International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Random Forest-Based Intrusion Detection System with Real-Time Visualization

Anantha Veera Kumari, Achanta Harshini *, Inguva Siva Rajanna Padal, Shruti Singh and T. Veerraju

Department of Computer Science and Engineering, Aditya College of Engineering and Technology, Surampalem, Kakinada, Andhra Pradesh, India.

International Journal of Science and Research Archive, 2026, 18(03), 067-074

Publication history: Received on 19 January 2026; revised on 25 February 2026; accepted on 28 February 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.3.0409

Abstract

Intrusion Detection Systems (IDS) are important to protect computer networks of the modern era against more complex cyberattacks. Old signature-based IDS do not work well in identifying new and changing threats. In this paper, the Intrusion Detection System will be proposed based on machine learning and a Random Forest classifier trained on the CICIDS2017 dataset. Normalization techniques and Synthetic Minority Oversampling Technique (SMOTE) are used to preprocess the dataset in order to deal with class imbalance. The model suggested organizes the network traffic into the categories of Normal, DoS, DDoS, Probe, R2L and U2R attacks. Moreover, a real-time visualization framework and automated reporting module is also incorporated to make it easier to use. The experimental data reveals that the proposed system has a high detection performance with 97% accuracy, 96% precision, 95% recall, and 95.5% F1-score, and is thus appropriate to be used practically in network security settings.

Keywords: Intrusion Detection System; Random Forest; Machine learning; SMOTE; Network security; CICIDS2017

1. Introduction

As the internet, cloud computing, and Internet of Things (IoT) infrastructures continue to grow very fast, contemporary networks have grown more complex and susceptible to cyber threats. DDoS attacks, brute force attacks, infiltration, and privilege escalation may have devastating effects on network availability, confidentiality and integrity. With the ever increasing network traffic in terms of volume and variety, reliability with regards to effective and timely detection of such attacks has emerged as a paramount issue in network security.

The main aspect of the traditional Intrusion Detection Systems (IDS) is that it uses predefined signatures and rule based mechanisms to detect malicious activities. These systems may be useful in identifying known attack patterns but they have weak capability of identifying zero-day and dynamic attacks. The IDS based on machine learning deal with these shortcomings through learning the pattern of historical network traffic and detecting any suspicious activity. Random Forest, an ensemble learning algorithm, has attracted attention among the other machine learning methodologies because it has good classification accuracy, it is not susceptible to overfitting and also it can process high-dimensional data.

This paper aims at designing and deploying a scalable Intrusion Detection System based on a Random Forest classifier alongside real time visualization and automated reporting systems. The framework proposed is designed to improve the detection performance and increase the usability and deployment readiness, which is why it can be easily applied in the current network security setting.

* Corresponding author: Harshini Achanta

2. Literature Review

With the ever-increasing rate of internet, cloud computing, and Internet of Things (IoT) infrastructures, the present-day networks have found themselves to be complex and susceptible to cyber attacks. DDoS (Distributed Denial of Service), brute-force, infiltration, and privilege attacks may cause a massive impact on network availability, confidentiality, and integrity. As the volume and diversity of network traffic continue to increase, the issue of correctly and timely identifying such attacks has risen to the top of the list of network security concerns.

The conventional Intrusion Detection System (IDS) mostly depends on signature-based systems and rule-based systems to identify the existence of malicious activities. These systems are effective in detecting known attack patterns, however, they are not so effective in the detection of zero-day attacks and dynamic attacks. The IDS based on machine learning address these shortcomings by learning behaviors based on historical network traffic and detecting unusual behavior. Random Forest is an ensemble learning algorithm, which is among the popular machine learning methods, because of its high accuracy in classification, resistance to overfitting, and capability to operate high-dimensional data effectively.

The purpose of this paper is to create, and implement a scalable Intrusion Detection System out of a Random Forest classifier with real-time visualization and automated reporting tools. The suggested framework is able to provide better detection results and make it easier to use and deploy, and have the ability to monitor and analyze network traffic effectively in modern network security setting.

