Issue 17, 2025

https://doi.org/10.23939/sisn2025.17.160

UDC 004.75

REAL-TIME ANOMALY DETECTION IN DISTRIBUTED IOT SYSTEMS: A COMPREHENSIVE REVIEW AND COMPARATIVE ANALYSIS

Pavlo Pustelnyk[1], Yevheniya Levus[2]

Lviv Polytechnic National University,
Department of Software Engineering, Lviv, Ukraine
[1] E-mail: pavlo.y.pustelnyk@lpnu.ua, ORCID: 0009-0005-5745-9941
[2] E-mail: yevheniia.v.levus@lpnu.ua, ORCID: 0000-0001-5109-7533

The rapid expansion of the Internet of Things (IoT) has resulted in a substantial increase of diverse data from distributed devices. This extensive data stream makes it increasingly important to implement robust and efficient real-time anomaly detection techniques that can promptly alert about issues before they could escalate into critical system failures. Anomaly detection in data is essential in today's interconnected landscape, as it facilitates the early identification of deviations from established baseline behavior that may indicate system malfunctions, security vulnerabilities, or operational inefficiencies. By promptly identifying these deviations, organizations can reduce downtime, optimize performance, and safeguard critical assets.

This article provides a comprehensive review and comparative analysis of modern methods for detecting anomalies in distributed IoT systems. It examines a wide range of techniques, including traditional statistical approaches, distance-based methods, machine learning models, deep learning architectures, and explainable AI frameworks. Each category is evaluated with respect to detection accuracy, computational efficiency, and interpretability. Real-world examples – ranging from predictive maintenance in industrial IoT and energy management in smart grids to fraud detection in financial networks – demonstrate the broad practical applications of these techniques.

The review further identifies current challenges and promising future research directions, including active learning-based approaches, which offer potential solutions to improve adaptability and reduce the reliance on large labeled datasets. The insights from this review provide a strong foundation for future research aimed at developing hybrid anomaly detection models that integrate advanced techniques to further enhance system adaptability and security in distributed IoT environments.

Key words: IoT, anomaly detection, real-time processing, machine learning, deep learning, explainable AI, distributed systems.

Problem Statement

The rapid expansion of the Internet of Things (IoT) has fundamentally transformed data generation and management across different industries (Idhalama, 2024). In today's interconnected world, devices such as sensors, meters, smart appliances, and industrial controllers continuously produce massive volumes of mixed data (Giannoni, 2018). This dynamic environment creates both opportunities and challenges. One of the most critical challenges is the detection of anomalies – data points or patterns that deviate from what is considered baseline behavior (Zakariah, 2023).

Anomalies can signal a variety of issues, from early signs of system malfunctions and equipment failures to potential cyber-attacks that threaten operational integrity (Chang, 2025). Although the introduction of advanced data processing and communication technologies has enabled more efficient management of large-scale systems, traditional security mechanisms that rely on static, rule-based approaches are increasingly insufficient (Ukil,

2016). The need for real-time, adaptive anomaly detection is evident as organizations aim to maintain system reliability, reduce downtime, and safeguard sensitive information (Alrashdi, 2019). For example, a sudden and unexplained spike in water consumption from smart metering data may indicate a leak, prompting immediate action to mitigate resource loss and potential damage.

This article provides a comprehensive review and comparative analysis of modern methods for real-time anomaly detection in distributed IoT systems. It examines a broad spectrum of techniques – from traditional statistical models and distance-based approaches to modern machine learning, deep learning, explainable AI, and active learning strategies. By analyzing the strengths and limitations of each approach, we aim to offer valuable insights into which methods are most promising for future research and practical implementation.

In order to conduct this review, a systematic approach to literature search and selection was adopted. Primary sources included major academic databases such as Scopus, Web of Science, IEEE Xplore, and Google Scholar. The search focused on studies addressing anomaly detection in IoT, emphasizing various modern methodologies and prioritizing publications from the last five years. Selected articles were subsequently subjected to a detailed analysis, concentrating on their methodological approaches, experimental outcomes, and overall contributions to the field.

Analysis of Recent Studies and Publications

The central challenge in real-time anomaly detection in distributed IoT systems is represented in processing and analyzing large volumes of heterogeneous data with minimal delay (Martins, 2022). Data are collected from devices utilizing different communication protocols, and their dynamic and diverse nature complicates the establishment of "baseline" behavior. Furthermore, many IoT devices have limited computational resources, making the deployment of resource-intensive algorithms problematic (Hichem, 2016). Additionally, for security-critical applications, the interpretability of detection outcomes is crucial (Cauteruccio, 2021). These factors necessitate methods that are not only accurate and efficient but also capable of providing transparent explanations for their decisions.

