International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 180-184, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE30
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

Outlier Detection for IoT Frameworks using Machine Learning Techniques

[1]V.Lakshmi Chaitanya, [2]G.Anju Sree, [3]D.Aisha Thabusum, [4]U.Sravani, [5]G.Sneha, [6]K.Jyotshna

[1]Assistant Professor, Department of Computer Science & Engineering, Santhiram Engineering College, Nandyal, A.P., India
[2,3,4,5,6]Student, Department of Computer Science & Engineering, Santhiram Engineering College, Nandyal, A.P., India

***Abstract* - In "Outlier Detection for IoT Frameworks Using Isolation Forest," it highlights the need of identifying anomalous data in Internet of Things (IOT) environments, where enormous volumes of sensor data are transmitted over wireless networks. In IoT frameworks, identifying anomalies is crucial for network security, problem solving, and efficient data management. Issues in this field include high data volume and velocity, resource constraints on IoT devices, fluctuating network conditions, and the difficulty of distinguishing between real outliers and assaults or sensor faults. These problems are addressed by a variety of machine learning algorithms, such as K-Means Clustering and DBSCAN for classifying similar patterns and identifying outliers, Isolation Forest and One-Class SVM for unsupervised anomaly detection, and Neural Networks and Auto encoders for deep learning-based anomaly detection in complex, high- dimensional IoT data. These techniques look at network traffic, sensor readings, and device behaviors to improve system security and efficiency. This approach guarantees intelligent, safe, and dependable IoT activities in wireless settings. Among its uses are traffic optimization in smart transportation networks, network anomaly detection in healthcare monitoring systems, fault prediction in industrial IoT, and intrusion detection in smart cities.**

Keywords: Outlier Detection, Machine Learning, Internet, Internet of Things (IOT), Wireless sensor networks (WSN).

I. INTRODUCTION

Any system, including Wireless Sensor Networks (WSN) and the Internet of Things (IoT), relies heavily on sensors to generate data. They bear the responsibility of detecting, managing, and preserving data. A number of sets of decisions depend on this data. Therefore, the dependability of all information is very important in every usage. Limited ressource and ability in the sensors often cause the results they generate to be unreliable and imprecise. These sensors run on batteries, and recharging them greatly increases the chance of getting inaccurate information. The operation of sensor nodes is also greatly affected by environmental conditions. Whether it is IoT or WSN, the main goal is to enable devices to transmit data without human involvement. Malicious attacks—where hostile forces exploit the data—so expose the sensors. The unreliability of sensor data results from all of these factors, which in the end affects the last decision-making process. Outliers could therefore be seen as a significant cause influencing the quality of information. Outliers—such as fraud and intrusion detection, weather monitoring, and sensor faults in heating, ventilating, and air conditioning (HVAC) systems—greatly affect any sector engaged in data collection. Among others traffic irregularity detection Machine learning has recently shown its use as an excellent means of outlier identification in sensor data. This brief research outlines a collection of articles that address the subject of outlier detection using machine learning methods. The structure of the rest of the document is as follows: Section II covers the fundamentals of anomalies in WSN and IoT. Section III discusses the various machine learning techniques used to identify outliers in WSN and IoT, while Section IV identifies important research topics that require more investigation. The paper is concluded in Section V.

II. OUTLIER IDENTIFICATION UTILIZING MACHINE LEARNING IN IOT AND WSN

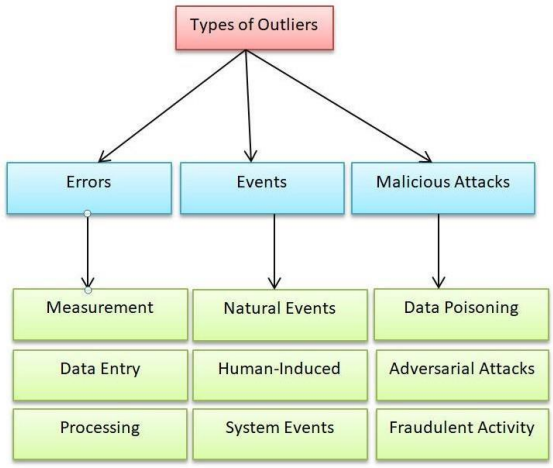
A number of preparatory actions must be conducted before the survey is completed in its entirety. One of the most commonly used definitions of an outlier is as follows:

Definition:

"An observation (or strange observation) small subset of data) that appears to be at conflict with the remainder of this data set".

