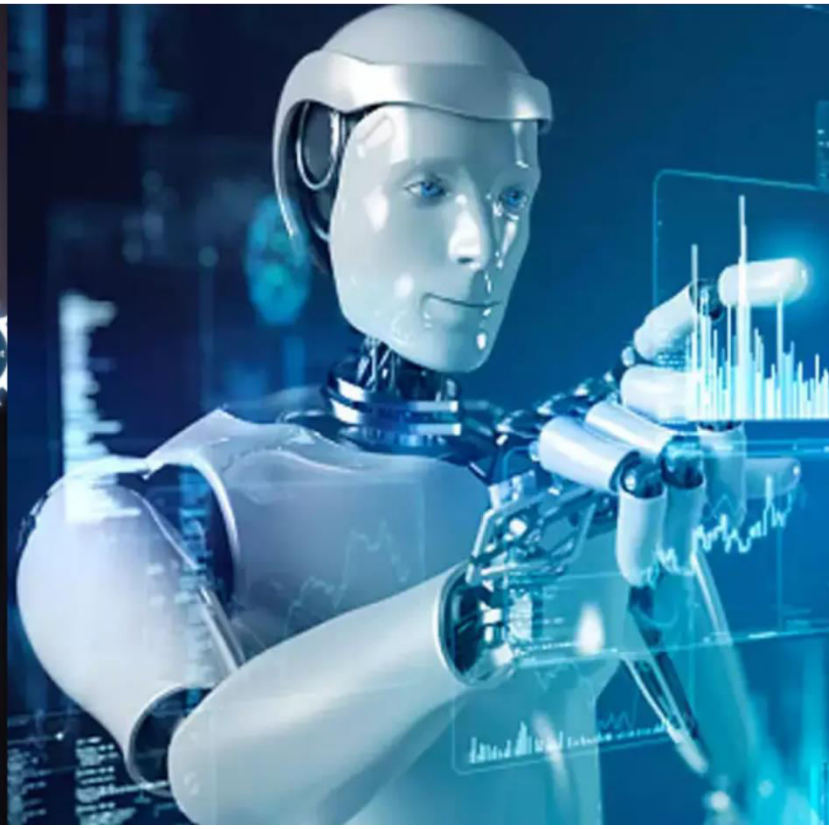
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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 15, Issue 2, February 2026

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 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502024|

Unsupervised Threat Detection using Isolation Forest on Synthetic Data

Shuchitha M N[1], Meghana Jain B M[2], Tulasi H M[3], Vaishnavi K A[4], Dr. Rajath A N[5]

Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India[1-4]

Associate Professor, Department of Computer Science & Engineering, GSSS Institute of Engineering & Technology for Women, Mysuru, Affiliated to VTU, Belagavi, Karnataka, India[5]

ABSTRACT: This paper presents an Unsupervised Threat Detection with Isolation Forest Using Synthetic Data System that uses synthetic network traffic and an Isolation Forest unsupervised learning model to detect anomalous network behavior. The system generates representative synthetic datasets that incorporate normal and anomalous traffic patterns, applies careful preprocessing and feature engineering, and trains an Isolation Forest to isolate anomalies without relying on labelled attack signatures. Experimental evaluation on multiple patterns reports model performance using accuracy, precision, recall, F1-score, support, macro average and weighted average results show the impact of scaling and feature transformations on detection effectiveness. We discuss limitations such as sensitivity to pattern selection and class imbalance, and outline improvements including ensemble strategies and hybrid deep-learning augmentation. The proposed framework aims to provide a scalable, interpretable, and adaptable approach for proactive intrusion detection across diverse network settings.

KEYWORDS: Anomaly Detection; Isolation Forest algorithm; Intrusion Detection System (IDS); Synthetic Network Traffic; zero-day Attack Detection; Machine learning; Network traffic

I. INTRODUCTION

The reliance on networked systems across sectors has increased the attack surface for cyber threats. Conventional signature-based intrusion detection systems (IDS) are limited against zero-day and polymorphic attacks. Machine learning offers an alternative by learning patterns of normal behavior and detecting as anomalies. This paper proposes an Unsupervised Threat Detection with Isolation Forest Using Synthetic Data System that leverages synthetic traffic generation and the Isolation Forest algorithm, an unsupervised method well suited to anomaly detection in high-dimensional spaces. We provide a framework that covers synthetic dataset creation, model development, performance evaluation, and results visualization through GUI.

Most working IDS tools rely on pattern checks, rough guesses, or fixed rules built into firewalls along with SIEM systems. Though they catch familiar dangers well, they demand frequent tweaks while producing lots of false alarms now and then. Besides, real-world data tagged for learning is rare and costly to gather, pushing interest toward self-learning methods plus artificial samples instead.