A systematic data anomaly detection literature overview (Jot, 2023) also provides an extensive insights of anomaly detection techniques in IoT. It overviews a broad range of methods and discusses the challenges related to real-time processing, scalability, and the adaptation of algorithms across different data domains. It concludes that although promising methods exist, further research is required to address persistent gaps in scalability and real-time processing efficiency.

Building on this foundation, study (Mutambik, 2024) focuses on deep learning approaches specifically applied to IoT devices. The paper examines CNN-based and RNN-based neural network architectures, along with autoencoder-based methods, to extract complex features from high-dimensional IoT data. It demonstrates that deep learning can achieve impressive detection accuracy, although these models demand considerable computational resources and often lack transparency.

In contrast, framework (Gad, 2025) uniquely combines threshold optimization with causal analysis using an explainable Random Forest model. By utilizing the linear non-Gaussian acyclic model, the framework uncovers causal relationships in network traffic data, thereby refining the detection thresholds and enhancing interpretability. The study reports near-perfect classification metrics on the CICIoT2023 dataset, illustrating the potential of integrating causal inference with machine learning for robust IoT security.

Study (Kaya, 2025) proposes a graph-based approach that transforms dynamic network data captured by Wireshark into visual graph representations. This method not only detects anomalies in real time but also provides an intuitive visualization of network behavior, enabling system administrators to quickly interpret and respond to deviations. The work demonstrates that graphical models can significantly enhance the understanding of complex network interactions.

A learning-driven framework for anomaly detection in IoT-based monitoring systems is presented in the article (Anusha, 2024). This research evaluates several machine learning models – including decision trees and ensemble methods – and finds that ensemble techniques, particularly Random Forests, offer high accuracy and efficiency. The experiments, conducted in real-world settings, confirm that an optimized learning-driven approach can substantially reduce downtime and improve system reliability.

Article (Aminu, 2024) focuses on the use of streaming data platforms for real-time anomaly detection. Focusing on architectures based on Apache Kafka and associated streaming APIs, this study demonstrates that event-driven data processing can dramatically reduce latency compared to traditional batch processing, although it may introduce challenges related to message ordering and additional computational overhead.

The unsupervised learning approaches for real-time data anomaly detection (Gupta, 2024) includes methods like clustering (k-means, DBSCAN), dimensionality reduction (PCA, autoencoders), density-based approaches (Isolation Forest, LOF), and one-class SVMs. The analysis concludes that unsupervised methods are highly adaptable for scenarios with limited labeled data, yet challenges such as concept drift and processing efficiency remain. Another study (Balega, 2024) focuses on optimizing IoT security through machine learning-based anomaly detection. It evaluates models including XGBoost, SVM, and deep convolutional neural networks across multiple datasets. Results reveal that XGBoost outperforms the other models in terms of accuracy (up to 99.98%) and training efficiency, making it a strong candidate for securing IoT applications.

As a method of enhancing anomaly detection in IoT systems through explainable AI techniques the XAI-IoT framework was introduced (Gummadi, 2023). The framework integrates both single and ensemble AI models with a suite of XAI tools (such as SHAP, LIME, and ALE) to elucidate feature importance. The evaluation on real-world datasets underscores that incorporating XAI improves transparency without sacrificing detection accuracy.

The literature on anomaly detection in IoT systems includes a broad array of approaches, each designed to address the unique challenges of processing heterogeneous data in real time. In reflecting on the methods proposed by researchers, several key categories emerge, each with its own advantages, disadvantages, and future potential.

Traditional Statistical and Distance-Based Methods rely on established statistical models and distance metrics to determine baseline behavior and identify outliers (Hu, 2020). Their simplicity and inherent interpretability are major advantages; they are straightforward to implement and often require less computing resources (Dickson, 2024). However, as several studies have noted, these methods tend to struggle with the high dimensionality and non-stationary nature of IoT data. Their performance is typically sensitive to the assumptions made about data distribution, which may not hold in dynamic real-world environments. In practice, while these methods can serve as useful baselines, their limited scalability and adaptability to complex, evolving data patterns restrict their long-term applicability.

Supervised and ensemble approaches, such as Random Forests, Support Vector Machines (SVMs), and XGBoost, have been extensively applied to anomaly detection in IoT contexts. These methods have demonstrated high accuracy in detecting subtle anomalies by learning complex decision boundaries from historical data (DeMedeiros, 2023). Ensemble techniques, in particular, help reduce overfitting and improve robustness. However, these models often require extensive data pre-processing, including feature selection and normalization, and they may demand significant computational resources (Tyagi, 2021). Moreover, many supervised approaches depend on large labeled datasets, which can be a major limitation in environments where anomalies are rare or evolving. The trade-off between accuracy and computational efficiency is a recurrent theme, with studies suggesting that while models like XGBoost offer an excellent balance, methods such as SVMs might become impractical when scaled to very large IoT networks (Lee, 2025).

