International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 6, Nov-Dec 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Anomaly Detection in Industrial IoT Sensor Data Using SVC and Random Forest: A Review

Somoshi Kavita M.[1], Bansode Rahul S.[2]

[1]ME Student (VLSI& ES), [2]Assistant Professor,
[1,2]SPCOE, Otur, Maharashtra, India

ABSTRACT

The Internet of Things (IoT) has transformed the way we interact with our surroundings by enabling large-scale connectivity among billions of devices. These connected devices continuously exchange data, resulting in massive and rapidly growing data volumes. However, this expansion also brings a major concern: ensuring the reliability and security of IoT-generated data. Detecting data anomalies is essential for identifying unusual patterns, system deviations, and potential cyber threats within IoT environments.

This paper reviews the latest developments, commonly used approaches, and ongoing challenges in IoT data anomaly detection. We examine the advantages and limitations of different detection techniques, including statistical models, machine learning approaches, and deep learning methods. The discussion also highlights the unique characteristics of IoT data—such as heterogeneity, scalability requirements, real-time processing demands, and privacy issues—that make anomaly detection particularly complex. By addressing these factors, this review aims to provide a comprehensive understanding of the difficulties involved and guide the development of more accurate and efficient anomaly detection solutions for IoT systems.

How to cite this paper: Somoshi Kavita M. | Bansode Rahul S. "Anomaly Detection in Industrial IoT Sensor Data Using SVC and Random Forest: A Review" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-6, December 2025, pp.436-442, URL: www.ijtsrd.com/papers/ijtsrd99872.pdf

IJTSRD99872

KEYWORDS: *Cloud Computing, PaaS, Library.*

I. INTRODUCTION

The Internet of Things (IoT) refers to a network of smart devices, sensors, and systems that collect, share, and respond to data in real time. These devices communicate with each other and with users through embedded sensors, software, and internet connectivity. IoT technology supports a wide range of applications, including smart homes, smart cities, industrial automation, transportation, and healthcare. By linking physical objects to the digital world, IoT improves operational efficiency, supports better decision-making, and provides new levels of automation and connectivity in everyday life. A typical IoT structure is made up of three essential layers: the perception layer, the network layer, and the application layer. The perception layer, also called the sensing or physical layer, contains devices and sensors that collect data from the environment. The network layer is responsible for transmitting this data between devices and systems using communication technologies such as Wi-Fi, Bluetooth, Zigbee, cellular networks, and lightweight IoT protocols like MQTT and CoAP. At the top, the application layer uses the collected data to deliver services, insights, automation, and control functions. The rapid growth of IoT has significantly transformed sectors such as healthcare, industry, transportation, agriculture, and education. However, as these devices generate vast amounts of data, ensuring the accuracy, security, and reliability of this data becomes a critical challenge. Abnormal data patterns, outliers, and unexpected behaviors can threaten system performance, expose security weaknesses, or indicate cyberattacks. Detecting such anomalies early is essential to prevent disruptions, protect sensitive information, and maintain trust in IoT systems. Effective anomaly detection helps identify system faults, equipment failures, and security breaches before they escalate. It also improves system performance, supports better decision-making, and ensures high-quality data. By identifying unusual patterns, organizations can optimize resources, improve operations, and gain valuable insights from IoT data. This is especially important in sensitive fields like education and agriculture, where IoT systems collect large volumes of data that support innovation and efficiency. Detecting anomalies protects these systems from malicious activities and operational risks, ensuring data integrity and maintaining system reliability. As IoT deployments continue to grow in scale and complexity, the need for advanced and efficient anomaly detection techniques becomes even more important. These techniques must handle large, diverse, and continuously changing data streams while providing real-time, accurate results. As a result, researchers are exploring new anomaly detection algorithms, improved data preprocessing methods, and the integration of anomaly detection with real-time analytics platforms.

