Journal of Inventive and Scientific Research Studies (JISRS)
www.jisrs.com Vol: II, Issue: 2 December 2024 ISSN: 2584-0630 (Online)

BEYOND NORMAL: A SURVEY OF NETWORK ANOMALY DETECTION TECHNIQUES

S. INDIRA
Assistant Professor, Department of Computer Science
G. Venkataswamy Naidu College (Autonomous), Kovilpatti.
E-mail: s.indirajune@gmail.com

Received: June 15, 2023, **Accepted**: August 21, 2024, **Online Published**: December 15, 2024

ABSTRACT

In many applications (e.g., healthcare, smart homes and industrial systems) the security of Internet-of-Things devices depends on detecting anomalies for protecting against cyber attacks. It was examined how the use of machine learning and deep learning techniques enhanced a anomaly detection in such domains. The primary objective of this review is to survey on approach that have been developed by researchers in detecting the anomalies present over network traffic and henceforth it also provide some insight for future enhancement. This work will help researchers in the fields of network security, IoT and machine learning. More broadly, Anomaly Detection (AD) plays an important part in the protection and integrity of Internet-connected devices within networks such as Social Network Services (SNSs) or enabling various aspects of Secure IoT. Here the recent trends of deployment machine learning (ML) and deep learning (DL) in anomaly detection (AD) were explored through this study. The purpose of this paper is to compile an extensive literature review related to AD, which can help in understanding the state-of-the-art practices in the field and analysing key research limitations as well as hint potential directions for future. Therefore, although this survey has provided important insights into the state-of-the-art of AD approaches and their application in IoT by examining literature published from 2018 to 2023 with a critical eye, it seems that realization/acceptance is the way forward.

Keywords: Deep Anomaly Detection, IoT Security, Deep Learning, Sensors, and Intrusion Detection.

Introduction

Newer challenges in securing and operating diverse IoT solutions along the lines of smart home, healthcare, industrial process or infrastructure too have surfaced with explosion in those areas of domain. To address these problems, anomaly detection (AD) has emerged as one of the key approaches for detecting security breaches, system failures and abnormal performance by observing patterns that differ from an expected norm. Conventional AD techniques like statistical analysis and clustering have advanced this field. However, the complexity and volume of data generated by networks today necessitate the use of more advanced methods, particularly with the assistance of machine learning (ML) and deep learning (DL), as illustrated in Figure 1, which provides an overview of the IoT and Anomaly Detection process.

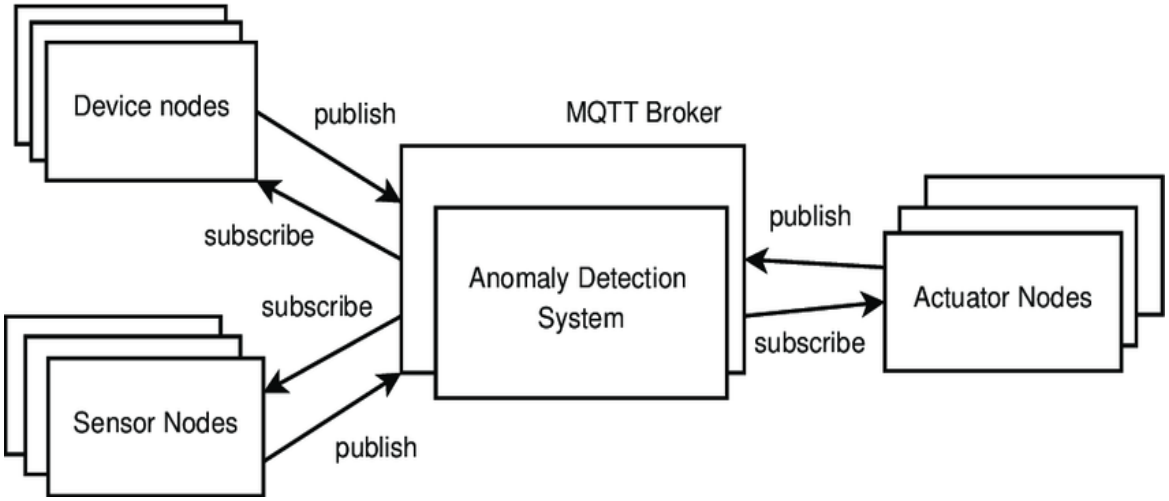

Figure 1: Overview of IoT and Anomaly Detection [1]