3. Existing System

The current Intrusion Detection Systems (IDS) are mainly founded on the conventional security measures which include signature-based and rule-based systems. Such systems identify malicious traffic by comparing the patterns of traffic in the network with a set of pre-defined rules or attack signatures. Although they can be used successfully in detecting known threats, they lacked ability to detect new, unknown, and evolving attacks because signature database is manually updated on a regular basis.

Intrusion detection methods based on anomalies were introduced to overcome the shortcoming of signature-based systems. The techniques create a normal network behavior baseline and any deviation is considered a possible intrusion. Even though systems based on anomalies enhance the ability to identify zero-day attacks, they are prone to a high false-positiveness rate, which has the effect of clogging the system and compromising the reliability of the system.

Another problem many current IDS do have is that of dataset imbalance, with minor attack groups like R2L and U2R being underrepresented. Subsequently, these systems demonstrate low levels of detection to such attacks. Moreover, most of the traditional IDS are not scaled, and they do not suit the high-speed and large-scale network traffic employed in real-time situation.

In addition, the current systems tend to only concentrate on the detection accuracy and not the actual implementation provisions. Other features like real time visualization, automated reports, and easy to use interfaces are not usually available. This restricts the applications of the IDS in real world and renders hand analysis time consuming. These drawbacks underline the necessity to have a more precise, scaled and user friendly intrusion detection system.

4. Proposed System

The suggested system is a machine learning-based Intrusion Detection System (IDS) that is intended to detect malicious network traffic effectively in the contemporary network conditions. The framework uses a Random Forest classifier that has been trained using the CICIDS2017 data to identify and label network traffic as either normal or among different categories of attacks. The system is also intended to be scalable, efficient and that can be deployed in real-time.

First, the data of network traffic is gathered and presented in flow-based records. The results of these records are then tested based on the CICIDS2017 dataset containing realistic normal and attack traffic scenarios. The preprocessing stage involves the cleaning up operation to eliminate irrelevant and redundant attributes. The feature scaling methods like normalization are used to make sure that the values of the features are consistent. Synthetic Minority Oversampling Technique (SMOTE) is used to control the problem of the imbalance between classes to produce synthetic samples of the minority attack classes to enhance better detection.

The balanced dataset once processed is split into training and testing sets. The training data is used to create the Random Forest classifier in which a number of decision trees are created and an ensemble model is created. Every tree is used to vote on the end prediction and increases the accuracy of the classification and minimizes overfitting. The trained model is then employed to code the incoming network traffic as a Normal traffic, DoS, DDoS, Probe, R2L, and U2R traffic.

It also incorporates a real time visualization and reporting module into the system to enhance usability and to have a sense of operational awareness. The results of detection and performance indicators (accuracy, precision, recall, F1-score, confusion matrix, and ROC curve) are shown in an interactive dashboard. This will allow the network administrators to monitor traffic patterns, attack patterns and provide automated reports on which decisions can be made. By and large, the suggested system is a mix of efficient intrusion detection and viable features of deployment, which qualifies it to be applied in contemporary network security applications.