Deep learning methods, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Autoencoders, have shown great promise in automatically extracting hierarchical features from complex data streams (Diro, 2021). These models are particularly effective at capturing non-linear relationships and temporal dependencies in time-series data, which are common in IoT applications (Cook, 2020). Despite their superior accuracy, deep learning approaches have several notable disadvantages: they require large amounts of training data, demand advanced computing infrastructure, and often operate with limited transparency (Jaiswal, 2024). These factors make them challenging to deploy in resource-constrained IoT environments, and their high computational cost can be a critical barrier in real-time applications (Sahu, 2020).

The integration of explainable AI methods has emerged as a crucial advancement, particularly for applications where understanding model decisions is as important as high accuracy (Abououf, 2023). Techniques such as SHAP, LIME, and ALE provide insights into which features contribute most significantly to the detection of anomalies. By making these complex models more transparent, XAI frameworks enable

stakeholders to trust and verify the decisions made by the anomaly detection system (Nguyen, 2024). However, this increased interpretability often comes at the cost of additional computational overhead and complexity in model integration (Abudurexiti, 2025). Despite these challenges, the promise of XAI in introducing more transparent models makes it a valuable direction for future research (Kalutharage, 2023).

Active learning represents a relatively recent approach in which the model iteratively selects the most informative samples for labeling (Nixon, 2024). This strategy is especially beneficial in the IoT context where labeled data are limited and anomalies are rare (Stradiotti, 2024). By reducing the manual labeling load and focusing on samples that are likely to improve the model, active learning can enhance detection performance in real-time environments (Yang, 2018). While the approach can adapt quickly to evolving data patterns and concept drift, its effectiveness heavily depends on the quality of the sample selection strategy and may require continuous human intervention, which can introduce variability and potential delays in model updates (Liao, 2022).

In summary, while traditional statistical and distance-based methods provide valuable simplicity and clear interpretability, their limitations in handling high-dimensional and real-time data suggest that more advanced techniques are necessary for robust anomaly detection in IoT systems (Table 1). Machine learning models, particularly ensemble methods like XGBoost, offer high accuracy but come with increased computational requirements and dependency on extensive pre-processing. Deep learning approaches surpass in feature extraction yet introduce challenges in terms of resource demands and model transparency. XAI frameworks and active learning strategies, meanwhile, provide promising approaches for enhancing model interpretability and adaptability, addressing some of the key disadvantages of other methods. Future research is expected to benefit from hybrid approaches that integrate these techniques, establishing pathways for more scalable, efficient, and transparent anomaly detection systems in IoT environments.

Table 1

Data anomaly detection method categories comparison

Method Category	Complexity	Accuracy	Efficiency	Scalability
Statistical & Distance-Based Methods	Low to moderate	Moderate, effective in simple settings	High, low computational cost	Limited in high-dimensional, non-stationary data
Machine Learning Models	Moderate to high	High, particularly with ensemble models	Moderate, requires extensive pre-processing	Good when optimized, may struggle with very large datasets
Deep Learning Approaches	High	Very high, when trained on large datasets	Low, computationally intensive	Challenging without high-end hardware
Explainable AI (XAI) Frameworks	Moderate to high	Comparable to underlying ML models	Moderate, additional overhead for explanation	Good if integrated with scalable ML models
Active Learning Strategies	Moderate	High with iterative model updates	High, reduces labeling burden over time	Excellent, adapts to evolving data patterns

Formulation of the Article's Objective

Recent studies have advanced the field of real-time anomaly detection in IoT systems by proposing innovative algorithms and frameworks that address the unique challenges created by distributed, high-velocity data. In the survey (Chandola, 2009), the authors present a comprehensive classification of anomaly detection techniques, categorizing methods into statistical, distance-based, and machine learning approaches. They conclude that while numerous techniques exist, significant challenges remain in scaling these methods

and in achieving a balance between accuracy and interpretability. In the article (Krzyszton, 2024) authors offer an experimental comparative analysis of anomaly detection methods in IoT networks. By evaluating both supervised and unsupervised techniques on benchmark datasets, they highlight the trade-offs between model complexity, resource demands, and detection performance. Their findings emphasize that ensemble methods often show the best overall performance while pointing out the limitations of centralized approaches in distributed environments.

The purpose of this article is to present a comprehensive review and comparative analysis of current methods for real-time anomaly detection in distributed IoT systems. By examining traditional statistical methods, distance-based techniques, machine learning models, deep learning approaches, and explainable AI frameworks, this work aims to assess the effectiveness of various approaches, identify the key challenges inherent in distributed IoT environments, and propose future research directions to improve scalability, resource efficiency, and adaptive learning capabilities.

The object of this research is the process of real-time anomaly detection in scalable distributed data processing systems for IoT devices. The subject of this research is real-time anomaly detection methods in distributed IoT systems, focusing on integrating diverse methodologies – from traditional statistical models to deep learning and explainable AI – to effectively address challenges related to data heterogeneity, scalability, and computational efficiency.