Modern research has also outlined several machine learning (ML) and deep learning (DL) based anomaly detection (AD) methods that are suitable for IoT and sensor networks. These techniques enhance effectiveness in managing the increased scale and flow of data in real-time, thereby offering greater precision and efficiency in identifying anomalies. This paper discusses these developments, specifically focusing on the usefulness, applicability, and transferability of these studies across different networks. The survey also highlights previous research limitations and provides recommendations for future studies concerning the integration of hybrid models and the design of improved and scalable AD systems.

Related Works and Approaches

Machine Learning-Based Anomaly Detection

As a robust technique of data analysis, machine learning (ML) has found its way to AD in network security; besides being versatile in relation to the type of data and network. SVM and decision tree are

some of the typical nominees of ML that has been widely used in AD since it can judge anomalies by comparing them to set patterns. Nevertheless, the accuracy of these algorithms decreases with the size of the data set and in the context of dynamically changing networks, these methods have been further advanced.

A flowchart outlining the typical steps involved in an ML-based anomaly detection process.

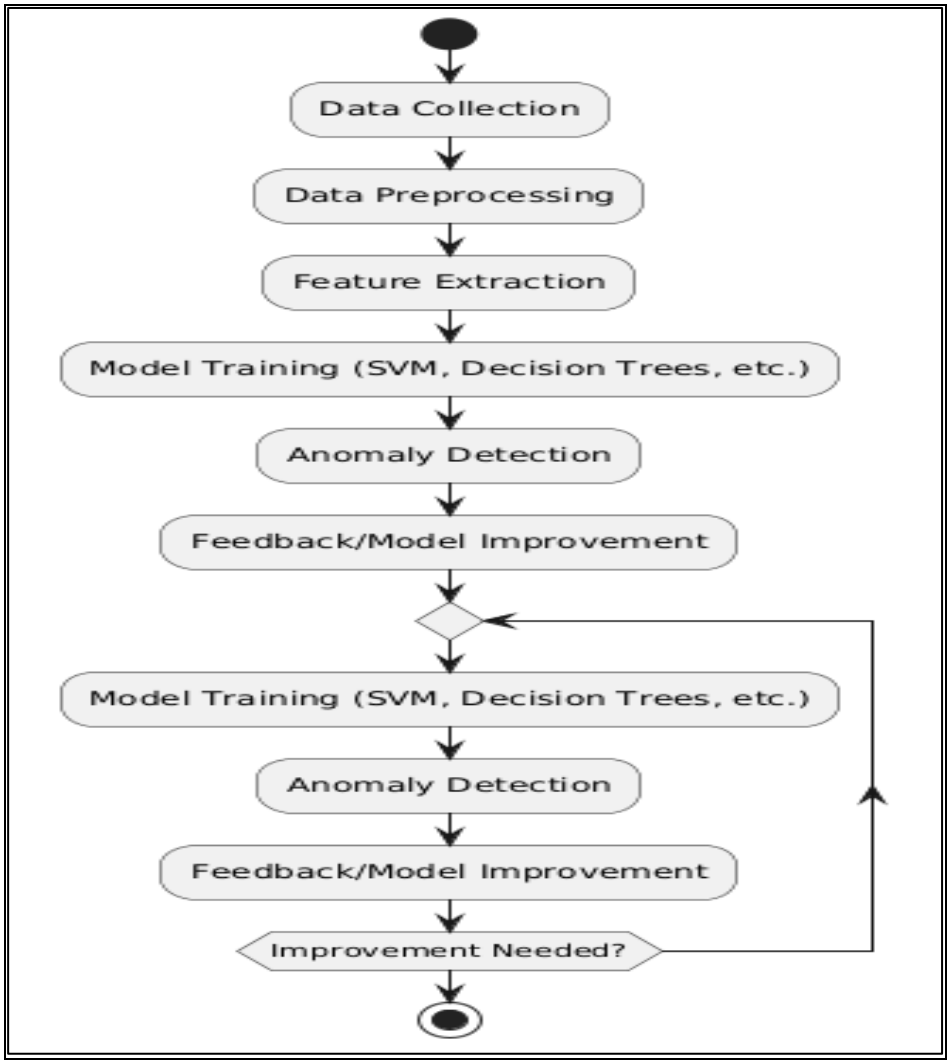

Figure 2: Flowchart of Machine Learning-Based Anomaly Detection Process

The typical steps involved in an ML-based anomaly detection process are outlined in Figure 2. As it has been investigated in the recent works, ensemble learning methods, which combine several of them, are shown to be more accurate and less sensitive to differences. For instance, Zhang et al. (2019) suggested the integration of random forests and gradient boosting machines to outcompete each other and single models in the classification of network intrusions [2]. Equally, the study by Kim et al. (2020) demonstrated how the use of ensemble

methods for IoT was beneficial, especially in cases of imbalanced data [3].

A comparison of different ML algorithms, such as SVM, Decision Trees, Ensemble Learning, and Autoencoders, is shown in Table 1. This comparison is based on criteria like accuracy, scalability, applicability to IoT, strengths, and weaknesses. For example, SVMs are effective on small datasets but less effective on large, dynamic datasets, while ensemble learning methods combine the strengths of multiple models, making them highly accurate and scalable.

