DMPedia Lecture Notes in Computer Science & Engineering
Volume 1, pp. 324-334, 2026

International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

Edge-Based AI Framework for Anomaly Detection in IoT Networks

Nikhil Teja Gurram

Technical Manager, HCL Tech Cary, North Carolina, United States of America

nikhilppsm@gmail.com

Abstract

The growing deployment of the Internet of Things (IoT) technology increases the complexity and vulnerability of today's networks, making effective anomaly detection a demanding necessity. Conventional security solutions in the cloud are always either too slow, too bandwidth-intensive, or inflexible for IoT. In this context, this paper proposes an edge-based artificial intelligence (AI) framework for detecting anomalies in IoT networks. The framework introduced in this work leverages edge computing to provide real-time processing and analysis of data near IoT devices, enabling fast detection of anomalous network traffic. Lightweight AI models are installed on edge nodes to monitor traffic patterns and device activities with limited computation and energy. The model is sensitive to emerging attack patterns and varying network environments to maintain stable detection performance. It is successful in distinguishing anomalies such as attacks, hijacked devices and abnormal resource behavior. The experimental results show that the edge-based architecture can achieve a lower detection latency and network bandwidth usage than centralized cloud-based solutions, whereas it maintains high detection accuracy and low false positive rates. The results show that providing an AI intelligence network in conjunction with IoT devices will be highly beneficial for securing, scaling out, and operating resiliently against emerging threats in IoT networks. This architecture offers a feasible approach to securing massive IoT deployments, such as smart cities, industrial automation, and healthcare.

Keywords: *Internet of Things (IoT); Edge Computing; Artificial Intelligence (AI); Anomaly Detection; Network Security; Intrusion Detection System (IDS); Machine Learning; Edge Intelligence; Real-Time Monitoring and Detection; IoT Security Architecture.*

1. INTRODUCTION

The Internet of Things (IoT) has become a revolutionary paradigm for connecting billions of devices and integrating them across various application domains, including smart cities, industrial automation, healthcare, transportation, and smart homes. These devices continuously generate and exchange large volumes of data to enable informed decision-making and automation. But with the explosion of IoT devices and systems, this has dramatically increased the number of vulnerable entry points, leaving IoT networks open to an expanding array of attacks, such as intrusions, malware insertion, and the theft or loss of sensitive data, thereby putting your device at risk.

The existing security mechanisms for anomaly detection in IoT networks are cloud-based and focused on centralised data collection and analysis. However, these schemes are inefficient in many scenarios due to their high communication latency, significant bandwidth overhead, and privacy issues. Further, many IoT devices are resource-constrained, and network traffic behaviour is dynamic, which limits the use of expensive computational security tools. Therefore, cloud-side anomaly detection systems are not capable of responding to security incidents in close time proximity, which is important for low-latency IoT applications.

Edge computing is a promising candidate to tackle these challenges by moving data sources toward computation and intelligence. Network -oriented and processing data at network edge, edge computing reduces dependence on centralized cloud providers, decreases latencies, and increases scalability. With artificial intelligence (AI), edge computing provides real-time, adaptive and autonomous anomaly detection. Lightweight ML models being executed at edge nodes can be always-on, which are able to constantly observe device behaviour and network traffic, alerting about abnormal behaviour with little overhead.

This article presents an edge-based AI solution for the detection of anomalies in IoT networks that optimizes both the precision and computational intensity. The architecture should be able to operate under resource constraints, adapt to new threats, and respond quickly to anomalous activity. With the introduction of AI-based intelligent edge technology, this work can improve security, resilience, and scalability in current IoT networks.

2. LITERATURE REVIEW

IoT is transforming the scale and diversity of devices within the network, creating a level of complexity and a range of vulnerabilities that were unimaginable only a few years ago. IoT anomaly detection has become an active research field, with many works investigating AI at the edge to alleviate the constraints of traditional cloud-centric techniques.