Main Results

Understanding Data Anomaly Detection

Data anomaly detection is the process of identifying patterns or individual data points that deviate significantly from expected behavior. In practice, this process is essential for the early identification of issues such as equipment malfunctions, cybersecurity breaches, and operational inefficiencies (Škvára, 2024). The real-life impact of anomaly detection is significant: early detection can prevent catastrophic failures and reduce downtime by ensuring the reliability of critical systems (Wang, 2024).

From a practical standpoint, anomaly detection involves several key steps: establishing a baseline of normal behavior, continuously monitoring real-time data streams, and flagging deviations that could indicate potential issues. This is particularly challenging in IoT environments, where data are often high-dimensional and exhibit non-stationary characteristics (Sakong, 2024). For example, sensor data from industrial IoT applications may fluctuate due to changes in operational conditions, making it difficult to distinguish between normal variability and true anomalies (Nizam, 2022). Fig. 1 shows a time series dataset from an IoT device with outliers that significantly deviate from the usual values, which therefore could be treated as data anomalies.

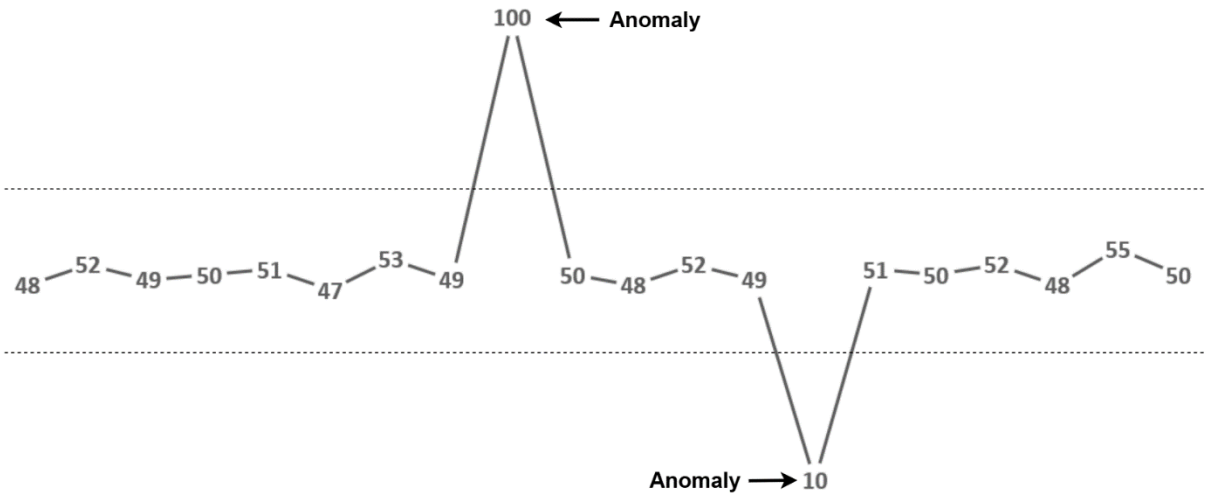

Fig. 1. Data points with significant deviations (anomalies)

A generic anomaly detection pipeline includes essential stages of processing data from IoT devices (Zeng, 2025). The pipeline begins with the collection of raw data from various sensors and devices, followed by data pre-processing where cleaning and normalization ensure data quality. This is succeeded by the feature extraction phase, which involves selecting and transforming relevant features. Subsequently, the model training phase employs machine learning or deep learning techniques to build a predictive model, which is then deployed in the real-time detection stage to monitor incoming data continuously (Fig. 2). Finally, the system transitions to post-processing where detected anomalies are aggregated and alerts are generated. This flowchart provides a clear and structured framework, serving as a blueprint for understanding how raw data is transformed into actionable insights within distributed IoT systems.