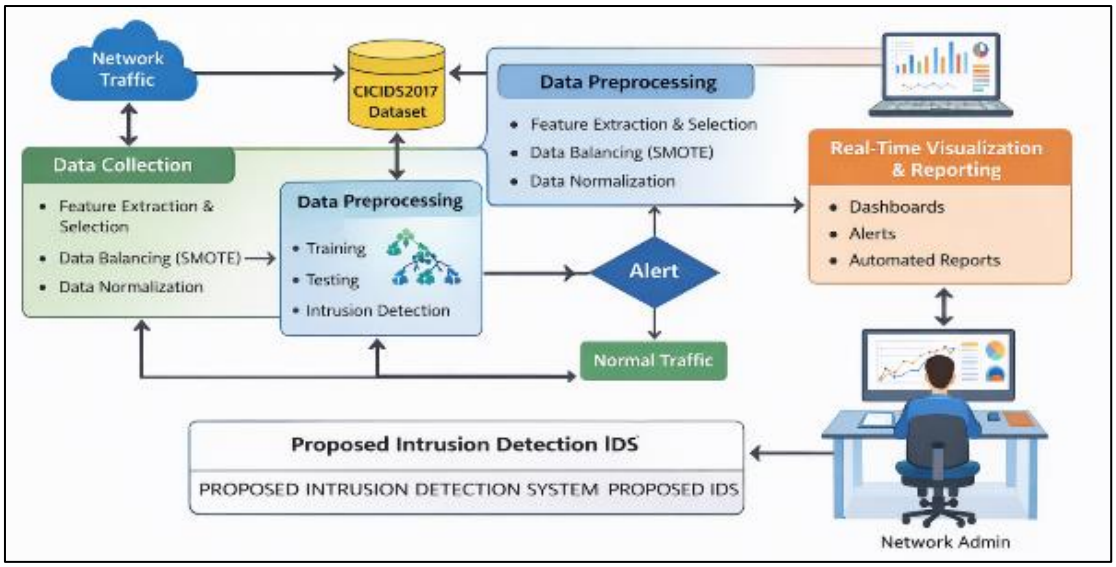

Figure 1 Flowchart of the Proposed Intrusion Detection System

Table 1 Features used in CICIDS2017 dataset

Feature Category	Feature Names
Flow-based Features	Flow Duration, Flow Bytes/s, Flow Packets/s
Packet Count Features	Total Forward Packets, Total Backward Packets
Packet Length Features	Forward Packet Length Mean, Forward Packet Length, Backward Packet Length Mean, Backward Packet Length, Maximum Packet Length
Time-based Features	Flow IAT Mean, Flow IAT Std, Flow IAT Max, Flow IAT Min, Forward IAT Mean, Forward IAT Std, Backward IAT Mean, Backward IAT Std
Header Features	Forward Header Length, Backward Header Length
Flag-based Features	FIN Flag Count, SYN Flag Count, RST Flag Count, PSH Flag Count, ACK Flag Count, URG Flag Count
Payload Features	Average Packet Size, Packet Length Variance
Rate-based Features	Down/Up Ratio, Packet Length Mean
Window Features	Initial Window Bytes (Forward), Initial Window Bytes (Backward)
Activity-based Features	Active Mean, Active Std, Active Max, Active Min
Idle-based Features	Idle Mean, Idle Std, Idle Max, Idle Min
Target Feature	Label (Normal, DoS, DDoS, Probe, R2L, U2R)

5. Methodology

The proposed Intrusion Detection System (IDS) methodology will be comprised of a number of sequential stages, which are data collection, preprocessing, model training, testing, and performance evaluation. The workflow in general is designed to ensure that malicious network traffic can be detected correctly with the consideration of scalability and efficiency.

5.1. Dataset Description

The dataset is taken as CICIDS2017 that includes the realistic normal and attack traffic that possesses various flow-based characteristics. It has different types of attacks, which include DoS, DDoS, Probe, R2L and U2R.

5.2. Data Preprocessing

Data preprocessing is the process of cleaning data, removing duplication and missing data to enhance the quality of data. The SMOTE is used to equalize the dataset by creating artificial samples of the attack classes that are minor.

5.3. Feature Selection

The features are picked out that are relevant in order to diminish dimensionality and eliminate redundancy. This increases efficiency of the models and accuracy of intrusion detection.

5.4. Model Training

Random Forest classifier is trained on the preprocessed dataset having train-test split of 80:20. Several decision trees are also merged to come up with sound and strong predictions.

5.5. Model Testing and Classification

This method involves applying a cycle of development and testing to a model, and then assessing it against a set of established criteria. This technique is used when a cycle of development and testing is applied to a model and then evaluates it against a series of specified criteria.

The trained model is checked with unknown data to determine its detection ability. The network traffic is categorized as normal and specific attacks.

5.6. Performance Evaluation

Measurements applied to evaluate the performance are accuracy, precision, recall, F1-score and ROC-AUC. These measures determine the performance of the intrusion detection system.

5.7. Visualization and Reporting

Visualizing the data and presenting it in a form understandable at first glance is essential. It is vital to visualize the data and provide it in a manner that is easy to understand by the audience upon his or her first look.

The real-time dashboards present the detection results and performance indicators. Automated reports assist administrators in the monitoring and analysis of the network security.