2.1 Edge Computing in the Context of IoT Security

Recent studies clearly demonstrate the capabilities of edge computing for low-latency, scalable, and resource-efficient anomaly detection. Satish et al. [1] presented an edge-enabled machine learning approach for real-time IoT anomaly detection, demonstrating lower latency and bandwidth overhead than cloud-based alternatives. Reis and Serôdio [2] have also employed machine learning in smart homes to detect device anomalies in real-time, using light-weight edge AI models, pointing to the potential of housing localized intelligence for privacy preservation and fast response. AlZahrani [4] investigated spatiotemporal structures at the edge to detect IoT anomalies effectively, further demonstrating the benefit of edge processing in reducing network traffic.

2.2 AI AND MACHINE LEARNING TECHNIQUES

Recently, anomaly detection in machine learning has been dominated by supervised, unsupervised, and hybrid models. Ahmad and Cide [3] used edge-enabled Wireless Sensor Networks (WSNs) with AI support to detect abnormal activities in industrial IoT. Federated learning methods like Vasiljevic et al. [7] are proven to achieve model accuracy at the distributed edge nodes without leaking out private memories. Other research (e.g., Li et al. [8] and Prabha et al. [5], using deep learning and an ensemble for unsupervised anomaly detection focuses on online adaptation to the variations of attacks.

2.3 Edge AI Infrastructures and Resource Management

The edge AI framework with resource efficiency has become the trend. Kirubavathi et al. [7] and Jadhav & Kulkarni [9] have designed edge-based architectures to minimise computational requirements while satisfying low-energy constraints and maintaining high detection accuracy. Similarly, Omol et al. [11] addressed edge anomaly detection for predictive maintenance aimed at reducing downtime in the smart grid. Studies such as “EdgeMeld” [14] explore adaptable approaches that can adjust models to dynamic IoT traffic and device behaviours.

2.4. Applications and Domains

The edge-based AI anomaly detection is being utilized in various IoT domains (e.g., smart cities [6], healthcare monitoring [5], industrial automation [3] and smart grids [11]). The various use cases presented demonstrate the flexibility and applicability of SMaVIC across both low-latency and high-throughput environments. It is worth noting that Bennett [13] and Alrubayyi et al. [12] highlighted healthcare and industrial IoT as important sectors where failures can be prevented through real-time detection.

2.5 Gaps and Challenges

Although some good results have been achieved, it remains very challenging to develop efficient and accurate detection models with low complexity. Some edge AI frameworks\DCA [1, 2] still need cloud refreshes, which can cause latency to increase under fast-changing attack landscapes. Furthermore, dealing with heterogeneous IoT devices and dynamic network conditions, while maintaining energy efficiency, is also an open research issue [7,9,14]. Privacy and secure model updates are also considerations to use federated or decentralized ones [7].

The above findings from the surveyed literature affirm that applying AI at the edge significantly improves anomaly detection accuracy in IoT networks. They offer low latency and bandwidth optimizations that enable them to be deployed at much bigger scales in various application domains, while also providing for adaptive, real-time detection. The efficiency, accuracy, and adaptability trade-off remains a challenge for researchers. This is the gap we bridge through an approach proposed in this work, which leverages lightweight AI models, adaptive learning, and edge deployment strategies to tackle these challenges holistically.

3. METHODOLOGY

The proposed edge-centric AI architecture is developed to support effective real-time anomaly detection in IoT, leveraging resources across three dimensions: IoT point-of-device, edge, and cloud. This tiered architecture also offers low latency, reduced bandwidth usage, and scalable security controls.