Algorithm	Accuracy	Scalability	Applicability	Strengths	Weaknesses
SVM	High	Moderate	Good	Effective on small datasets	Less effective on large, dynamic datasets
Decision Trees	Moderate	High	Good	Simple to implement	Prone to overfitting
Ensemble Learning	Very High	High	Excellent	Combines strengths of multiple models	Complex to implement
Autoencoders	High	High	Excellent	Handles unlabeled data	Requires large datasets for training

Table 1: Comparison of Machine Learning Algorithms for Anomaly Detection

Moreover, development in autoencoders and clustering algorithms, which comes under Unsupervised learning have made the methods efficient to identify unknown noise. Wang et al proposed an approach for anomaly detection in network traffic based upon autoencoder that extracts a low-level feature of the network traffic without requiring labeled data for training [4]. Being an unsupervised learning model, this has been especially valuable in constantly evolving IoT contexts due to the rarity of labeled data.

Deep Learning-Based Anomaly Detection

DL is currently crucial in AD as it offers effective methods for processing extensive and multi-feature data. CNNs and RNNs are some of the most frequently employed DL architectures in this field. CNNs are more suitable for handling data with a spatial relationship like images and sensor data while RNNs are ideal in handling data with temporal dependency, which is important in analyzing the time-series data that is characteristic of the IoT systems.

The combination of DL with other conventional methods of ML has continue to improve the performance of AD systems. For example, Xu et al. (2020) suggested a CNN-SVM model to improve the results of the traditional approach in

detecting anomalies in industrial IoT networks [5]. This approach leverages CNNs for feature extraction and SVMs for classification, resulting in a more effective AD system.

Apart from CNNs and RNNs, there is another type of network known as the generative adversarial networks (GANs), which appears to be useful for AD. GANs are more useful in providing synthesized data for training purposes, in sequence solving the problem of scarcity of data in AD applications. Li et al. (2022) showed that GANs are useful for In power gird networks anomaly detection where labeled data is scarce and hard to obtain [6]. With the help of GAN-based model the AD system enhanced the detection accuracy and the model was more robust when dealing with real life data than synthetic data.

