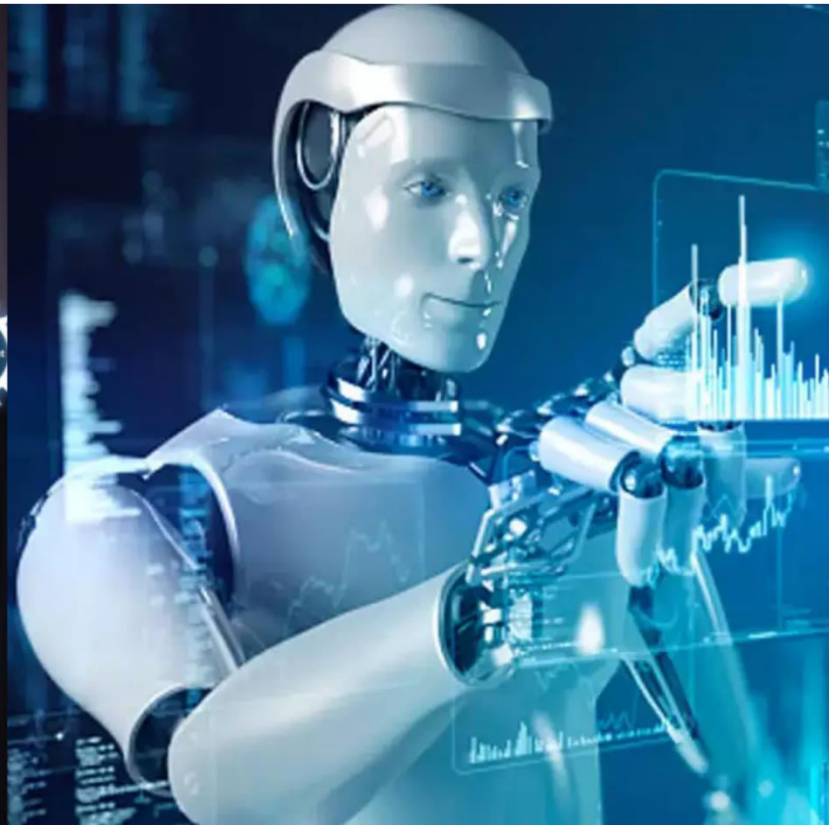
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 12, December 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412106|

Unsupervised Learning Based Network Anomaly Detection in IoT Systems

Mehanthika S, Dr Sathyavathi S, Shobana G

P.G Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

Assistant Professor, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

Assistant Professor, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

ABSTRACT: Due to the increasing number of devices connected to the Internet of Things and the growing prevalence of advanced cyber attacks, network anomaly detection methods are becoming essential. However, supervised techniques are constrained by a lack of labeled data. Thus, in this paper, an unsupervised learning- based IoT network traffic anomaly detection system is developed utilizing One-Class Support Vector Machine (OC- SVM), Isolation Forest and Local Outlier Factor (LOF) algorithms which recognize the behavior of normal traffic and identify the deviation as anomalies. In preprocessing network traffic data, feature scaling and encoding can improve model performance. The experiment found that for detecting network anomalies, Isolation Forest outperformed LOF in terms of detection efficiency, while LOF is considered to outperform in detecting local anomalies. This comparison indicates the applicability of each model in the context of Io T security applications, and the proposed approach can be scaled for an intrusion detection use case without the requirement of labeled data, suitable for a real- time intrusion detection system.

KEYWORDS: Network Anomaly Detection, IoT Security, Unsupervised Learning,One-Class SVM, Isolation Forest, Local Outlier Factor

I. INTRODUCTION

The Internet of Things (IoT) could be defined as a major approach to the current state of technology related to the Internet. IoT helps and enables the communication of IoT devices for different purposes like managing smart cities, healthcare services, industry management, and home services. Even though the starting phase of IoT technology introduces efficiency as well as automation to different processes, the current state of IoT technology reveals some severe security threats because of the degree of decentralization and limited resource availability.