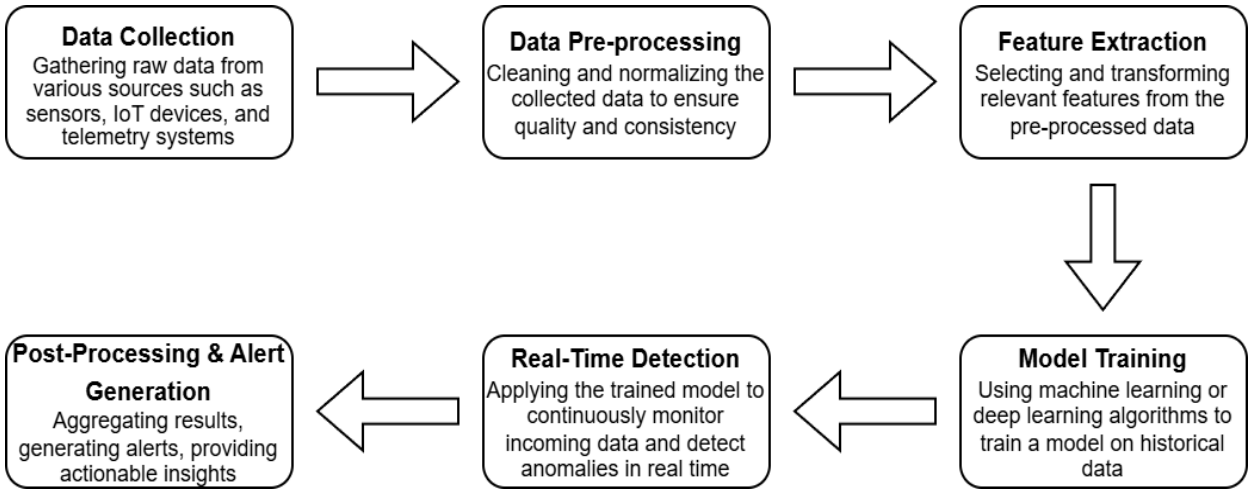

Fig. 2. Generic Anomaly Detection Pipeline Flowchart

Various methods have been proposed to address the challenges of data anomaly detection in a distributed IoT systems:

- **Statistical Methods**

These techniques model data distributions and identify outliers as deviations from the norm. They are computationally light and easily interpretable, but commonly have difficulty with high-dimensional data or non-Gaussian distributions.

- **Distance- and Density-Based Methods**

Approaches such as k-NN and LOF detect anomalies by measuring the distance or density of data points relative to their neighbors. These methods work well in capturing local irregularities; however, they can become computationally expensive and sensitive to the choice of parameters when applied to large datasets.

- **Machine Learning Models**

Supervised classifiers (e.g., SVM, Random Forests, XGBoost) and ensemble methods have demonstrated high accuracy in detecting anomalies by learning complex decision boundaries from historical data. Their major drawback is the need for extensive pre-processing and large labeled datasets, which are not always available in dynamic IoT scenarios (Vajda, 2024).

- **Deep Learning Approaches**

Techniques such as CNNs, RNNs, and autoencoders excel at automatically extracting intricate features from raw data. While these models can achieve impressive detection performance, they demand significant computational resources.

- **Explainable AI (XAI) Frameworks**

Integrating methods like SHAP, LIME, and ALE with traditional ML models helps clarify which features drive anomalies, thereby enhancing confidence in the system's outputs. This transparency is crucial in critical applications, although it may add computational overhead.

- **Active Learning Strategies**

These methods iteratively select the most informative data samples for labeling, thereby reducing the reliance on large labeled datasets and adapting quickly to new or evolving anomalies. Active learning shows promise in environments with dynamic real-time data but depends heavily on an effective sample selection strategy

.

Discussions and Further Research

The reviewed literature indicates that while significant advances have been made, several challenges remain in real-time anomaly detection for distributed IoT systems. Scalability remains a pressing issue as systems must manage ever-increasing data volumes without compromising detection speed (Odoh, 2022). Moreover, methods such as deep learning, although highly accurate, are resource-intensive and less interpretable. Future research should focus on hybrid models that integrate the strengths of statistical, ML, deep learning, and XAI approaches. Federated learning and active learning represent promising avenues to enhance adaptability and reduce labeling load in dynamic environments (Nguyen, 2019). Real-world case studies underscore the practical importance of these methods and the need for continued innovation in this field (Iturbe, 2023).

Conclusions

This article has provided a comprehensive overview of real-time anomaly detection methods in distributed IoT systems. Traditional statistical and distance-based approaches offer simplicity but have disadvantages in handling the complexity of IoT data. In contrast, machine learning and deep learning methods deliver high accuracy yet require significant resources and often lack interpretability. The integration of explainable AI techniques has started to reduce this gap by enhancing the transparency of model decision processes. Finally, active learning strategies offer a promising solution to address the challenges of limited labeled data.

A summary of the reviewed method categories is as follows:

- **Traditional Statistical and Distance-Based Methods** computationally light and highly interpretable due to their reliance on established statistical models and distance metrics, but limited in their ability to handle complex, high-dimensional, and evolving data patterns.
- **Machine Learning Model** sachieve high accuracy in detecting subtle anomalies through complex decision boundaries, yet they often require extensive data pre-processing and large labeled datasets, which can be impractical in dynamic IoT scenarios.
- **Deep Learning Approaches** surpass in automatically extracting intricate features from raw data and capturing non-linear relationships, but their high resource demands limit their deployment in resource-constrained environments.
- **Explainable AI (XAI) Frameworks** significantly enhance transparency by clarifying feature importance, but it increases computational overhead and integration complexity.
- **Active Learning Strategies** reduce the dependency on large labeled datasets and adapts quickly to evolving anomalies, but its effectiveness is heavily dependent on the quality of the sample selection process.