The main contributions of this study are as follows:

- To provide a detailed review of current trends and techniques used for anomaly detection in IoT systems.
- To identify major challenges and limitations related to anomaly detection in IoT environments.
- To evaluate existing methods and algorithms, including statistical techniques, machine learning approaches, and anomaly-scoring models.
- To examine how characteristics of IoT data—such as high dimensionality, heterogeneity, and dynamic behavior—affect anomaly detection performance.
- To propose strategies for improving the accuracy, efficiency, and scalability of IoT anomaly detection methods.
- To highlight open research issues and future directions to guide researchers and practitioners working in this field.

The rest of the paper is organized as follows. Section II reviews various data anomaly detection techniques. Section III outlines the challenges associated with anomaly detection in IoT environments. Section IV provides a discussion, while Section V presents future research directions. Section VI concludes the paper.

II. DATA ANOMALY DETECTION STRATEGIES IN IOT

Data anomaly detection in the Internet of Things (IoT) focuses on identifying unusual or abnormal patterns in the data produced by connected devices. Since IoT systems continuously generate large volumes of information, detecting irregularities is essential for identifying security threats, equipment malfunctions, system failures, or unexpected behavior. Several techniques are used for anomaly detection, with statistical methods being one of the most fundamental approaches. Statistical techniques identify anomalies by examining how data deviates from expected patterns. One common method involves modeling IoT data using probability distributions such as Gaussian or Poisson distributions. When real-world data significantly differs from the expected pattern—such as showing unusually high, low, or inconsistent values—these deviations are marked as potential anomalies. This helps detect abnormal activities that may indicate faults or security intrusions. Time-series analysis is another widely applied statistical approach, especially because IoT data typically has a temporal nature. Models such as ARIMA or exponential smoothing are used to study and forecast normal behavior based on historical data. Any extreme difference between predicted values and collected data points may signal abnormal behavior, enabling early detection of faults or unusual system activity. In general, an anomaly is any observation that differs significantly from the normal behavior of a system. These unusual points can appear in different forms: a single abnormal reading, a context-specific deviation, or a pattern affecting an entire group of data points. Anomalies often occur due to external influences, device failures, environmental changes, or malicious attacks. Therefore, the main goal of an anomaly detection algorithm is to identify these abnormal occurrences accurately and, when possible, determine their underlying causes. Selecting the appropriate model is essential, especially in binary classification scenarios where data must be labeled as either normal or abnormal. Because IoT systems vary widely in their purpose, environment, and data characteristics, customized detection approaches are often required for different applications. This ensures that the algorithm can effectively capture the unique behavior of the system being monitored. IoT anomaly detection strategies can be grouped into four general categories based on their problem-solving approach, application domain, methodological type, and response time. These classifications provide a structured understanding of how different detection algorithms are designed and implemented. Anomalies themselves are commonly divided into three main types: point anomalies, contextual anomalies, and collective anomalies.

- **Point anomalies** occur when a single data point significantly differs from expected behavior. For example, a sudden spike in a sensor reading may signal a malfunction.
- **Contextual anomalies** depend on the situation or environment. A data point may be normal in one context but abnormal in another, such as

temperature readings that vary based on location or time of day.

- **Collective anomalies** involve a group of data points that appear unusual when viewed together. For example, analyzing an entire electrocardiogram waveform can reveal irregularities that are not visible in individual readings.

Understanding these anomaly types and detection strategies is crucial for developing reliable IoT systems capable of responding to risks and maintaining trustworthy operations. Anomaly detection in IoT focuses on identifying unusual or unexpected patterns in the data generated by sensors and connected devices. These irregularities can indicate faults, cyberattacks, sensor failures, or abnormal system behavior. IoT anomaly detection strategies can be grouped based on application, computational approach, and the type of algorithms employed.

A. Anomaly Categorization by Application

IoT anomaly detection applications generally fall into three categories: **constructive**, **destructive**, and **data-cleaning** applications.

Constructive Applications Constructive applications provide positive or beneficial outcomes. They enhance safety, efficiency, and productivity across different domains. Examples include monitoring daily activities of elderly individuals to detect falls, optimizing UAV operations in smart agriculture, or improving smart home performance through intelligent learning models. These applications typically rely on machine learning models such as multilayer perceptron networks, k-nearest neighbors, support vector machines, and reinforcement learning.