The IoT network is found to be a highly susceptible targeted network to different attacks such as Distributed Denial of Service (DDoS) and any form of unauthorized access or modification.The traditional way of network intrusion detection techniques either employs the concept of signature-based detection techniques or supervised learning techniques. In the signature- based detection technique, the attack could be ruled out on the basis of a pattern. However, this might go wrong in a situation involving a new pattern of attacks. In the supervised learning techniques, a huge amount of data needs to be gathered. This, in fact, poses a problem in the context of the evolving IoT network, where the data continues to keep changing. It can be seen that due to the limitations in the respective methods of anomaly detection, there was interest in unsupervised learning methods for anomaly detection in filling the gap.

Unsupervised learning methodology includes the benefit that no labeling is required. The normal flow in the network could be directly identified with the aid of flow patterns. The pattern that varies from the recognized knowledge is recognized as an anomaly; therefore, the method could be applied appropriately for the detection of unknown anomalies as well. The fact shows that unsupervised learning methodology is very relevant in the context of the IoT due to the changing patterns and the absence of the required labeling.

Among the various methodologies related to unsupervised learning, the algorithms that have proven their effectiveness in the domain of identifying the anomalies in the respective network are One-Class Support Vector Machine (OC-SVM), Isolation Forest, and Local Outlier Factor (LOF). In the OC- SVM method, the boundary can be formed using

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412106|

the normal points of data, in Isolation Forest method, the isolation of the data can be performed using the concept of the random partitioning method, whereas in LOF method, the difference in density can be calculated. Each of the methodologies mentioned works efficiently for respective types of anomalies.

The paper proposes a new approach in the domain of the unsupervised learning methodology for the detection of anomalies in the IoT. The new approach uses the domains related to the One-Class Support Vector Machine, Isolation Forest, and LOF. The proposed approach intends to provide a new approach within the domain related to identifying the preprocessing steps in the respective IoT network, and new models related to the unsupervised approach are developed. Additionally, the respective comparison methodology among the various approaches takes place with the intention of identifying the new approach's efficacy.

II. RELATED WORK

Anomaly detection in a network has been one of the successful means of enhancing security in IoT and networked systems. conventional intrusion detection systems mainly makes use of signature detection as a technique. The technique lacks effective detection of familiar attacks; on the other hand, zero-day attacks cannot be detected using such approaches. In view of these issues, scientists have shifted their focus to developing intrusion detection based on ML algorithms, namely unsupervised algorithms.

Use of One-Class Support Vector Machine has been proved in different researches for anomaly detection of network traffic. The learning process of OC-SVM models takes place considering only normal instances, and they formulate the decision boundary based on normal instances. OC-SVM has been applied by Kumar and Singh for detecting anomalous activities on critical infrastructural networks, and they found based on their research that OC-SVM has potential for global anomaly detection, but their proposed technique was sensitive to parameters and was also unable to address the dynamically varying behavior of the network.

Isolation Forest has been greatly acclaimed for its computational efficiency and scalability. Gupta & Batra did a comparative study on some anomaly detection methods in the IoT domain and concluded that Isolation Forest performed better than other methods in terms of the time taken for execution and accuracy of the detection process. The Isolation Forest algorithm follows the structure of trees, in which the anomaly points in the data could be isolated effectively with less consumption of time. Hence, Isolation Forest is very efficient for larger- scale and high-dimensional network traffic. But even Isolation Forest has some limitations – it is unable to isolate the local anomaly in the dense part of the data.

LOF is an anomaly detection method that works on the basis of density between the data point along with its neighborhood points. Alshammari et al. demonstrated the efficacy of LOF in detecting an anomaly point in IoT networks. Even then, LOF gets degraded in handling global anomalies in data points or in scenarios with variable data point density. More recently, some research has also focused on the hybrid approaches that involve deep learning, including the use of autoencoders, ensemble techniques, and so on, which try to further enhance the accuracy of the detection process.Although the above approaches seem rather promising, some of them consume considerable amounts of computational power and are often not suitable for IoT systems due to the complexity of the model parameters and tuning. Although some related works try to propose that unsupervised learning could be effectively used for anomaly detection, few of these works provide comprehensive comparisons of their techniques. This research attempts to address this gap by making a comparative analysis of the OC-SVM, Isolation Forest, and LOF algorithms for network anomaly detection in the context of IoT systems regarding the performance, scalability, and real-time efficiency of the techniques used.