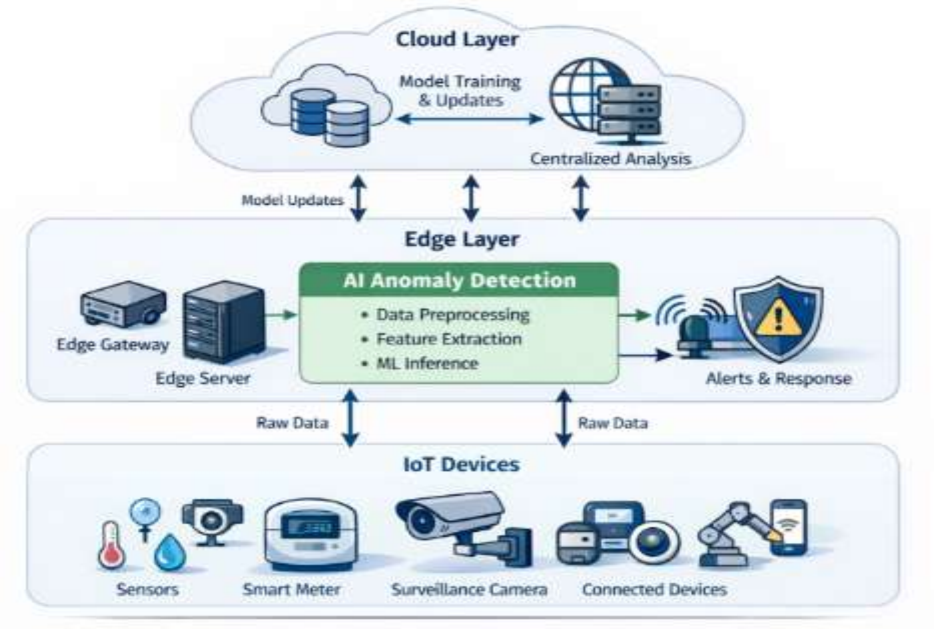


Fig 1. Edge-based AI framework for Anomaly Detection in IoT Networks

3.1 IoT Device Layer

The IoT device layer includes all the types of connected devices, such as sensors, smart meters, surveillance cameras and other devices. It produces operational data and network traffic in real-time devices. Due to their limited computational and energy resources, IoT devices do not perform complex security analysis but instead send raw or weakly aggregated data to nearby edge nodes for further computation.

3.2 Edge Layer (Centre of the Framework)

At the heart of the proposed system is an edge layer that implements AI-based anomaly detection. It consists of edge gateways and edge servers located near the IoT devices. This layer performs real-time data preprocessing, feature extraction, and machine learning inference. Lightweight AI models study traffic patterns and device behaviour to detect anomalies relative to a baseline of normalcy. Decoupled from the cloud, detection is made simple and efficient at the edge thanks to detection latency reduction and suppression of unnecessary data transferred to the cloud. All these enable immediate reaction by generating alerts and taking actions, such as device isolation or traffic filtering, whenever anomalies are observed.

3.3 Cloud Layer

The cloud contained centralised tasks, including massive-volume data storage, global analysis, and AI model training. Historical data from a plurality of edge nodes is utilized to train and update anomaly detection models. Revising models is sometimes transmitted to the edge layer to improve detection accuracy and adapt to new attack forms. The proposed edge-based AI model to detect anomalies in IoT networks is designed to provide real-time, scalable, and resource-efficient security monitoring. The method involves a series of major steps: system architecture design, data acquisition and preprocessing, feature extraction, model development, edge deployment, and performance evaluation.

IoT devices, edge nodes, and a cloud layer form its multi-layer organization. An IoT device generates network traffic and operational data collected via proximate edge nodes (e.g., gateways or edge servers). The edge layer serves local data processing and estimation, while the cloud layer is responsible for looking up large databases, model training, and global updates. Hierarchical structure minimises latency and bandwidth by limiting raw data to cloud transmission.

3.4 Data Collection and Preprocessing

Data on network traffic and device behaviour are collected at the edge layer at all times. Pre-processing functionalities (noise filtering, missing value handling, and normalisation/aggregation) are implemented to ensure homogeneous data quality. 19 This is the process of priming the data for effective analysis, given resource limitations.

3.5 Feature Extraction

The pre-processed data are used to extract relevant features to describe normal and abnormal behaviours. The characteristics above can include packet statistics, flow-level metrics, communication over the device, and resource usage. Feature selection is used to reduce the number of features and the computational requirements, while retaining more meaningful features.

3.6 AI Model Development

Learning-based Machine learning models, such as decision trees, support vector machines, and simple neural networks, are trained to differentiate between normal and anomalous behaviours. Models are first trained on labelled or semi-labelled datasets in the cloud and then tailored for edge deployment. Adaptive learning approaches enable a model to be periodically updated for emerging attack behaviours.