Destructive Applications Destructive applications involve behaviors intended to disrupt normal operations or compromise IoT systems. These may include cyberattacks targeting IoT networks, unauthorized data manipulation, or disruption of business-critical processes. Detection strategies for destructive activities often rely on early prevention and post-event identification. Systems designed for this purpose typically use auto encoders, intrusion detection models, and defensive machine learning techniques that can recognize malicious patterns.

Data-Cleaning Applications Data cleaning focuses on improving data quality by removing noise, spikes, or corrupted values from sensor streams. Techniques such as deep convolutional neural networks and other filtering models help eliminate irregularities before the data is used for analysis or decision-making. Clean, reliable data is essential for accurate IoT system performance, prediction, and control.

B. Online and Offline Detection Approaches

Anomaly detection algorithms also differ based on how and when they process data.

Online Detection Online algorithms analyze data in real time, processing one point at a time or within a sliding window. They do not require access to the entire dataset. These approaches are ideal for continuous monitoring systems, such as those used in smart homes, healthcare devices, and network intrusion detection. Online methods often rely on lightweight statistical models, fuzzy clustering, or ensemble techniques.

Offline Detection Offline algorithms analyze the complete dataset after it has been collected. They tend to use more complex and computationally intensive techniques, such as deep learning models, probabilistic modeling, and large-scale clustering. While they cannot provide instantaneous results, they offer higher accuracy and deeper insights. Recent advances have blurred the boundary between both approaches, enabling models trained offline to be deployed for real-time online detection.

C. Classification by Method Type

Detection techniques are commonly divided into **geometrical**, **statistical**, and **machine learning** approaches.

Geometrical Approaches Geometrical methods are based on the idea that normal and abnormal data points occupy different regions in a feature space. Distance-based and density-based approaches identify anomalies by measuring how far a data point lies from dense regions or clusters. Points located in sparse regions are flagged as anomalous. Thresholds are used to determine whether deviations are significant. These models are useful for sensor networks and systems where spatial patterns are important.

Statistical Approaches Statistical methods aim to model normal system behavior through probability distributions, mathematical functions, or predictive models. Techniques such as minimal volume estimators or exponential smoothing forecast expected values. Any significant deviation from the predicted pattern is treated as an anomaly. These methods work well for systems with predictable behavior or stable historical patterns.

Machine Learning and Deep Learning Approaches Machine learning and deep learning techniques have become increasingly popular due to their ability to learn complex, non-linear patterns from large datasets.

- **Machine learning methods** (such as SVM, k-NN, random forests, and clustering models) identify anomalies by learning from labeled or unlabeled data. Supervised learning is effective when labeled examples exist, while unsupervised approaches such as clustering are useful when anomalies are unknown in advance.
- **Deep learning methods**, including LSTM networks, transformer models, autoencoders, and CNNs, are highly effective for analyzing high-dimensional, sequential, or unstructured data. LSTM and GRU models detect anomalies in time-series and streaming data, while CNNs excel with image-based or spatial sensor data. These models automatically learn features from raw data and can adapt to new patterns over time.

Modern IoT systems often combine multiple techniques to achieve higher accuracy, robustness, and adaptability.

D. Machine Learning for IoT Anomaly Detection

Machine learning plays a central role in anomaly detection due to its ability to analyze large, heterogeneous datasets. These algorithms learn hidden patterns from historical data and classify new observations as normal or abnormal. Machine learning is particularly valuable in IoT because:

- It can manage large volumes and varieties of data, including sensor readings, logs, and network traffic.
- It can detect previously unseen anomalies that traditional rule-based methods may miss.
- It adapts to changing environments by continuously updating models.
- It reduces the need for manual inspection, enabling automated real-time monitoring.

Supervised models work well when labeled data exists, while unsupervised models such as clustering and outlier detection are used when labels are unavailable.