III. PROPOSED SYSTEM

The proposed research proposes the implementation of an unsupervised learning – based model for identifying anomalies in IoT networks. The main purpose of the correct identification of anomalous network activity without the need for supervised learning, which is not always possible and available in actual IoT networks. The main purpose of the proposed model is the correct identification of anomalous network activity without the need for supervised learning, which is not always possible and available in actual IoT environments.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412106|

The proposed system involves a multi-stage process comprising data collection, preprocessing, feature scaling, modeling, anomaly detection, and result analysis. The process begins with the collection of data from the IoT network traffic based on parameters such as packet size, type of packet, duration of communication flow, source and destination address, and traffic statistics. These parameters enable a detailed understanding of the network. During the preprocessing phase, the raw data is processed by eliminating the missing, duplicate, or inconsistent data. Categorical variables are transformed into their corresponding numerical form. While dealing with the noise issues in the data, the RobustScaler function is used, which normalizes the variables by reducing the effect of noise in the data, as the data constitutes the network traffic.

Following preprocessing, three types of anomaly detection algorithms using unsupervised learning are used: One- Class Support Vector Machine (OC-SVM), Isolation Forest, and Local Outlier Factor (LOF). The OC- SVM is trained only on existing normal traffic data to establish a boundary within which any deviations are considered anomalies. The Isolation Forest algorithm relies on isolating points through random splitting in a tree structure, where anomalies are isolated in fewer splits than regular points. The LOF technique is based on comparing the density of an object to those of nearest neighbors to establish anomalies in network traffic.

The trained models are then employed to inspect the network inputs and detect whether a particular input belongs to the normal or anomaly class. The results obtained from the models will then be inspected to select the best approach for IoT network anomaly detection. The proposed framework architecture allows smooth integration with existing network inspection frameworks to facilitate learning from collected network inputs. On the whole, it can be stated that the proposed work presents an effective approach for addressing unknown and emerging threats in an IoT network through the application of unsupervised learning techniques without relying on any labeled data sets.

IV. METHODOLOGY

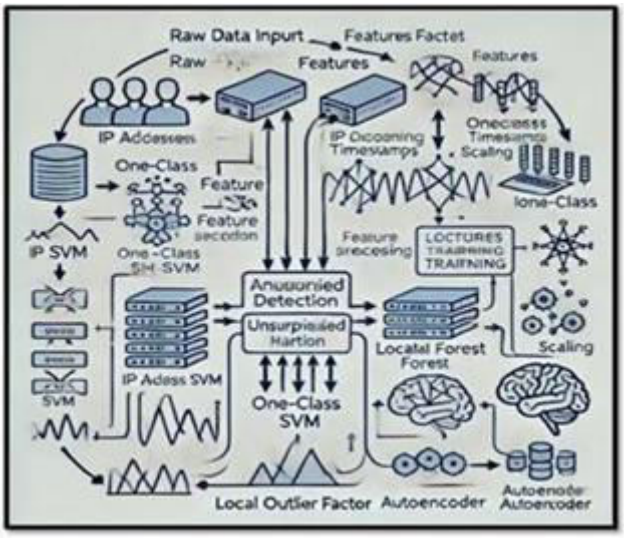


FIG 1.0

The proposed system's approach to the methodology is systematic and structured with respect to the detection of the irregular or anomalous network transactions that do, in fact, exist in the IoT system through unsupervised learning methods. The different stages involved in the complete procedure include data collection, data preprocessing, scaling, training phase, anomaly detection, and evaluation phase. All these stages are involved in ensuring that the anomalies are detected accurately.