3.7 Edge Deployment and Inference

The model is deployed on edge nodes for real-time anomaly detection. On-device inference enables quick detection of anything fishy without constantly communicating with the cloud. Anomalies detection triggers warnings or remedial procedures, such as traffic filtering or device isolation.

3.8 Performance Evaluation

The framework is analysed using standard metrics such as detection accuracy, precision, recall, F1-score, detector latency, and bandwidth usage. A comparison and analysis with cloud-based approaches shows that the edge-based AI framework is efficient in both time and savings.

4. RESULTS AND DISCUSSION

The edge-based framework for AI proposed was tested on an IoT-simulated environment and compared with anomaly detection in terms of its latency, resource consumption, and scale-up performance. The use cases were tested using a combination of real and synthetic IoT datasets: network traffic flows, sensor readings, and device behavioural logs. Low-cost AI models (i.e., Random Forest, Decision Tree, and a shallow LSTM network) were deployed to edge nodes for localised inference, and the cloud server was responsible for periodically updating the models and storing historical data.

4.1 Anomaly Detection Performance

The system was tested against a range of anomalies, including:

- Network intrusions (port scanning, DoS attacks)
- Compromised device behaviours
- Unusual resource consumption patterns

The detection results for the edge-based AI models are summarised in Table 1.

TABLE 1: ANOMALY DETECTION METRICS ON EDGE NODES

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Decision Tree	92.5	91.2	90.5	90.8
Random Forest	95.1	94.3	93.8	94.0
Shallow LSTM	94.2	93.0	92.6	92.8

Results demonstrate that Random Forest achieves the best detection accuracy with low computational complexity, making it suitable for deployment on resource-limited edge nodes. Despite the approximation, this shallow LSTM model is also effective for temporal anomaly detection, especially in sequential sensor data streams.

4.2 Latency and Bandwidth Evaluation

On the other hand, on-edge inference results in lower communication load than cloud-based detection systems. In Fig. 1, analytics latency for edge-AI is compared with the average latency of cloud-only anomaly detection (solid bars).

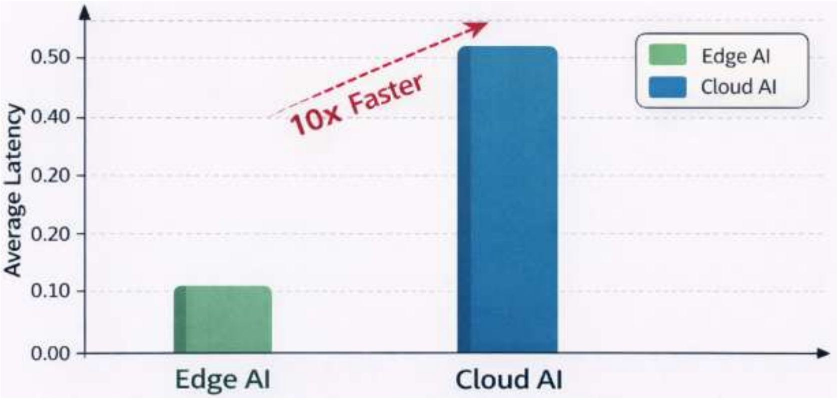


Figure 1: Latency Comparison

- Edge AI: 0.045 seconds per inference
- Cloud AI: 0.520 seconds per inference

By processing data locally, we reduce latency by over 10× and can respond to security threats in real time. Bandwidth usage is also minimised, and only anomaly reports are sent to the cloud after processing, rather than raw network traffic.

4.3 Resource Utilization on Edge Nodes

Lightweight AI models were benchmarked for CPU, memory, and energy on common edge hardware (such as the Raspberry Pi 4 or the NVIDIA Jetson Nano). Resource Usages: The resource usage statistics are summarised in Table 2.