Hybrid Approaches and Cross-Domain Applications

Hybrid approaches of AD have been introduced due to the increasing uncertainties and complexities of IoT environments and diversification of AD techniques. These approaches make use of the features of the mentioned methods while overcoming the weaknesses of each. For instance, the integration of statistical methods with the ML and DL has also displayed effectiveness in identifying complicated patterns or anomalies that are difficult to detect using a single method, as illustrated in Figure 3.

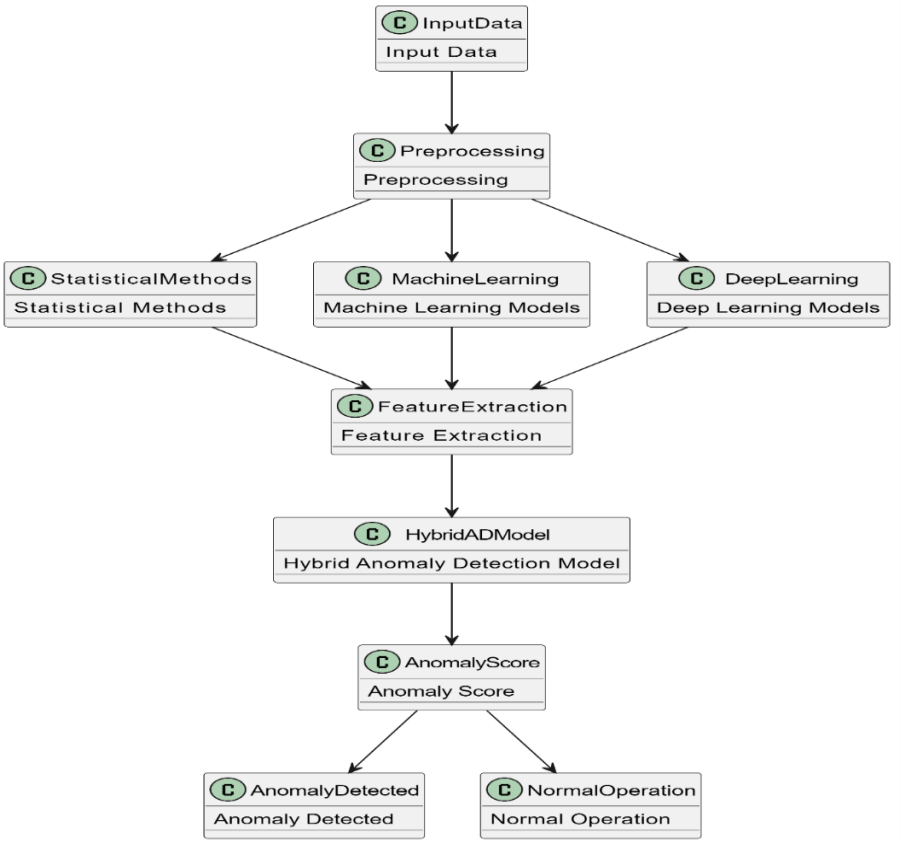

Figure 3: Hybrid Anomaly Detection Model

Another research by Chen et al. (2021) has investigated incorporating SPC along with RNNs which resulted in enhancements in terms of anomalies identification in the industrial IoT [7]. The SPC component oversees the system's general functioning, while the RNN component looks for distinct temporal patterns associated with undesired behaviors. The integration of these two led to the general and faster AD system which is capable of both known and unknown AD.

Table 2 comparing single anomaly detection models (e.g., ML-only, DL-only) with hybrid models.