A. Data Collection

This process involves gathering the data of IoT network communication traffic through network logs. This dataset has many features, which include size of packets, type of protocol, duration of the flows, sources, and destinations, as well

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412106|

as statistics of flows concerning the traffic. These features represent normal and anomalies that exist while experiencing IoT network communication. This data serves as the input for the process of carrying out anomalies.

B. Data Preprocessing

Data preprocessing is a critical process in the task of removing and enhancing the quality of the input data. In the process of pre-processing, the dataset gets rid of all the null values and the repeating ones. The features having attributes of the categories type, for example, the type of protocols in the dataset, undergo processes of encoding, which helps in the conversion of the attributes into their numeric counterparts. The benefit accrued from the process is that all the attributes in the dataset are machine learning- friendly, and this implies a reduced level of noise in the dataset.

C. Feature Scaling

The data from the networks contains several outliers. The range of value for some variables keeps changing. The data undergoes RobustScaler. The main objective of the RobustScaler is data normalization. The model employed in the RobustScaler is the median and the interquartile range. The approach is highly effective in the presence of outliers. Scaling the feature assists in avoiding the domination of the learning process.

D. Model Training

First, the models designed for the detection of anomalies are trained using unsupervised machine learning techniques on the processed data. There are three models that are applied in this research work:

One-Class Support Vector Machine: It relies on normal patterns in network traffic flow to learn from them in order to establish what kind of boundary should encompass these normal patterns. Any input beyond this boundary is considered an anomaly.

Isolation Forest Algorithm: Isolation Forest is built upon random feature selection, in which isolation is achieved through recursive partitioning. The number of splits for anomalous points is lower; therefore, it resides in a region which is highly computationally efficient.

Local Outlier Factor (LOF): In this case, the algorithm compares the local density of each point with the local density of its neighboring points, enabling the detection of local differences and anomalies. Each model will be trained separately to test the efficiency of the anomaly detection task concerning the IoT network traffic.

E. Anomaly Detection:

This network traffic data will be processed by the trained models. Thus, each set of data will be labelled depending on whether it was normal traffic or an anomaly, based on the trends the model trained on.
Anomaly Score: This shows how abnormal the network traffic activity is, depending on the model output.

F. Performance Appraisal

Precision, recall, and F1-value are some traditional metrics which could be used for measuring the efficiency of the new approach. The metrics are appropriate for the kind of system developed for anomaly detection because imbalance is normally present in these kinds of systems. The value attained from the models shall be used for determining their efficiencies in determining the most efficient unsupervised learning technique for IoT networks.

V. RESULTS AND DISCUSSION

The performance of the proposed framework based on unsupervised learning algorithms for anomaly detection is measured using typical parameters for classification problems, such as precision, recall, and F1 measure. The parameters are very popular among schemes evaluating anomaly detection systems due to the imbalanced nature of traffic, where anomalous. Traffic occupies only a small fraction of network traffic. From the experimental results,it can be concluded that all three unsupervised learning algorithms, namely One-Class Support Vector Machine (OC-SVM), Isolation Forest and Local Outlier Factor (LOF), are able to identify anomalous network activity in **the IoT environment. Their performance depends upon the type of anomalies and the detection method.**

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412106|

The Isolation Forest model yields the best results in terms of overall performance among the tested algorithms. The tree- based model in the Isolation Forest algorithm has a high capacity to isolate points in a dataset by recursively dividing the space of the features. This attribute of the Isolation Forest algorithm yields the best precision, recall, and F1 score, implying an optimal trade-off between the true positive and negative rates. This informs the suitability of the algorithm for a wide range of IoT networks, particularly in large environments that demand instant isolation of anomalies in the IoT network.

One-Class SVM performs steadily with consistency for identifying global outliers. By learning the decision boundary around normal traffic patterns, OC-SVM accurately detects abnormal patterns that fall outside normal patterns. Yet, it proves quite vulnerable to the choice of kernels. This vulnerability could make it less adaptable for frequent changes in network patterns of IoT environments.