In summary, while each method category has distinct strengths, significant trade-offs exist between accuracy, computational efficiency, scalability, and interpretability. Future research should explore hybrid models that integrate the simplicity and interpretability of traditional methods with the high accuracy of machine learning and deep learning techniques, while also incorporating explainable AI elements to enhance transparency without incurring prohibitive computational costs. Scalability and efficiency must remain central concerns, with particular attention given to distributed and federated learning frameworks that can manage the increasing data volumes generated by IoT systems.

REFERENCES

1. Abououf, M., Singh, S., Rabeb Mizouni, & Hadi Otrok. (2023). Explainable AI for Event and Anomaly Detection and Classification in Healthcare Monitoring Systems. IEEE Internet of Things Journal, 1–1. https://doi.org/10.1109/jiot.2023.3296809
2. Abudurexiti, Y., Han, G., Zhang, F., & Liu, L. (2025). An explainable unsupervised anomaly detection framework for Industrial Internet of Things. Computers & Security, 148, 104130. https://doi.org/10.1016/j.cose.2024.104130
3. Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., & Ming, H. (2019). AD-IoT: Anomaly Detection of IoT Cyberattacks in Smart City Using Machine Learning. IEEE Xplore. https://doi.org/10.1109/CCWC.2019.8666450
4. Aminu, M., Akinsanya, A., Oyedokun, O., Dickson, A., & Dako. (2024). Enhancing cyber threat detection through real -time threat intel
   Technology and Research, 13, 11–27. https://doi.org/10.7753/IJCATR1308.1002
5. Anusha, R. S., Dadavali, S. P., Akash, D., Vinay, M. G., Tapkire, M., & Manjunath, N. (2024). Efficient learning - driven anomaly detection and classification for IoT -based monitorir
   20(11), 3749–3758. https://doi.org/10.52783/jes.8237
6. Balega, M., Farag, W., Wu, X.-W., Ezekiel, S., & Good, Z. (2024). Enhancing IoT security: Optimizing anomaly detection through machine learning. Electronics, 13(11), 2148. https://doi.org/10.3390/electronics13112148
7. Cauteruccio, F., Cinelli, L., Corradini, E., Terracina, G., Ursino, D., Virgili, L., Savaglio, C., Liotta, A., & Fortino, G. (2021). A framework for anomaly detection and classification in Multiple IoT scenarios. Future Generation Computer Systems, 114, 322–335. https://doi.org/10.1016/j.future.2020.08.010
8. Chandola, V., Banerjee, A., & Kumar, V. (F). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58. https://doi.org/10.1145/1541880.1541882
9. Cook, A. A., Mısırlı, G., & Fan, Z. (2020). Anomaly detection for IoT time-series data: A survey. IEEE Internet of Things Journal, 7(7), 6481–6494. https://doi.org/10.1109/JIOT.2019.2958185
10. DeMedeiros, K., Hendawi, A., & Alvarez, M. (2023). A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks. Sensors, 23(3), 1352. https://doi.org/10.3390/s23031352
11. Dickson, S. M. (2024). Detection of anomalies in Internet of Things (IoT) devices and sensors. Radinka Journal of Science and Systematic Literature Review, 2(3), 475–481. https://doi.org/10.56778/rjslr.v2i3.347
12. Diro, A., Chilamkurti, N., Nguyen, V.-D., & Heyne, W. (2021). A Comprehensive Study of Anomaly Detection Schemes in IoT Networks Using Machine Learning Algorithms. Sensors, 21(24), 8320. https://doi.org/10.3390/s21248320
13. Gad, I. M. (2025). TOCA-IoT: Threshold optimization and causal analysis for IoT network anomaly detection based on explainable random forest. Algorithms, 18, 117. https://doi.org/10.3390/a18020117
14. Giannoni, F., Mancini, M., & Marinelli, F. (2018). Anomaly Detection Models for IoT Time Series Data. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.1812.00890
15. Gummadi, A. N., Napier, J. C., & Abdallah, M. (2023). XAI-IoT: An explainable AI framework for enhancing anomaly detection in IoT systems. IEEE Access. https://doi.org/10.1109/ACCESS.2023.0322000
16. Gupta, P., & Tripathy, P. (2024). Unsupervised learning for real-time data anomaly detection: A comprehensive approach. SSRG International Journal of Computer Science and Engineering, 11(10), 1-11. https://doi.org/10.14445/23488387/IJCSE-V11I10P101
17. Hu, X., Xu, Q., & Guo, Y. (2020). Trajectory anomaly detection based on the mean distance deviation. Communications in Computer and Information Science, 140–147. https://doi.org/10.1007/978-3-030-63820-7_16
18. Idhalama, O., & Oredo, J. (2024). Exploring the next generation Internet of Things (IoT) requirements and applications: A comprehensive overview. Information Development. https://doi.org/10.1177/02666669241267852
19. Iturbe, J., & Rifà-Pous, H. (2023). Anomaly-based cyberattacks detection for smart homes: A systematic literature review. Internet of Things, 22, 100792. https://doi.org/10.1016/j.iot.2023.10079
20. Jaiswal, A., & Koupaei, A. N. (2024). Deep comparison analysis: Statistical methods and deep learning for network anomaly detection. International Journal of Computer Science and Information Security, 22. https://doi.org/10.5281/zenodo.14051106
21. Jot, J., & Sharma, L. (2023). Study of anomaly detection in IoT sensors. International Journal for Research in Applied Science and Engineering Technology, 11, 767–774. https://doi.org/10.22214/ijraset.2023.55226
22. Kalutharage, C. S., Liu, X., Chrysoulas, C., Pitropakis, N., & Papadopoulos, P. (2023). Explainable AI-Based DDOS Attack Identification Method for IoT Networks. Computers, 12(2), 32. https://doi.org/10.3390/computers12020032