Conventional setups often depend on signatures, struggle to adjust, besides costing a lot to keep running. This work tackles that issue - spotting anomaly behavior without needing tons of labeled data and learns patterns.

II. LITERATURE SURVEY

Researchers have explored multiple intelligent approaches for intrusion detection using machine learning, deep learning, and anomaly-based techniques. While these methods often achieve high accuracy and automation, many rely heavily on labelled datasets, static rules, and high computational resources. Such dependencies limit their ability to detect zero-day attacks, adapt to dynamic network environments, and scale effectively in large or IoT-based networks. Additionally, issues such as false positives, lack of interpretability, dataset imbalance, poor generalization, and vulnerability to adversarial manipulation remain open challenges.

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 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502024|

Earlier IDS research primarily focused on supervised and deep learning models, which perform well on known attacks but struggle with unseen threats and real-world variability. Recent studies emphasize the need for lightweight, explainable, and unsupervised approaches that can operate on unlabeled or synthetic data. Several works highlight the advantages of anomaly detection, the importance of Explainable AI, and the role of synthetic data in overcoming data scarcity and imbalance.

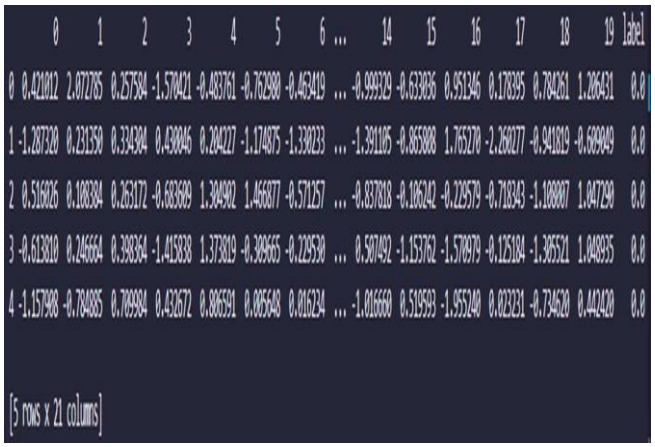
Motivated by these findings, this work adopts an unsupervised Isolation Forest model trained on synthetic network traffic. The approach reduces dependence on labelled datasets, lowers computational overhead, improves adaptability to novel threats, and provides interpretable anomaly scores, making it suitable for real-time and resource-constrained environments.

III. METHODOLOGY

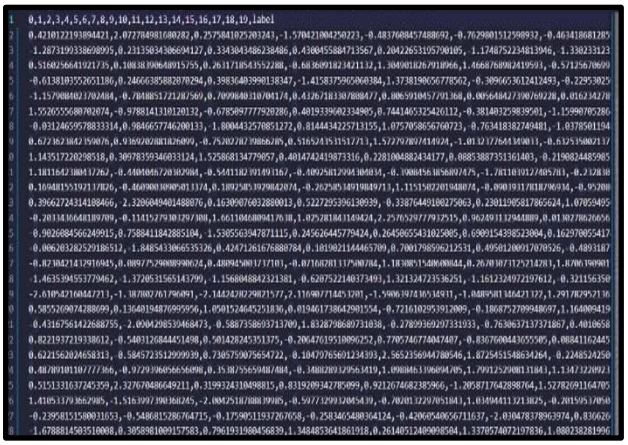
This section explains the overall methodology used in the system, covering synthetic data generation, model training, anomaly detection, and evaluation. Synthetic network traffic data is generated to represent both normal and anomalous network behavior, ensuring realistic traffic patterns that include regular communication as well as abnormal activities. The Isolation Forest algorithm is selected for model training due to its effectiveness in identifying anomalies in complex, high-dimensional data. As an unsupervised learning approach, Isolation Forest does not require labelled data; instead, it learns normal behavior directly from the dataset and detects anomalies as deviations from common patterns. The model isolates outliers by randomly selecting features and split values, allowing unusual data points to be identified efficiently. After training, the model performs anomaly detection by assigning anomaly scores to new data samples based on their deviation from learned normal behavior. A threshold is then defined on these scores to classify traffic as normal or anomalous, indicating potential network threats. The system's performance is evaluated using metrics such as precision, recall, F1-score, support, macro average, and weighted average, providing an assessment of overall training and testing accuracy.