Model Type	Detection Accuracy	Robustness	Computational Efficiency	Flexibility
ML-only	Moderate	Low	High	Low
DL-only	High	Moderate	Moderate	High
Hybrid	Very High	High	Moderate	Very High

Table 2: Comparison of Single vs. Hybrid Anomaly Detection Models

Other emerging area of research in AD techniques is implementing AD techniques in other areas of electronics or computer science. These applications entail the migration of AD models constructed in one domain to another with the use of similarities in data patterns and network topology. For instance, Zhang et al. (2020) have shown that an AD model developed on the processing of healthcare data can be directly applied in a smart home setting with reasonable levels of accuracy and with only a tiny amount of fine-tuning required [8]. This approach gives emphasis on investigation of cross-domain AD techniques that ultimately lead to the vigorous reduction of the new model development time in the new environment.

Challenges and Limitations

However, the existing ML and DL-based AD techniques are also associated with some problems and limitations. The first problem is scalability The idea of scalability is important here because Wikipedia currently comprises hundreds of thousands of entries, and someone who wants to use the site to find information has to search through all these entries to find the necessary data. IoT networks can contain a very large number of devices, belonging to different categories, that send a very large number of messages which have to be processed quickly. These environments can prove challenging for Traditional AD techniques because as the number of endpoints increases, getting a hold of them becomes problematic and there is usually a longer time taken for an attack to be detected, and the number of false positives rises.

Figure 4 summarizing the main challenges in anomaly detection for IoT, such as scalability, data diversity, real-time processing, and detecting sophisticated attacks.

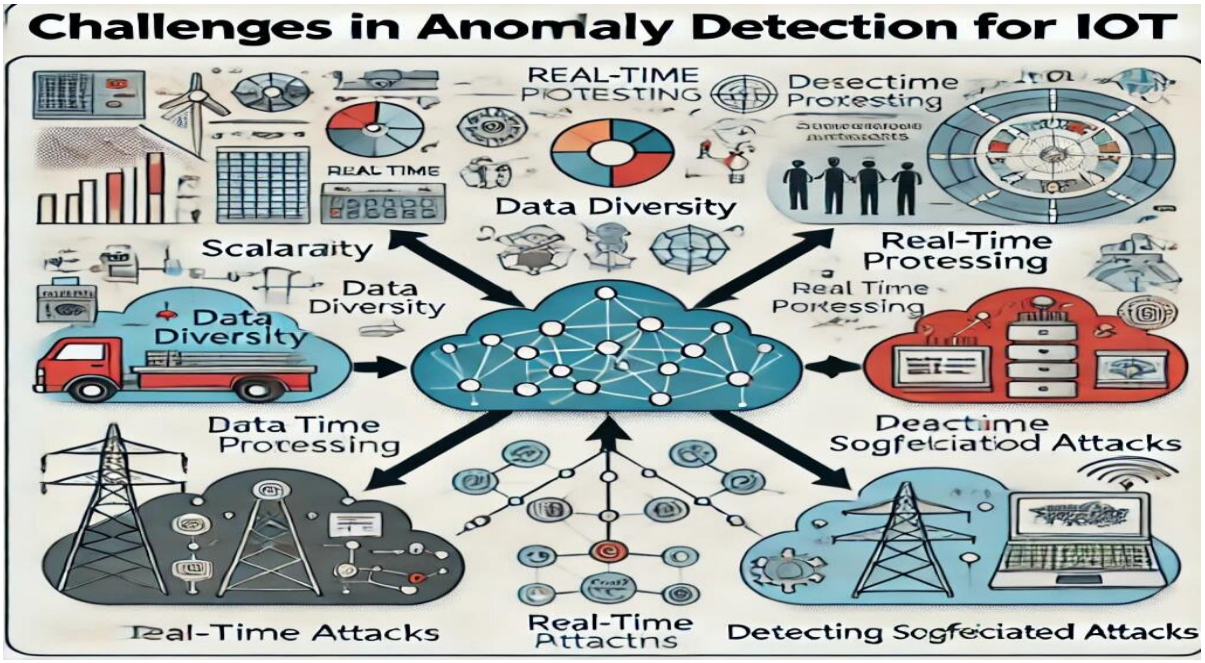

Figure 4: Challenges in Anomaly Detection for IoT

In order to overcome this problem, scholars have investigated various forms of distributed AD techniques, particularly forms of AD models applied in multiple nodes. For instance, Zhou et al. (2021) developed a federated learning-based distributed AD framework in which multiple models are trained locally on edge devices and are then, combined together to form a global model [9]. It also decreases the load of computation to the central servers that makes the scalability of system improved in the AD system.