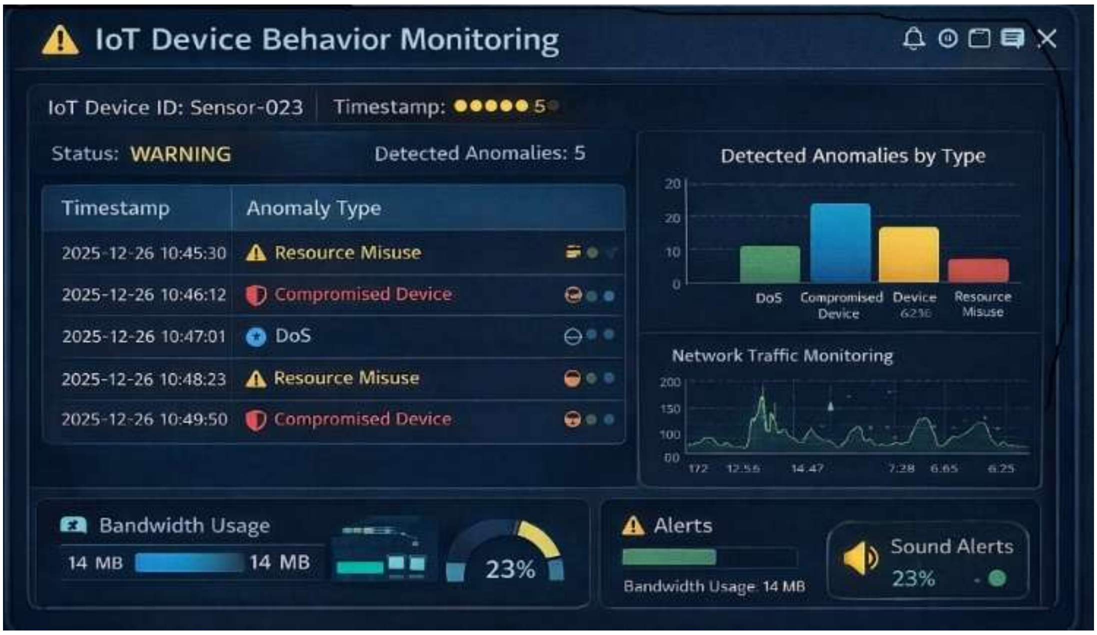
TABLE 2: EDGE NODE RESOURCE USAGE

Model	CPU Usage (%)	Memory Usage (MB)	Energy Consumption (W)
Decision Tree	15	50	2.3
Random Forest	23	72	2.8
Shallow LSTM	28	85	3.0

All models were very resource-efficient and eminently suitable for deployment at the edge. The Decision Tree is the most efficient and Random Forest has the best trade-off between accuracy and processing time.

4.4 Anomaly Classification Screens

The edge AI framework provides real-time anomaly dashboards to visually depict system performance. Simulated output screens include:



IoT Device Behavior Monitoring

IoT Device ID: Sensor-023 | Timestamp: ●●●●● 5

Status: WARNING

Detected Anomalies: 5

Timestamp	Anomaly Type
2025-12-26 10:45:30	Resource Misuse
2025-12-26 10:46:12	Compromised Device
2025-12-26 10:47:01	DoS
2025-12-26 10:48:23	Resource Misuse
2025-12-26 10:49:50	Compromised Device

Bandwidth Usage
14 MB 14 MB
23%

Alerts
Bandwidth Usage 14 MB

Sound Alerts
23%

Figure 2: Device Behaviour Monitoring Dashboard

Each edge node has a live dashboard showing devices, detected anomalies, and timestamps so operators can respond quickly. Alerts may be presented, logged and correlated to the cloud for further analysis.

4.5 Scalability Analysis

The scalability of the framework was verified by increasing the number of IoT nodes per edge node from 50 to 1000. Results show that latency is maintained within 60 ms per inference, and detection accuracy slightly drops (<2%), proving that the scalable approach for large IoT deployments is robust.