Local Outlier Factor is effective in finding local anomalies, where it calculates the density of data points compared to their neighboring points. It is effective in finding anomalies when compared to models that need a general understanding of the data. However, the performance of Local Outlier Factor is relatively low when it comes to the detection of global anomalies because of changes in the density of data. There may also be high processing complexity when handling larger data.On the whole, the experiment outcome has justified the use of unsupervised machine learning algorithms in anomaly detection within the IoT network. From the various algorithms used in this study, it can be concluded that the Isolation Forest algorithm presents the optimal balance between accuracy, efficiency, and scalability. The need for the correct selection of the algorithm based on the size of the network cannot be ignored.

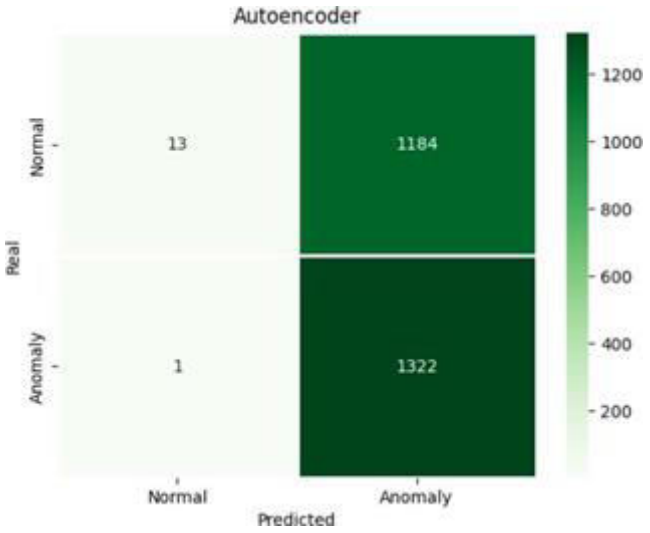


FIG 1.1

VI. CONCLUSION AND FUTURE ENHANCEMENT

This paper has presented a detailed framework for unsupervised learning-based network anomaly detection for Internet of Things (IoT) networks, using One-Class Support Vector Machine, Isolation Forest, and Local Outlier Factor. The proposed system performs well by learning the normal activities of the network using any unlabeled network traffic and then identifying anomalies based on properties of deviations, hence being highly relevant to real-world IoT network setups, where there is no availability of labeled examples.The experimental research clearly depicts that the procedure of Isolation Forest performs better with respect to efficient anomaly isolation and scalability for high- dimensional network traffics, along with overall global anomaly detection by the One-Class SVM procedure and efficient local anomalies of network traffics by Local Outlier Factor.

It has clearly shown that unsupervised learning techniques exhibit efficient scalability for real-time intrusion detection techniques for IoT setups with much less human interference. Hence, for further research work, the proposed system

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 12, December 2025

|DOI: 10.15680/IJIRSET.2025.1412106|

may be further advanced by developing using other models for Deep Learning techniques, further addition of methods for strategy learning techniques, and further addition of online learning techniques for real-time processing. Further, addition of concepts for federated learning techniques for real-time processing for large distributed IoT setups may also be useful for further improving real-time scalability and security for distributed network security.

REFERENCES

[1] C. C. Aggarwal, Outlier Analysis, 2nd ed. Cham, Switzerland: Springer, 2017.
[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," ACM Comput. Surv., vol. 41, no. 3,pp. 1–58, July 2009
[3] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," J. Netw. Comput. Appl., vol. 60, pp. 19–31, Jan. 2016.
[4] . Alshammari, L. Basalelah, and W. Rukbah, "Malware detectionfor IoT using one-class classification," unpublished.
[5] T. Khan and A. Mehmood, "A hybrid unsupervised learning approach for enhanced cybersecurity in IoT," IEEE Internet Things J., in press.
[6] F. Li and X. Zheng, "Federated PCA on Grassmann manifold for IoT anomaly detection," IEEE Trans. Ind. Informat., vol. 20, no. 3, pp. 2150–2159, Mar.2024.
[7] R. Gupta and H. Batra, "A comparative analysis of anomaly detection methods in IoT networks," Future Gener. Comput. Syst., vol. 148, pp. 120–132, Feb. 2024.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details