In the context of WSN and IOT, an anomaly is any data statistic that deviates significantly from the average of the data sensed. Three categories can be distinguished based on the way in which failures, events, and malicious assaults in networks, as well as anomalies related to the Internet of Things, are classified: Clarence W. The techniques for identifying these different defects are described.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 180-184, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE30
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)



2.2 Kinds of Outliers:

The three types of outliers referred to earlier were explored in greater detail:

Errors: An error refers to the data produced by an inaccurate sensor data measurement. Errors cause outliers in data to be significantly more common than those from activities. These inaccuracies influence the quality of data and detection capabilities, thus permitting effective use of the data by negating them. However, Very uncommon anomalies are eliminated in order to save energy in the resource-constrained sensors that are the foundation of all IOT and WSN systems. Defect detection in wireless sensor networks (WSN) and the Internet of Things (IOT) is the focus of the majority of the research that has been discussed thus far.

Events: People perceive it as an act or behavior that alters the essence of the physical environment, including forest fires and air pollution, among other factors. This type of anomaly endures over a significantly longer time frame than errors and usually alters the information patterns. It should be noted, however, that flawed sensors may also create this kind of lasting mistakes, thereby complicating the distinction between the two. Therefore, the current situation emphasizes the crucial significance of spatial correlation, as incorrect sensor data is spatially distinct, while event measurements are spatially interconnected.

Malicious Attacks: People are familiar with a variety of violent assaults everywhere. risks associated with the Internet of Things and wireless sensor networks. In order to deceive other nodes, an attacked node will pose as a valid node. As a result, network security is seriously jeopardized. Because IoT devices frequently use wireless connectivity, they are vulnerable to security breaches during data transmission. One of the main components of intrusion detection is the identification of damaging attacks. In order to find these kinds of irregularities in any sensor network, intrusion detection systems might be developed.

III. LITERATURE REVIEW

The growing dependence on sensor data for decision-making across various applications, identifying outliers within Internet of Things (IoT) systems and Wireless Sensor Networks (WSN) constitutes a significant area of investigation. Serving as the foundation of IoT and WSN systems, sensors are tasked with data collection, processing, and transmission. However, factors such as environmental influences, device malfunctions, and cyber threats could render sensor data unpredictable. Mistakes, occurrences, or intentional disruptions can lead to outliers in sensor data, making their detection crucial for maintaining data integrity and reliability. Applications like weather monitoring, fraud detection, HVAC systems, and traffic anomaly detection may suffer adverse effects due to outliers. In contrast to conventional methods such as threshold-based or rule-based techniques, machine learning (ML) has emerged as a potent means of detecting anomalies in sensor data with enhanced adaptability and precision.

IV. EXISTING METHODOLOGY

Manual Threshold – Based Detection:

- A certain limit is set for the parameters to be measured by the sensors such as temperature, pressure, or humidity levels. Anything above or below that limit is detected as an abnormality.
- For example, if the set limit for temperature is 30 C̊, then any measurement beyond the limit of 30 C̊ is abnormal.
- Drawback: This method lacks adaptability and cannot handle dynamic changes in sensor readings effectively.

Rule Based- System:

- Based on their knowledge of the system, domain experts provide precise rules to spot anomalies. These rules, which record expected actions, are frequently "if-then" formulations.
- If the temperature exceeds 30°C and humidity falls below 20%, mark it as an anomaly.
- Drawback: These systems require continuous updates and struggle to detect complex patterns of anomalies.

Moving Average & Simple Statistical Methods:

- The sensor data's mean, median, and standard deviation are computed. Data points that substantially differ from the surrounding data points are known as outliers.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 180-184, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE30
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

- If the temperature reading is over 2 standard deviations from the moving average, it's considered an outlier.

Time-Series Analysis (Basic Trend Detection):

- Past data is studied to identify patterns, seasonality, and trends. Anomalies occur when a point of data significantly deviates from the expected pattern or trend.
- Example: If the weather typically follows a daily pattern but drops at an unusual time, it is labeled as an anomaly.
- Drawback: This method struggles with unpredictable variations and requires continuous recalibration.