Another large problem is the problem with discovering modern and hidden threats, including APTs. These attacks will usually bypass conventional AD techniques by mimicking the norm of the network or leveraging on the system weakness. As a result, using DL-based approaches, for instance, deep reinforcement learning (DRL), could be beneficial for solving this problem since defense approaches are gotten through training with the environment. Wang et al. (2022) presented a study that investigated the application of DRL in identifying APTs in IoT environment and establish that DRL provides better detection rates than conventional detection techniques [10].

In addition, extension of the DL models for AD is another significant problem for researchers, and interpretability is an essential point. However, DL models achieve high accuracy; however, they are black boxes and, therefore, lack interpretability. Such opacity can become an issue to the

application of DL-based AD systems especially in high risk areas that require extensive explanation. However, to achieve this, researchers have come up with what is known as explainable AI, or XAI, which are methods that give a closer look into the decision-making process done by DL models. For instance, Li et al. (2021) initially proposed an XAI framework for AD in industrial IoT concerning visualization of contributions of the features measured to the model's decisions, enhancing the reliability of the system [11].

Applications of Anomaly Detection

Anomaly detection is one of the most important functions present in numerous IoT applications such as smart city, healthcare, industrial, and environment. In smart cities, AD techniques also prove effective at observing and controlling elements of the city's framework like traffic, energy, and safety systems. For instance, Zhang et al. (2022) designed CNN based AD system to identify abnormalities in traffic flow dynamics so as to control the traffic flow effectively, thus, avoiding traffic congestions [12].

Table 3 listing different application domains (e.g., smart cities, healthcare, industrial automation, environmental monitoring) and the corresponding AD techniques or models used in each domain, along with their benefits.

Application Domain	AD Technique/Model	Benefits
Smart Cities	CNN-based models	Real-time traffic management
Healthcare	RNNs for ECG analysis	Early detection of arrhythmias
Industrial Automation	Hybrid SPC-DL models	Enhanced operational reliability
Environmental Monitoring	Deep Learning for Water Quality	Early detection of pollution

Table 3: Applications of Anomaly Detection in Various Domains

In healthcare, AD is crucial for keeping vital signs and diagnosing symptoms of illnesses on clients' statuses. Wearable devices and remote monitoring systems provide constant Physiological signals to Analyze by applying AD techniques to detect out of Norm Physiology of health ailments. Kim et al. (2022) used RNNs in the diagnosis of cardiac abnormalities from ECG signals and the method developed a high accuracy in diagnosing arrhythmias [13]. This technique underscores how AD in the case of patient ABC has the propensity of having better outcomes of the patient when diagnosed earlier.

Industrial automation is another domain that has significant application of AD where it is used as the ultimate guarantee in determining the reliability and safety of the operations. Industry, especially manufacturing, is one of the sectors wherein IoT devices are vastly employed wherein they perform the role of monitoring equipment and identify signs of possible failure. Chen et al. (2022) put forward a hybrid AD model to integrate SPC and DL approaches for identifying the changes of performance of industrial robots during operation and the study achieved a satisfactory outcome [14]. But besides enhancing the detection accuracy, it also offered solutions for the subsequent preventive maintenance.

Environmental monitoring is yet another domain

where techniques of AD are widely used. Smart devices used in environmental applications measure and transmit data of different attributes including temperature, humidity, presence of gases, and water levels, among others. All these data streams are processed in real-time with the intentions of trying to identify irregularities that are likely to point to natural disasters or environmental risks. For instance, Zhang et al., in their study in 2021 designed and constructed an AD system for water quality data anomaly detection in order to promptly identify cases of pollution [15].