E. Deep Learning for IoT Anomaly Detection

Deep learning leverages neural networks to detect anomalies by modeling complex patterns in IoT data. Its strengths include:

- Ability to process high-dimensional and unstructured data such as images, audio, and time series.
- Automatic feature extraction without the need for manual engineering.
- Support for transfer learning, enabling pretrained models to be adapted to new tasks.
- Real-time detection capability through sequential models like LSTM and GRU networks.

Deep learning improves the reliability, security, and performance of IoT systems by detecting hidden patterns and enabling quick intervention when anomalies occur.

F. Comparative Analysis of Techniques

Different machine learning and deep learning algorithms offer unique strengths:

- **Support Vector Machines (SVM):** High accuracy and well-suited for high-dimensional data; however, they require careful selection of kernel functions.
- **Random Forests:** Robust and accurate with good feature ranking capabilities; less interpretable than simple decision trees.
- **K-Nearest Neighbors (k-NN):** Simple and intuitive, effective for local anomaly detection but sensitive to distance metrics and choice of k.
- **Recurrent Neural Networks (RNN):** Good for sequential patterns but prone to training challenges such as vanishing gradients.
- **LSTM Networks:** Handle long-term dependencies well, ideal for time-series data; require longer training times.
- **Convolutional Neural Networks (CNN):** Effective for image and sensor data, capable of capturing spatial dependencies; typically require large labeled datasets and high computational power.

Overall, the choice of technique depends on the data characteristics, system requirements, computational constraints, and the need for real-time or offline processing.

III. CHALLENGES IN DATA ANOMALY DETECTION FOR IoT

Detecting anomalies in IoT data is difficult because IoT systems generate complex, large-scale, and continuously changing data. Several major challenges make anomaly detection in IoT environments more demanding:

High Dimensionality IoT devices collect data from many sensors and sources, resulting in high-dimensional datasets. Anomaly detection algorithms must process numerous features and understand complex relationships between them. This increases computational cost and difficulty. Dimensionality reduction or feature selection techniques are often needed to manage this issue.

Scalability IoT systems produce massive streams of data every second. Effective anomaly detection models must scale to handle large volumes and fast data arrival rates. Achieving real-time processing requires efficient algorithms and powerful computing infrastructure capable of managing both storage and computation demands.

Imbalanced Data In most IoT datasets, normal behavior is much more common than anomalies. This imbalance can cause models to be biased toward normal patterns, leading to poor detection of rare abnormal events. Techniques such as resampling, anomaly-focused learning, or cost-sensitive methods are needed to improve performance on minority anomaly classes.

Concept Drift IoT environments are highly dynamic, and data patterns can change over time. When the underlying distribution shifts, models trained on old data may no longer perform well. Continuous learning, updating models, and adapting to new patterns are essential to maintain accurate anomaly detection in evolving environments.

Lack of Labeled Data Supervised anomaly detection requires labeled examples of normal and abnormal behavior. However, anomalies in IoT systems are rare, unpredictable, and often not labeled. Collecting and labeling such data is labor-intensive and sometimes impractical. As a result, unsupervised or semi-supervised methods are often required.

Privacy and Security Concerns IoT data frequently contains sensitive personal or operational information. Anomaly detection algorithms must ensure privacy and avoid exposing confidential data during processing. Secure computation, federated learning, and privacy-preserving models are necessary to balance detection accuracy with data protection.

Real-Time Detection Requirements Many IoT applications—such as health monitoring, industrial automation, and security systems—require immediate detection of abnormal behavior. Delivering real-time results is challenging due to high data velocity and the computational cost of advanced algorithms. Efficient models and optimized system architectures are needed to support rapid analysis.

Interpretability In many scenarios, understanding why a particular event is labeled as an anomaly is crucial for decision-making and system maintenance. However, advanced machine learning and deep learning models often act as “black boxes,” offering limited transparency. Achieving a balance between accuracy and interpretability remains an important challenge, especially in safety-critical applications.