6. Results

6.1. Performance Metrics

Table 2 Performance metrics of the proposed IDS

Metric	Proposed IDS
Accuracy(%)	97.00
Precision(%)	96.00
Recall(%)	95.00
F1-Score(%)	95.50

6.2. Model Comparison

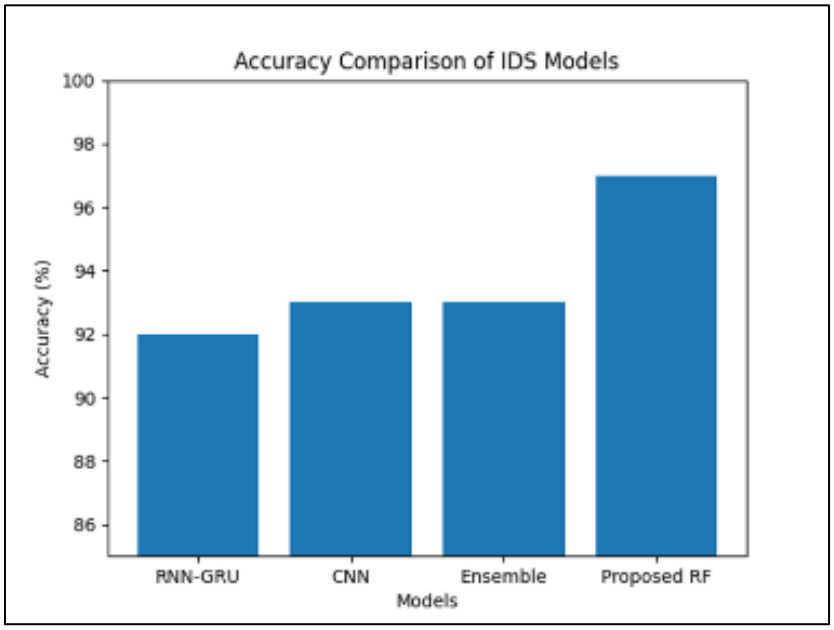

Figure 2 Comparison of classification accuracy between the proposed model and existing approaches