IV. EXPERIMENTAL RESULTS

Figures show the results of Unsupervised Threat Detection with Isolation Forest Using Synthetic Data. Figs. (a) Preview of training datasets used for Isolation Forest Models (b) Features of Synthetic Network Data (c) Classification report of the metrics (d) Graphical representation of precision metric for normal and anomaly traffic (e) Graphical representation of recall metric for normal and anomaly traffic (f) Graphical representation of F1 – score metric for normal and anomaly traffic (g) Graphical representation of support metric for normal and anomaly traffic (h) Graphical representation of Macro Average Metric for normal and anomaly traffic (i) Graphical representation of Weighted Average Metrics for normal and anomaly traffic



(a)



(b)

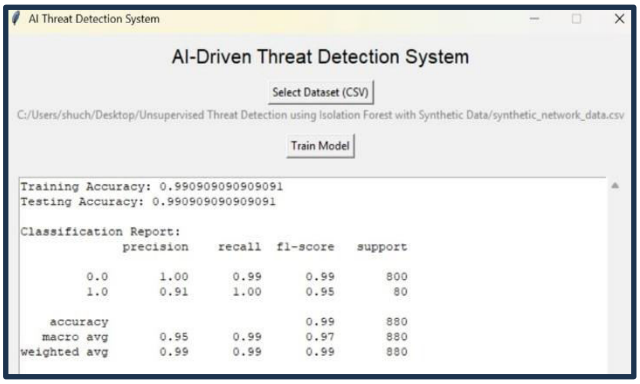
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

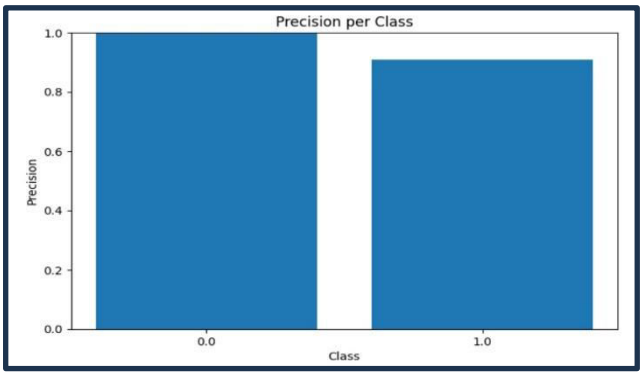
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 2, February 2026

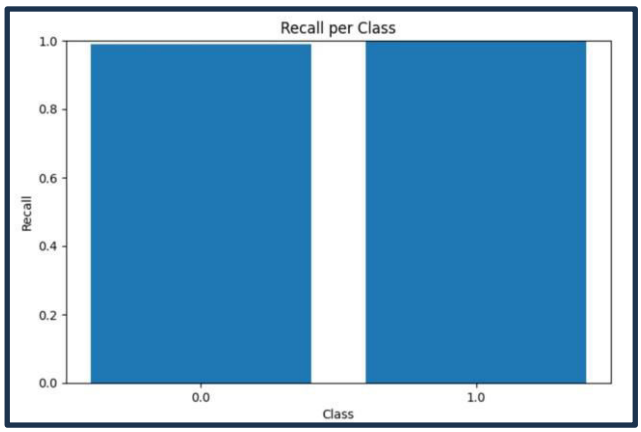
|DOI: 10.15680/IJIRSET.2026.1502024|



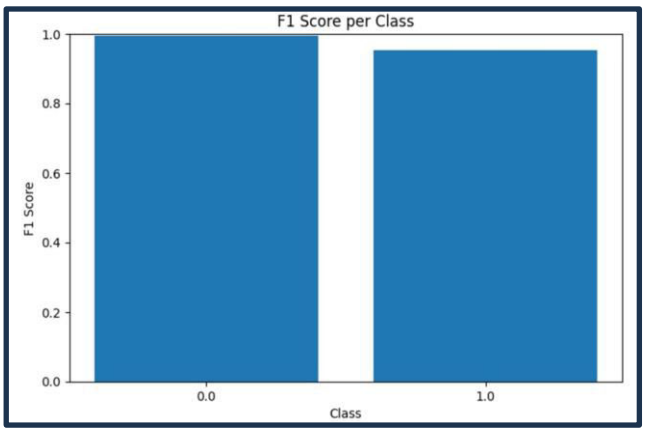
(c)



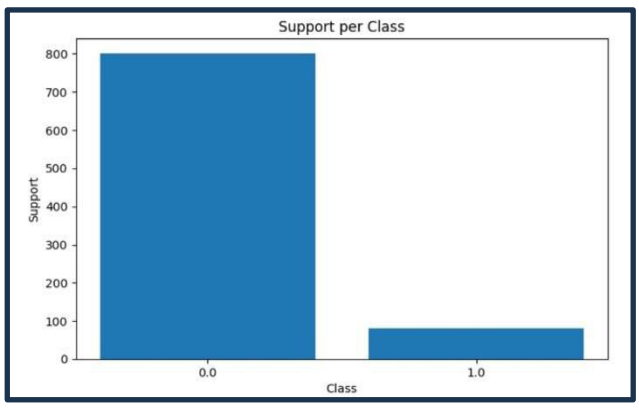
(d)