V. PROPOSED METHODOLOGY

Supervised Learning:

In supervised learning, prior knowledge regarding the data classes is available, meaning that both the inputs and their corresponding outputs are provided. Several well-known supervised learning algorithms include:

1) Support Vector Machine (SVM): A classification technique called the Support Vector Machine (SVM) uses the idea of a hyperplane to differentiate between two classes. SVM aims to maximize the margin and divide the classes with the fewest possible mistakes in order to identify the hyperplane. One well-liked supervised learning technique for identifying inaccuracies in sensor readings is SVM. The decision to identify the malfunctioning sensor is made by a cluster head.

2) k-nearest Neighbor (k-NN): The process of predicting a new data point using k-nearest Neighbor (k-NN) involves examining the k nearest neighbors using several distance metrics, the most popular of which is the Euclidean distance. For real-time anomaly detection in WSNs. The authors of this work have suggested a novel approach to the k-NN lazy learning problem that is based on hypergrid intuition.

3) Neural Networks: Neural networks mimic the functions of the human brain in an effort to identify the underlying relationships in a set of data. The method is used to address complicated issues and consists of multiple levels. The use of neural networks in resource-constrained IoT and WSN is hindered by the complexity of the computations, despite the fact that various research use neural networks to identify outliers. To find anomalies in sensors an advanced neural network technique known as auto encoder neural networks. To demonstrate the viability of their proposed approach, they also put it into practice on a test bed.

4) Bayesian Learning Method: The probability distribution is used as the learning parameter in the Bayesian learning approach. Therefore, in order to apply the Bayesian learning approaches, prior knowledge about the data set is required. This new intrusion detection technique is based on a slightly modified version of the Bayesian likelihood test. When compared to other state-of-the-art techniques now in use, the performance demonstrates that the suggested intrusion detection method performs better.

Unsupervised Learning:

Giving input data without any corresponding output variables is known as unsupervised learning. It seeks to comprehend the data's distribution in order to improve understanding. Learning models arrange the inputs according to their attributes. Below is a list of some of the most popular unsupervised learning techniques:

1) k-means Clustering: Finding the k-means number of clusters in the data is the goal of this method. By continually allocating each input data point to one of the k clusters according to their inherent feature similarities, the algorithm operates. To guarantee network security, a genetic k-means algorithm for intrusion detection.

2) Principal Component Analysis (PCA): PCA, or the computation of the projection of the original data set onto a new dimension with fewer components, is one method of breaking up large data sets into smaller ones. PCA is one of the most widely used methods for identifying irregularities in sensor data. Recursive PCA, a variation of simple PCA, was used in this research by the authors to merge duplicate process data and identify outliers in sensor data.

By building a classification model, classification-based techniques typically produce an extremely precise list of outliers. However, selecting the appropriate kernel function and the processing cost of SVM are two significant disadvantages. Furthermore, it is quite difficult to create an accurate Bayesian model when there are a lot of variables

VI. RESULT AND DISCUSSION

Machine learning techniques often aim for high accuracy when identifying sensor outliers. High detection rates and low false-positive rates are essential for all outlier detection. The percentage of outliers that are taken into account and correctly identified as such is implied by the detection rate. The fraction of normal data that were mistakenly labeled as outliers is known as the false positive rate. The tradeoff between these two elements is ascertained using the Receiver Operating Characteristic (ROC) curve. A machine learning algorithm performs better if the area under the curve is greater. There are still several areas of outlier detection that require more research, even though there have been numerous noteworthy

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 180-184, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE30
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

studies on the topic in WSN and IOT. These consist of the following:

- Using ensemble methods to identify outliers in IOT and WSN utilizing semantics and machine learning techniques to detect faults and occurrences.
- Combining online and offline learning strategies to improve outlier identification.

VII. CONCLUSION

In this study, a range of machine learning techniques for detecting outliers in IoT and WSN have been examined. Sensors are essential for generating raw data and for detecting changes in the environment in IoT and WSN frameworks. Detecting outliers is essential for accurately interpreting the sensor-generated, error-free data. Furthermore, a collection of studies that aid in identifying different types of outliers in sensor readings has been assembled. It can be inferred from the discussion that classification algorithms are the most commonly utilized learning techniques for outlier detection in IoT and WSN. Given the current limitations in both IoT and WSN, more effective outlier detection techniques for univariate and multivariate data need to be developed. When creating new machine learning approaches, node mobility and changes in network topology must also be carefully considered.