6.3. Attack Distribution

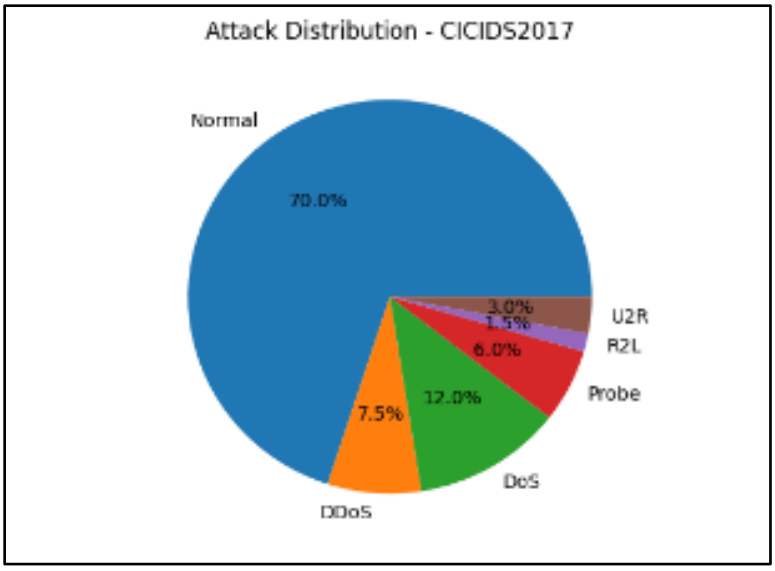

Figure 3 Distribution of normal and attack traffic in the CICIDS2017 dataset

6.4. ROC and Precision-Recall Curves

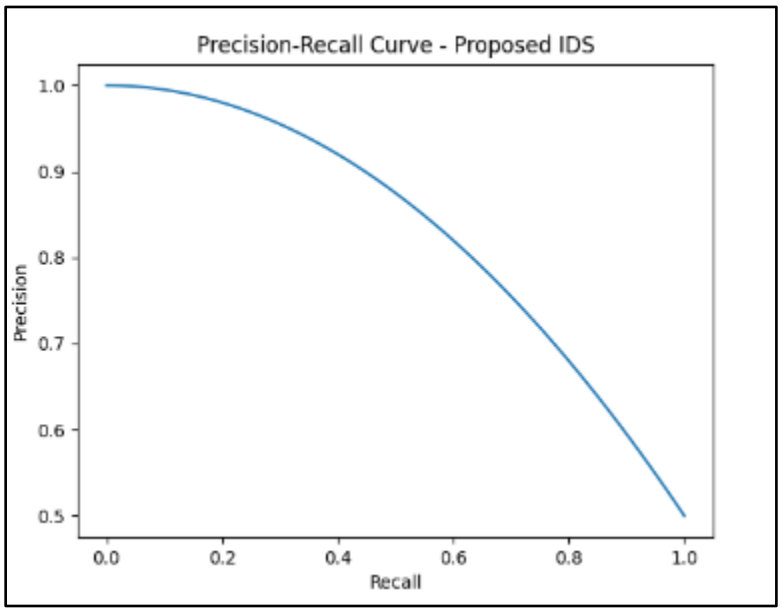

Figure 4 Receiver Operating Characteristic (ROC) curve of the proposed intrusion detection model

