514-0_6 (accessed on 17 March 2025).
13. Zyrianoff, I.; Borelli, F.; Biondi, G.; Heideker, A.; Kamienski, C. Scalability of Real-Time IoT-Based Applications for Smart Cities. In Proceedings of the 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, Brazil, 25–28 June 2018; pp. 00688–00693.
14. Abusitta, A.; Silva de Carvalho, G.H.; Abdel Wahab, O.; Halabi, T.; Fung, B.C.M.; Al Mamoori, S. Deep Learning-Enabled Anomaly Detection for IoT Systems. *Internet Things* **2022**, *21*, 100656. [CrossRef]
15. AI and IoT in Smart Cities Security. Available online: https://www.truehomeprotection.com/leveraging-ai-and-iot-for-next-generation-security-systems-in-smart-cities/ (accessed on 17 March 2025).
16. Sarbhukan (Bodade), V.V.; More, J.S.; Jadhav, Y. Smart City Infrastructure Monitoring Using AI and IoT Technologies. *Int. J. Intell. Syst. Appl. Eng.* **2024**, *12*, 1687–1695.
17. Sáez-de-Cámara, X.; Flores, J.L.; Arellano, C.; Urbieta, A.; Zurutuza, U. Clustered Federated Learning Architecture for Network Anomaly Detection in Large Scale Heterogeneous IoT Networks. *Comput. Secur.* **2023**, *131*, 103299. [CrossRef]
18. Ortiz, O.O.; Pastor Franco, J.Á.; Alcover Garau, P.M.; Herrero Martín, R. Innovative Mobile Robot Method: Improving the Learning of Programming Languages in Engineering Degrees. *IEEE Trans. Educ.* **2017**, *60*, 143–148. [CrossRef]
19. Elshaiekh, N.E.M.; Hassan, Y.A.A.; Abdallah, A.A.A. The Impacts of Remote Working on Workers Performance. In Proceedings of the 2018 International Arab Conference on Information Technology (ACIT), Sultan Qaboos University, Muscat, Oman, 28–30 November 2018; pp. 1–5.
20. Goswami, M.J. AI-Based Anomaly Detection for Real-Time Cybersecurity. *Int. J. Res. Rev. Tech.* **2024**, *3*, 45–53.
21. Ngo, M.V.; Luo, T.; Quek, T.Q.S. Adaptive Anomaly Detection for Internet of Things in Hierarchical Edge Computing: A Contextual-Bandit Approach. *ACM Trans. Internet Things* **2021**, *3*, 1–23. [CrossRef]
22. Liu, Y.; Zhou, Y.; Yang, K.; Wang, X. Unsupervised Deep Learning for IoT Time Series. *IEEE Internet Things J.* **2023**, *10*, 14285–14306. [CrossRef]
23. Nguyen, T.-A.; Le, L.T.; Nguyen, T.D.; Bao, W.; Seneviratne, S.; Hong, C.S.; Tran, N.H. Federated PCA on Grassmann Manifold for IoT Anomaly Detection. *IEEEACM Trans. Netw.* **2024**, *32*, 4456–4471. [CrossRef]
24. Zolanvari, M.; Ghubaish, A.; Jain, R. ADDAI: Anomaly Detection Using Distributed AI. In Proceedings of the 2021 IEEE International Conference on Networking, Sensing and Control (ICNSC), Xiamen, China, 3–5 December 2021; Volume 1, pp. 1–6.
25. Subasi, O.; Cree, J.; Manzano, J.; Peterson, E. A Critical Assessment of Interpretable and Explainable Machine Learning for Intrusion Detection. *arXiv* **2024**, arXiv:2407.04009.
26. Garcia, E. Smart City and Iot Data Collection Leveraging Generative Ai. Available online: https://philpapers.org/rec/GARSCA-9 (accessed on 15 May 2025).
27. Manthena, H.; Shajarian, S.; Kimmell, J.; Abdelsalam, M.; Khorsandroo, S.; Gupta, M. Explainable Artificial Intelligence (XAI) for Malware Analysis: A Survey of Techniques, Applications, and Open Challenges. *IEEE Access* **2025**, *13*, 61611–61640. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.