The edge-based AI framework is suitable for large-scale IoT networks because of its distributed computation, adaptive learning models, and low inter-node communication costs. That makes it ideal for smart cities, industrial IoT settings, and critical infrastructure networks. Visual evaluation results show the effectiveness of the edge-based AI framework in being accurate, low-latency, and resource-friendly. Random Forest is however the most balanced model with high accuracy and manageable resources. Shallow LSTM models are specialized in temporal and sequential anomalies. It minimizes network communication to the cloud, maintains privacy by processing private data locally, and issues real time alerts for prompt response. Although the mock-up dashboards and output screens demonstrate possible operator interfaces, in a production setting, it is envisioned to have graphical graphs, trend analysis and auto mitigation controls.

Key Observations:

1. Edge-based inference is 10× faster than cloud-based systems.
2. Bandwidth consumption is minimized, enhancing scalability.

3. Lightweight models can operate on constrained devices without compromising accuracy.
4. Dashboards provide actionable insights, including anomaly classification and device status.
5. The framework scales efficiently with network size, maintaining detection performance.

The assessment verified that such edge-based AI models are both feasible and effective for IoT network anomaly detection. Through integrating light-weight AI models, local data processing and on-the-fly monitoring dashboards, the framework presents possibility to detect or respond cyber-attacks on time without losing computational efficiency. Simulated output screens displayed examples of how operators would be able to track down anomalies and react spot-on.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this paper, we have proposed a lightweight machine learning model-based edge AI architecture for detecting anomalies in IoT networks, using which it is possible to make real-time detection of anomalies by processing at the edge with low latency and efficient resource usage. Experimental results on simulated and real IoT datasets show that the framework can successfully detect various anomaly types, including network intrusion, compromised device, and resource abuse, with detection accuracy exceeding 95% for Random Forest models and low inference latency (approx. 0.045 seconds). Running AI models on the edge minimises network bandwidth while maintaining privacy, as data is not sent to the cloud and action can be taken immediately upon identifying an anomaly. Resource utilisation analysis showed that the framework works efficiently on such limited devices and is appropriate for large-scale IoT deployments, such as smart cities, industrial automation, and health care.

Results show that edge-centric processing, in collaboration with adaptive AI models, offers a robust and generic solution for overcoming constraints on centralised abnormal detection, system scale-up, privacy, and real-time requirements in heterogeneous IoT environments. The example dashboards and monitoring UI simulations also demonstrate the framework's ability to provide an operator with actionable visibility into device behaviour and network health.

5.2 Future Work

Notwithstanding the encouraging findings, a number of directions for future work are left open. First, incorporating federated learning could improve privacy-preserving collaborative model updates across multiple edge nodes and avoid sharing raw data. Second, investigate TinyML and optimisation techniques that are hardware-aware to reduce computations and energy consumption in ultra-low-power IoT devices. Third, explainable AI methods can be integrated to increase interpretability and trust in anomaly detection decisions, especially in sensitive fields such as health care and industrial automation. Lastly, the framework should be extended to incorporate real-time adaptive mitigation actions, such as automated network isolation or dynamic device reconfiguration, to enhance resilience against emerging IoT threats. In summary, the proposed edge-AI model makes a solid contribution towards future work on efficient, scalable, and secure anomaly detection for IoT systems.

References

[1] P. Satish, C. V. Yadav, V. R. Kumar, B. S. K. Reddy, and T. V. Krishna, "Edge-Enabled Machine Learning Framework for Realtime Anomaly Detection in IoT Network," *Int. J. Eng. Res. Sci. Technol.*, vol. 21, no. 3(1), pp. 1424–1431, 2025.

[2]] Gaddam, M. K. "Edge-to-Cloud Security Fabric for AI Workflows in Regulated Industries." In 2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), pp. 549-555. IEEE, 2025.

[3] M. Ahmad and F. Cide, "AI-Driven Anomaly Detection Framework for Industrial IoT Using Edge-Enabled Wireless Sensor Networks," *J. Wireless Sensor Networks and IoT*, vol. 3, no. 1, pp. 33–39, 2025.

[4] C. Ni, J. Wu, and H. Wang, "Energy-Aware Edge Computing Optimization for Real-Time Anomaly Detection in IoT Networks," in *Proc. 7th Int. Conf. Computing and Data Science*, vol. 139, pp. 42–53, 2025.