IV. DISCUSSION

IoT data anomaly detection plays a crucial role across many industries, helping organizations improve operational efficiency, strengthen security, and support timely decision-making. Real-world case studies show how different sectors use anomaly detection techniques to monitor systems, identify unusual patterns, and prevent failures before they occur. In manufacturing, IoT sensors are embedded in machines to measure temperature, vibration, speed, and energy consumption. When anomaly detection algorithms analyze this data, they can identify unusual behavior that may signal equipment faults or early signs of failure. Detecting these issues in real time allows maintenance teams to take action before a breakdown occurs. This predictive approach reduces downtime, extends equipment lifespan, and lowers maintenance costs. Modern industrial platforms use such techniques to support smart manufacturing and improve production reliability. Smart homes also benefit from IoT-based anomaly detection. Devices such as motion sensors, cameras, smart door locks, and window sensors continuously generate data about household activities. Anomaly detection systems analyze this data to identify unusual behavior, such as unexpected movement or unauthorized entry. When an anomaly is detected, homeowners receive immediate alerts, allowing them to respond quickly to possible security threats. These systems enhance safety and give users greater control and peace of mind, especially when they are away from home.

In healthcare, IoT devices and wearable’s generate continuous streams of patient information, including heart rate, blood pressure, physical activity, and medication adherence. Anomaly detection techniques help identify patterns that deviate from normal health conditions. Early detection of irregularities allows healthcare providers to respond quickly, offer personalized treatment, and prevent complications. Remote monitoring systems equipped with anomaly detection contribute to improved patient outcomes and more efficient healthcare service delivery. Overall, these case studies highlight the broad impact of IoT anomaly detection across different fields. Whether it is predicting machine failures, detecting intrusions in smart homes, or monitoring patient health, anomaly detection enables smarter and safer IoT ecosystems. As IoT continues to expand, its role in ensuring reliability, security, and timely intervention will become even more essential.

V. FUTURE RESEARCH DIRECTIONS

Future research in IoT data anomaly detection will continue to focus on overcoming current limitations and developing more intelligent, efficient, and secure detection methods. Several promising directions are outlined below.

Real-time and edge-based anomaly detection: As IoT networks expand, the demand for real-time analysis grows. Future work will aim to design lightweight and efficient algorithms that can operate directly on edge devices. Processing data closer to the source will reduce latency, improve responsiveness, and support faster detection and mitigation of abnormal events.

Adaptation to dynamic IoT environments: IoT systems constantly evolve as devices, data patterns, and network conditions change. Future research will explore adaptive models that can learn from new data on the fly. Techniques such as online learning, incremental learning, and transfer learning will help anomaly detection systems remain effective even as conditions shift.

Multi-modal anomaly detection: IoT systems collect information from many different data types, such as sensor readings, images, audio, and video. Future studies will focus on multi-modal approaches that combine these diverse data sources. Integrating multiple modalities will help detect complex anomalies that may not be noticeable when each type of data is analyzed separately.

Explainable AI for anomaly detection: As anomaly detection methods become more advanced, understanding their decisions becomes increasingly important. Future research will emphasize explain ability, enabling users to understand why specific anomalies were flagged. This includes generating clear explanations, visual representations, and identifying the most influential features behind each decision.

Privacy-preserving anomaly detection: IoT data often contains sensitive personal information, making privacy protection essential. Future efforts will investigate techniques that allow anomaly detection without exposing private data. Approaches such as federated learning, differential privacy, and secure multi-party computation will help maintain confidentiality while still enabling effective analysis.

Adversarial anomaly detection: With the growing risk of cyberattacks on IoT systems, future research will develop methods to detect anomalies caused by malicious activities. This includes identifying data manipulation, spoofing, and adversarial inputs designed to fool detection models. Robust and resilient algorithms will be crucial to protect IoT infrastructure from targeted attacks.

Advancing research in these areas will enable more accurate, secure, and scalable anomaly detection solutions. These improvements will help IoT systems operate reliably, protect sensitive data, and support efficient decision-making across a wide range of applications.