REFERENCES

[1] V. Reppa, P. Papadopoulos, M. M. Polycarpou, and C. G. Panayiotou, "A distributed architecture for hvac sensor fault detection and isolation," IEEE Transactions on Control Systems Technology, vol. 23, no. 4, pp. 1323–1337, July 2015.

[2] M. Alotaibi, "Security to wireless sensor networks against malicious attacks using hamming residue method," EURASIP Journal on Wireless Communications and Networking, vol. 2019, no. 1, p. 8, Jan 2019.

[3] Sandhya G and A. Julian, "Intrusion detection in wireless sensor network using genetic k-means algorithm," in 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, May 2014, pp. 1791–1794.

[4] M. Xie, J. Hu, S. Han, and H. Chen, "Scalable hypergrid k-nnbased online anomaly detection in wireless sensor networks," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 8, pp. 1661–1670, Aug 2013.

[5] Mahammad, Farooq Sunar, et al. "Key distribution scheme for preventing key reinstallation attack in wireless networks." AIP Conference Proceedings. Vol. 3028. No. 1. AIP Publishing, 2024.

[6] Suman, Jami Venkata, et al. "Leveraging natural language processing in conversational AI agents to improve healthcare security." Conversational Artificial Intelligence (2024): 699-711.

[7] Sunar, Mahammad Farooq, and V. Madhu Viswanatham. "A fast approach to encrypt and decrypt of video streams for secure channel transmission." World Review of Science, Technology and Sustainable Development 14.1 (2018): 11-28.

[8] Mahammad, Farooq Sunar, Karthik Balasubramanian, and T. Sudhakar Babu. "A comprehensive research on video imaging techniques." All Open Access, Bronze (2019).

[9] Mahammad, Farooq Sunar, and V. Madhu Viswanatham. "Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach." The Journal of Supercomputing 76.4 (2020): 2275-2288.

[10] Devi, M. Sharmila, et al. "Extracting and Analyzing Features in Natural Language Processing for Deep Learning with English Language." Journal of Research Publication and Reviews 4.4 (2023): 497-502.

[11] Devi, M. Sharmila, et al. "Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection." Journal of Algebraic Statistics 13.3 (2022): 112-117.

[12] Mandalapu, Sharmila Devi, et al. "Rainfall prediction using machine learning." AIP Conference Proceedings. Vol. 3028. No. 1. AIP Publishing, 2024.

[13] Chaitanya, V. Lakshmi, et al. "Identification of traffic sign boards and voice assistance system for driving." AIP Conference Proceedings. Vol. 3028. No. 1. AIP Publishing, 2024.

[14] Chaitanya, V. Lakshmi. "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System." journal of algebraic statistics 13.2 (2022): 2477-2483.

[15] Chaitanya, V. Lakshmi, and G. Vijaya Bhaskar. "Apriori vs Genetic algorithms for Identifying Frequent Item Sets." International journal of Innovative Research & Development 3.6 (2014): 249-254.

[16] Parumanchala Bhaskar, et al. "Incorporating Deep Learning Techniques to Estimate the Damage of Cars During the Accidents" AIP Conference Proceedings. Vol. 3028. No. 1. AIP Publishing, 2024.

[17] Parumanchala Bhaskar, et al "Cloud Computing Network in Remote Sensing-Based Climate Detection Using Machine Learning Algorithms" remote sensing in earth systems sciences (springer).

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 180-184, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE30
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

[18] Parumanchala Bhaskar, et al. "Machine Learning Based Predictive Model for Closed Loop Air Filtering System." Journal of Algebraic Statistics 13.3 (2022): 416-423.

[19] Paradesi Subba Rao, "Detecting malicious Twitter bots using machine learning" AIP Conf. Proc. 3028, 020073 (2024), https://doi.org/10.1063/5.0212693.

[20] Paradesi Subba Rao, "Morphed Image Detection using Structural Similarity Index Measure"M6 Volume 48 Issue 4 (December 2024), https://powertechjournal.com

Citation of this Article:

V.Lakshmi Chaitanya, G.Anju Sree, D.Aisha Thabusum, U.Sravani, G.Sneha, & K.Jyotshna. (2025). Outlier Detection for IoT Frameworks using Machine Learning Techniques. In proceeding of International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25), published by *IRJIET*, Volume 9, Special Issue of INSPIRE'25, pp 180-184. Article DOI https://doi.org/10.47001/IRJIET/2025.INSPIRE30

****