Conclusion

However, with advancements in machine learning and deep learning anomaly detection systems for IoT and sensor networks have come a long way. The proposed schemes provide additional accuracy, scalability, and configurability making them attractive for various applications such as smart cities, healthcare systems; industrial automation, and environmental monitoring. However, there are many challenges to address such as low scaling capabilities of AD systems, detection of advanced attacks, and interoperability of DL models. These challenges can be addressed with ongoing research that focuses on developing hybrid systems combining the key elements of various AD technologies. Additionally, cross-domain applications and the incorporation of explainable AI techniques are prospective directions for future work in anomaly detection with such complex and dynamic IoT environments.

References

M. A. Bin Ahmadon and S. Yamaguchi, “Verification method for accumulative event relation of message passing behavior with process tree for IoT systems,” *Information (Switzerland)*, vol. 11,

no. 4, Apr. 2020, doi: 10.3390/INFO11040232.

Y. Zhang, H. Liu, J. Li, and H. Zhang, "Ensemble learning for network intrusion detection based on random forest and gradient boosting," *IEEE Access*, vol. 7, pp. 125201–125210, 2019. doi: 10.1109/ACCESS.2019.2939266.

M. Kim, J. Kim, and H. Lee, "Ensemble learning for anomaly detection in imbalanced IoT data," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6719–6730, 2020. doi: 10.1109/JIOT.2020.2971760.

J. Wang, Q. Zhang, and Y. Cao, "An autoencoder-based anomaly detection method for network traffic," *IEEE Access*, vol. 9, pp. 120019–120029, 2021. doi: 10.1109/ACCESS.2021.3108976.

Y. Xu, F. Zhang, and Z. Wang, "A hybrid deep learning model for anomaly detection in industrial IoT," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5744–5754, 2020. doi: 10.1109/JIOT.2019.2955785.

Y. Li, H. Chen, and X. Wang, "GAN-based anomaly detection in power grid networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2391–2400, 2022. doi: 10.1109/TII.2021.3118734.

Y. Chen, H. Sun, and W. Liu, "A hybrid anomaly detection model for industrial IoT systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5441–5450, 2021. doi: 10.1109/TII.2021.3059967.

Y. Zhang, H. Zheng, and Q. Li, "Cross-domain anomaly detection in IoT environments: A case study in smart homes," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8850–8861, 2020. doi: 10.1109/JIOT.2020.2979075.

Z. Zhou, X. Wu, and S. Li, "A distributed anomaly detection framework based on federated learning for IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3616–3626, 2021. doi: 10.1109/JIOT.2020.3045635.

J. Wang, L. Liu, and Y. Li, "Deep reinforcement learning for advanced persistent threat detection in IoT networks," *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1565–1577, 2022. doi: 10.1109/TNSM.2022.3164951.

Y. Li, H. Zhang, and F. Wu, "Explainable AI for anomaly detection in industrial IoT systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5548–5557, 2021. doi: 10.1109/TII.2021.3054339.

Y. Zhang, Z. Liu, and H. Wang, "A CNN-based anomaly detection system

for real-time traffic management in smart cities," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 2851–2860, 2022. doi: 10.1109/TITS.2021.3071479.

M. Kim, J. Kim, and S. Lee, "RNN-based cardiac anomaly detection using ECG signals," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 152–160, 2022. doi: 10.1109/TBME.2021.3108174.

Y. Chen, Z. Zhang, and Q. Wang, "Hybrid anomaly detection for industrial robots using statistical process control and deep learning," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 5, pp. 4774–4783, 2022. doi: 10.1109/TIE.2021.3071451.

Y. Zhang, L. Wang, and J. Liu, "Anomaly detection in water quality data using deep learning and IoT," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 2901–2910, 2021. doi: 10.1109/TII.2020.3036965.