23. Kaya, M. O., Ozdem, M., & Das, R. (2025). A novel approach for graph-based real-time anomaly detection from dynamic network data listened by Wireshark: A novel approach for graph-based real-time anomaly detection. EAI Endorsed Transactions on I ndustrial Networks and Intelligent Systems, 12. https://doi.org/10.4108/eetinis.v12i2.7616
24. Krzyszton, E., Rojek, I., & Mikołajewski, D. (2024). A comparative analysis of anomaly detection methods in IoT networks: An experimental study. Applied Sciences, 14, 11545. https://doi.org/10.3390/app142411545
25. Lee, C.-Y., & Maceren, E. D. (2025). Physics-informed anomaly and fault detection for wind energy systems using deep CNN and adaptive elite PSO-XGBoost. IET Generation, Transmission & Distribution, 19(1). https://doi.org/10.1049/gtd2.13289
26. Liao, N., & Li, X. (2022). Traffic Anomaly Detection Model Using K-Means and Active Learning Method. International Journal of Fuzzy Systems, 24(5), 2264–2282. https://doi.org/10.1007/s40815-022-01269-0
27. Martins, I., Resende, J. S., Sousa, P. R., Silva, S., Antunes, L., & Gama, J. (2022). Host-based IDS: A review and open issues of an anomaly detection system in IoT. Future Generation Computer Systems, 133, 95–113. https://doi.org/10.1016/j.future.2022.03.001
28. Mutambik, I. (2024). Enhancing IoT security using GA-HDLAD: A hybrid deep learning approach for anomaly detection. Applied Sciences, 14(21), 9848-9848. https://doi.org/10.3390/app14219848
29. Nguyen, M.-D., La, V.-H., Mallouli, W., Cavalli, A. R., & Oca, E. M. de. (2023). Toward Anomaly Detection Using Explainable AI. CyberSecurity in a DevOps Environment, 293–324. https://doi.org/10.1007/978-3-031-42212-6_10
30. Nguyen, T. D., Marchal, S., Miettinen, M., F ereidooni, H., Asokan, N., & Sadeghi, A.-R. (2019). DÏoT: A Federated Self-learning Anomaly Detection System for IoT. 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). https://doi.org/10.1109/icdcs.2019.00080
31. Nixon, C., Sedky, M., Champion, J., & Hassan, M. (2024). SALAD: A split active learning based unsupervised network data stream anomaly detection method using autoencoders. Expert Systems with Applications, 248, 123439. https://doi.org/10.1016/j.eswa.2024.123439
32. Nizam, H., Zafar, S., Lv, Z., Wang, F., & Hu, X. (2022). Real-Time Deep Anomaly Detection Framework for Multivariate Time-Series Data in Industrial IoT. IEEE Sensors Journal, 1–1. https://doi.org/10.1109/jsen.2022.3211874
33. Odoh, K. (2022). Real-time Anomaly Detection for Multivariate Data Streams. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2209.12398
34. Ukil, A., Bandyoapdhyay, S., Puri, C., & Pal, A. (2016). IoT Healthcare Analytics: The Importance of Anomaly Detection. 2016 I EEE 30th International Conference on Advanced Information Networking and Applications (AINA). https://doi.org/10.1109/aina.2016.158
35. Sahu, N. K., & Mukherjee, I. (2020). Machine Learning based anomaly detection for IoT Network: (Anomaly detection in IoT Network). IEEE Xplore. https://doi.org/10.1109/ICOEI48184.2020.9142921
36. Sakong, W., Kwon, J., Min, K., Wang, S., & Kim, W. (2024). Anomaly Transformer Ensemble Model for Cloud Data Anomaly Detection. IEEE Transactions on Cloud Computing, 12(4), 1305–1313. https://doi.org/10.1109/TCC.2024.3466174
37. Sedjelmaci, H., Senouci, S., & Al-Bahri, M. (2016). A lightweight anomaly detection technique for low-resource IoT devices: A game-theoretic methodology. HAL (Le Centre Pour La Communication Scientifique Directe). https://doi.org/10.1109/icc.2016.7510811
38. Stradiotti, L., Perini, L., & Davis, J. (2024). Combining active learning and learning to reject for anomaly detection. In Frontiers in Artificial Intelligence and Applications. https://doi.org/10.3233/FAIA240749
39. Škvára, V., Smidl, V., & Pevný, T. (2024). Anomaly detection in multifactor data. Neural Computing and Applications, 36(34), 21561–21580. https://doi.org/10.1007/s00521-024-10291-2
40. Tyagi, H., & Kumar, R. (2021). Attack and Anomaly Detection in IoT Networks Using Supervised Machine Learning Approaches. Revue d'Intelligence Artificielle, 35(1), 11–21. https://doi.org/10.18280/ria.350102
41. Vajda, D. L., Do, T. V., Bérczes, T., & Farkas, K. (2024). Machine learning-based real-time anomaly detection using data pre-processing in the telemetry of server farms. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-72982-z
42. Wang, C., & Zhu, H. (2024). Enhancing data for hard anomaly detection. In Universal Behavior Computing for Security and Safety, 2, 45–56. https://doi.org/10.1007/978-981-97-9014-2_2
43. Yang, K., Ren, J., Zhu, Y., & Zhang, W. (2018). Active Learning for Wireless IoT Intrusion Detection. IEEE Wireless Communications, 25(6), 19–25. https://doi.org/10.1109/mwc.2017.1800079
44. Zeng, F., Wang, M., Pan, Y., Lv, S., Huiyu, M., Han, H., & Yuan, X. (2025). Distributed data privacy protection via collaborative anomaly detection. Electronics, 14(2), 295. https://doi.org/10.3390/electronics14020295
45. Zakariah, M., & Almazyad, A. S. (2023). Anomaly detection for IoT systems using active learning. Applied Sciences, 13(21), 12029. https://doi.org/10.3390/app132112029