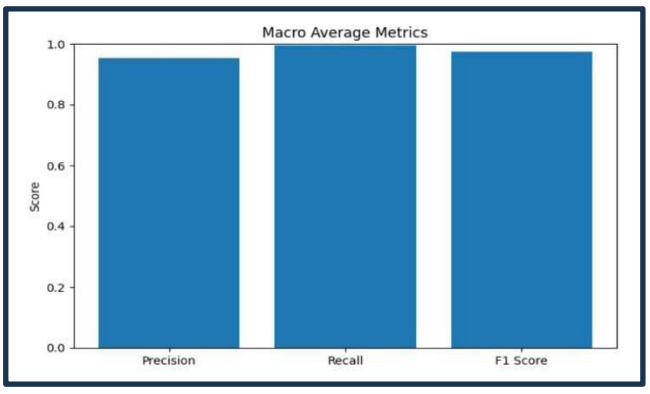
(e)



(f)



(g)



(h)

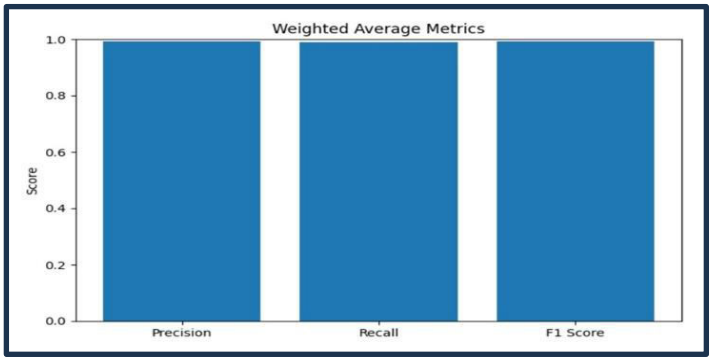
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

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

Volume 15, Issue 2, February 2026

|DOI: 10.15680/IJIRSET.2026.1502024|



(i)

V. CONCLUSION

The project "Unsupervised Threat Detection Using Isolation Forest on Synthetic Data" shows how machines can spot weird network activity - no tagged examples needed. Because actual hacked traffic is tough to get, patterns of network traffic which includes both normal and anomaly behavior generated. The Isolation Forest method spots odd data by separating unusual points from regular patterns. However, tests using accuracy, precision, recall, or F1-score proved the system works fine at detecting anomaly network traffic.

REFERENCES

1. Abraham, Biju, et al., "Intrusion detection system using machine learning algorithms." International Conference Emerging Trends in Information Technology and Engineering (ic-ETITE), pp. 1-6, (2020).
2. Albara Awajan, “A Novel Deep Learning-Based Intrusion Detection System for IoT Networks”, Computers, 12(2):34, E-ISSN 2073-431X, (2023).
3. B. Susilo, A. Muis & R. F. Sari, “Intelligent Intrusion Detection System Against Various Attacks Based on a Hybrid Deep Learning Algorithm”, Sensors, 25(2): 580, ISSN 1424-8220, (2025).
4. Euclides Carlos Pinto Neto, Scott Buffett, “Deep learning for intrusion detection in emerging technologies: a comprehensive survey and new perspectives”, Artificial Intelligence Review, Vol. 58, article number 340, (2025).
5. Hu, Jing, et al., "An intrusion detection model based on ensemble learning algorithm in cloud computing.", Future Generation Computer Systems, vol. 111, pp. 855-864, (2020).
6. Hozouri A., Mirzaei A. & Effatparvar M., “A comprehensive survey on intrusion detection systems with advances in machine learning, deep learning and emerging cybersecurity challenges”, Discover Artificial Intelligence, 5:314 (article number), DOI: 10.1007/s44163-025-00578-1, (2025).
7. Johnson, L., Brown, K., & Lee, S, “Machine Learning Applications in Cybersecurity”. Journal of Information Security, 15(3), 245-261, (2020).
8. Kim, Yeonuk, et al., "Deep learning-based network intrusion detection system for software-defined networking-enabled IoT environments.", IEEE Access, vol. 8, pp. 168195-168207, (2020).
9. Kwangjo Kim, Muhamad Erza Aminanto & Harry Chandra Tanuwidjaja, “Network Intrusion Detection using Deep Learning: A Feature Learning Approach”, Springer, Softcover ISBN: 978-981-13-1443-8; eBook ISBN:978-981-13-1444-5, (2018).
10. Laskov, Pavel, et al., "Practical evasion of a learning-based classifier: A case study." Journal of Machine Learning Research, vol. 9, 2008, pp. 187-205, (2008)

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