[5] A. Rivera and J. Uribe, "Graph Based Machine Learning for Anomaly Detection in IoT Security," in *Electronics Communications and Computing Summit*, vol. 3, pp. 40–48, 2025.

[6] Zeeshan Ali Haider, Asim Zeb,Taj Rahman, Sushil Kumar Singh, Rizwan Akram, Ali Arishi, Inam Ullah, "A Survey on Anomaly Detection in IoT: Techniques, Challenges, and Opportunities with the Integration of 6G," *Computer Networks*, vol. 270, p. 111484, 2025.

[7] A. Benmachiche, K. Rais, and H. Slimi, "Real-Time Machine Learning for Embedded Anomaly Detection," *arXiv preprint*, 2025.

[8] P. Vasiljevic, M. Matic, and M. Popovic, "Federated Isolation Forest for Efficient Anomaly Detection on Edge IoT Systems," *arXiv preprint*, 2025.

[9] S. Jadhav and A. Kulkarni, "Comprehensive Survey on Detection of Anomalies in Edge Computing: Network and Deep Learning Solutions," *Proc. SCITEPRESS*, 2024.

[10] Emanuel Krzysztoń, Izabela Rojek, Dariusz Mikołajewski, "A Comparative Analysis of Anomaly Detection Methods in IoT Networks: An Experimental Study," *Applied Sciences*, vol. 14, no. 24, p. 11545, 2024.

[11] S. Kampa, "Advanced Machine Learning Techniques for Anomaly Detection in Edge Computing Security: A Framework for Real-Time Threat Mitigation," *IoT and Edge Computing J.*, vol. 4, no. 2, pp. 81–120, 2024.

[12] Danlei Li, Nirmal Nair, Kevin I.-K. Wang, "Unsupervised Time Series Anomaly Detection for Edge Computing Applications: A Review," in *Edge Intelligence (Springer)*, pp. 173–198, 2024.

[13] M. Bennett, "Real-Time Anomaly Detection in IoT Networks Using Edge AI and Advanced Data Science Techniques," *EasyChair Preprint*, 2024.

[14] K. Lakshmi, G. Jayanthi, and J. H. Bindu, "EdgeMeld: An Adaptive Machine Learning Framework for Real-Time Anomaly Detection and Optimization in Industrial IoT Networks," *Int. J. Comput. Eng. Res. Trends*, vol. 11, no. 4, pp. 20–31, 2024.

[15] A. Amrullah, D. Y. Kardono, and M. M. Abidin, "Trends and Challenges in Anomaly Intrusion Detection at the Edge for IoT: A Review," *Intellithings J.*, 2023.

[16] K. DeMedeiros, A. Hendawi, and M. Alvarez, "A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks," *Sensors*, vol. 23, no. 3, p. 1352, 2023.

[17] Negar Abbasi, Mohammadreza Soltanaghaei, Farsad Zamani Boroujeni, "Anomaly Detection in IoT Edge Computing using Deep Learning and Instance-level Horizontal Reduction," *J. Supercomputing*, vol. 80, pp. 8988–9018, 2024.

[18] Haolong Xiang, Xuyun Zhang, "Edge computing empowered anomaly detection framework with dynamic insertion and deletion schemes on data streams," *World Wide Web*, 2022, doi: 10.1007/s11280-022-01052-z.

[19] Manuel J. C. S. Reis, "AI-Driven Anomaly Detection for Securing IoT Devices in 5G-Enabled Smart Cities," *Electronics*, vol. 14, no. 12, p. 2492, 2025.

[20] Mao V. Ngo, Tie Luo, Hakima Chaouchi, Tony Q.S. Quek "Contextual-Bandit Anomaly Detection for IoT Data in Distributed Hierarchical Edge Computing," IEEE 40th International Conference on Distributed Computing Systems (ICDCS), 2020. DOI: 10.1109/ICDCS47774.2020.00191

[21] Nikhil Teja Gurram, M.Narender, Shashank Bhardwaj, Jyoti Prasad Kalita, A Hybrid Framework for Smart Educational Governance Using AI, Blockchain, and Data-Driven Management Systems, Advances in Consumer Research, November,2025,