VI. CONCLUSION

Data anomaly detection is a vital component of the IoT ecosystem, as it helps identify unusual behavior, potential system failures, and security threats. This review has examined current trends, techniques, and challenges in IoT anomaly detection, with a particular focus on machine learning and deep learning approaches. Methods such as ensemble learning, recurrent neural networks, and convolutional neural networks have demonstrated strong capabilities for handling the high dimensionality and complexity of IoT data, resulting in more accurate and reliable detection outcomes. Unsupervised learning and real-time processing have also gained importance, offering the ability to detect anomalies without labeled data and enabling rapid responses to abnormal events. In addition, the integration of multiple data sources and the growing emphasis on explainable AI represent key developments that enhance the transparency, usability, and trustworthiness of anomaly detection systems. Despite these advancements, several challenges remain. Future work must focus on designing efficient real-time and edge-based detection methods, improving model robustness in continuously changing IoT environments, and advancing multi-modal detection techniques. Addressing privacy concerns and developing defenses against adversarial attacks are also essential to ensure secure and dependable IoT operations.

References

[1] P. Kamat and R. Sugandhi, "Anomaly detection for predictive maintenance in industry 4.0-A survey," in E3S web of conferences, 2020, vol. 170: EDP Sciences, p. 02007.

[2] S. K. Bose, B. Kar, M. Roy, P. K. Gopalakrishnan, and A. Basu, "ADEPOS: Anomaly detection based power saving for predictive maintenance using edge computing," in Proceedings of the 24th asia and south pacific design automation conference, 2019, pp. 597-602.

[3] E. Gultekin and M. S. Aktas, "A Business Workflow Architecture for Predictive Maintenance using Real-Time Anomaly Prediction On Streaming IoT Data," in 2022 IEEE International Conference on Big Data (Big Data), 2022: IEEE, pp. 4568-4575.

[4] A. Chehri and G. Jeon, "The industrial internet of things: examining how the IIoT will improve the predictive maintenance," in Innovation in Medicine and Healthcare Systems, and Multimedia: Proceedings of KES-InMed-19 and KES-IIMSS-19 Conferences, 2019: Springer, pp. 517-527.

[5] M. Yamauchi, Y. Ohsita, M. Murata, K. Ueda, and Y. Kato, "Anomaly detection in smart home operation from user behaviors and home conditions," IEEE Transactions on Consumer Electronics, vol. 66, no. 2, pp. 183-192, 2020.

[6] A. Lara, V. Mayor, R. Estepa, A. Estepa, and J. E. Díaz-Verdejo, "Smart home anomaly-based IDS: Architecture proposal and case study," Internet of Things, vol. 22, p. 100773, 2023.

[7] S. Ramapatruni, S. N. Narayanan, S. Mittal, A. Joshi, and K. Joshi, "Anomaly detection models for smart home security," in 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 2019: IEEE, pp. 19-24.

[8] X. Dai, J. Mao, J. Li, Q. Lin, and J. Liu, "HomeGuardian: Detecting Anomaly Events in Smart Home Systems," Wireless Communications and Mobile Computing, vol. 2022, 2022.

[9] S. Hadjixenophontos, A. M. Mandalari, Y. Zhao, and H. Haddadi, "PRISM: Privacy Preserving Internet of Things Security Management," arXiv preprint arXiv:2212.14736, 2022.

[10] C.-R. Su, J. Hajiyev, C. J. Fu, K.-C. Kao, C.-H. Chang, and C.-T. Chang, "A novel framework for a remote patient monitoring (RPM) system with abnormality detection," Health Policy and Technology, vol. 8, no. 2, pp. 157-170, 2019.

[11] M. L. Sahu, M. Atulkar, M. K. Ahirwal, and A. Ahamad, "Cloud-based remote patient monitoring system with abnormality detection and alert notification," Mobile Networks and Applications, vol. 27, no. 5, pp. 1894-1909, 2022.

[12] D. Gupta, M. Gupta, S. Bhatt, and A. S. Tosun, "Detecting anomalous user behavior in remote patient monitoring," in 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), 2021: IEEE, pp. 33-40.