ВИЯВЛЕННЯ АНОМАЛІЙ У РЕАЛЬНОМУ ЧАСІ В РОЗПОДІЛЕНИХ IOT-СИСТЕМАХ: КОМПЛЕКСНИЙ ОГЛЯД ТА ПОРІВНЯЛЬНИЙ АНАЛІЗ

Павло Пустельник[1], Євгенія Левус[2]

Національний університет "Львівська політехніка",
кафедра програмного забезпечення, Львів, Україна
[1] E-mail: pavlo.y.pustelnyk@lpnu.ua, ORCID: 0009-0005-5745-9941
[2] E-mail: yevheniia.v.levus@lpnu.ua, ORCID: 0000-0001-5109-7533

Стрімке поширення технології Інтернету речей (IoT) призвело до безпрецедентного росту обсягів неоднорідних даних з розподілених пристроїв. Цей величезний потік даних робить все більш важливим впровадження надійних і ефективних методів виявлення аномалій в режимі реального часу, які можуть попередити про проблеми у розподілених системах. Виявлення аномалій даних є критично важливим у сучасному світі, оскільки воно дозволяє на ранній стадії виявляти відхилення, які можуть свідчити про збої в роботі системи, порушення безпеки або операційну неефективність. Вчасне виявлення цих відхилень може скоротити час простою, оптимізувати продуктивність і захистити критично важливі активи.

Стаття містить огляд і порівняльний аналіз сучасних методів виявлення аномалій у розподілених системах, заснованих на технології Інтернету речей. У ній розглядається широкий спектр методів, таких як традиційні статистичні підходи, дистанційні методи, моделі машинного навчання, алгоритми глибинного навчання і методи пояснювального штучного інтелекту. Кожна категорія оцінюється з точки зору точності виявлення, обчислювальної ефективності та інтерпретованості. Реальні приклади – від прогнозування технічного обслуговування в промисловому IoT та управління енергією в розумних мережах до виявлення порушень у фінансових мережах – демонструють широке практичне застосування цих методів.

В огляді також визначено поточні виклики і перспективні напрямки майбутніх досліджень, зокрема федеративне навчання і підходи, засновані на активному навчанні, які пропонують потенційні рішення для підвищення адаптивності і зменшення залежності від великих маркованих наборів даних. Висновки, зроблені в цьому огляді, створюють основу для майбутніх досліджень, спрямованих на розроблення гібридних моделей виявлення аномалій, які інтегрують передові методи для подальшого підвищення адаптивності та безпеки систем в динамічних середовищах Інтернету речей.

Ключові слова: Інтернет речей, виявлення аномалій, опрацювання в реальному масштабі часу, машинне навчання, глибинне навчання, пояснювальний ШІ, розподілені системи.