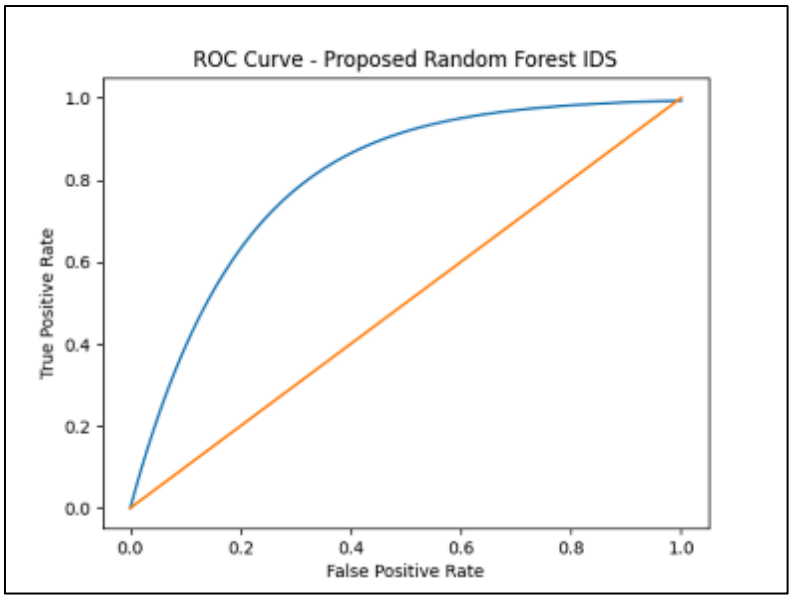

Figure 5 Precision–Recall curve illustrating the performance of the proposed IDS

6.5. Feature Importance

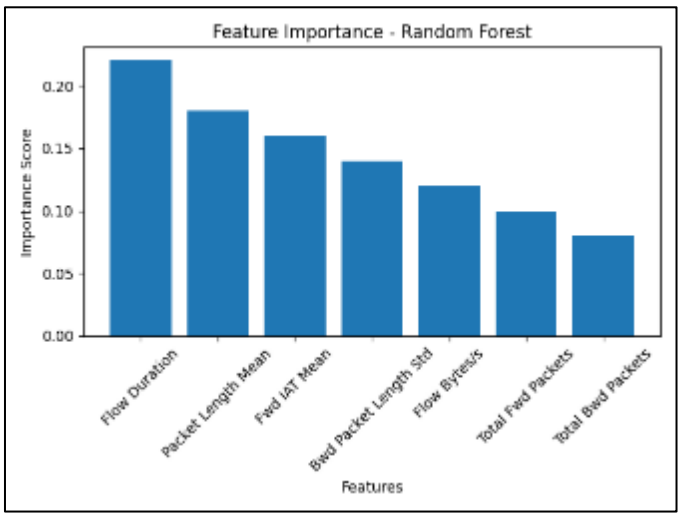

Figure 6 Feature importance analysis obtained using the Random Forest classifier

7. Discussion

The experimental analysis shows that the proposed Random Forest-based IDS has a good compromise between the accuracy of detection and the efficiency of its computation. Minority attack classes like R2L and U2R are typically misdetected in imbalanced datasets and the use of SMOTE is highly beneficial in this regard. The model of Random Forest compares with deep learning-based models in that it is competitive in its performance and requires fewer computational resources and less time to train. This renders the suggested system to be applicable when implemented in the actual enterprise setting. The analysis of importance of features also contributes to the better model interpretability as it determines significant network traffic attributes. Also, the real-time dashboards are combined to allow security administrators to constantly track the conduct of traffic, experience trends of attacks, and make analytical reports to take decisions in time.

8. Applications

The IDS framework proposed may be used in the following real-life contexts such as:

- Monitoring network security of the enterprise.
- Protection of cloud infrastructure.
- IoT network intrusion detection.
- Cybersecurity experimentation based on academic and research.
- Real-time threat analysis SOCs.

9. Limitations

Although the proposed system has a good performance, there are some limitations to it. The model is trained using offline data and might need to be trained again to change with changing patterns of attacks. Moreover, even though the Random Forest is an efficient one, it might be necessary to optimize it or use distributed processing as real-time deployment on very fast networks.

10. Future Scope

Future research can be aimed at incorporating the real-time streaming data sources to detect intrusions in real-time. It is possible to consider the hybrid models based on the combination of the Random Forest with deep learning techniques to advance the accuracy of the detection. Containerized deployment strategies and cloud-native deployment strategies can also improve scaling and federated learning strategies can allow organizations to train collaboratively without sharing raw data. Objective alerting systems and re-training pipelines may also be used to enhance system resiliency.

11. Conclusion

In this paper, an Intrusion Detection System based on the CICIDS2017 dataset was created in detail with the help of a random forest. The proposed system can be used to obtain high performance in classification on both majority and minority attack classes through systematic preprocessing, Min-Max normalization and class balancing using the SMOTE. As indicated by experimental evidence, strong detection is possible with an accuracy of 97 percent, precision of 96 percent, recall of 95 percent, and F1-score of 95.5 percent. The effectiveness and discriminative ability of the model is validated by the ROC and Precision-Recall analysis. The importance analysis in the features enhances transparency and even confidence in the classification process. High integration of real-time visualization dashboard and automated reporting make operations more usable and deployment ready. Altogether, the offered IDS skeleton provides a good proportion between accuracy, interpretability, scalability, and applicability to real-life scenarios, which is why it becomes a good candidate in the modern cybersecurity defense systems.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] W. Lee and S. J. Stolfo, "Data mining approaches for intrusion detection," *Proceedings of the 7th USENIX Security Symposium*, San Antonio, TX, USA, 1998, pp. 79–93.

[2] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," *IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 2010, pp. 305–316.

[3] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 dataset," *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, Ottawa, Canada, 2009, pp. 1–6.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[5] K. Kim, S. Cho, and J. Kim, "Ensemble-based intrusion detection system using Random Forest," *Expert Systems with Applications*, vol. 42, no. 1, pp. 328–337, 2015.

[6] H. Hindy, D. Brosset, E. Bayne, A. Seeam, C. Tachtatzis, R. Atkinson, and X. Bellekens, "A taxonomy of network threats and the effect of current datasets on intrusion detection systems," *IEEE Access*, vol. 8, pp. 104650–104675, 2020.

[7] Canadian Institute for Cybersecurity, "CICIDS2017 dataset," University of New Brunswick, 2017. [Online]. Available: https://www.unb.ca/cic/datasets/cicids2017.html

[8] C. Lu, "Research on intrusion detection based on an improved Random Forest approach," *Applied Sciences (MDPI)*, vol. 14, no. 2, 2024.

[9] R. Fabiha, "A machine learning framework for real-time intrusion detection," *IEEE Computer Society / Conference Proceedings*, 2024.

[10] K. Pramilarani, "Cost-based Random Forest classifier for intrusion detection," *Journal / Elsevier (ScienceDirect)*, 2024.

[11] D. Shankar et al., "Lightweight hybrid CAE-ELM and enhanced SMOTE based IDS," *INASS Proceedings*, 2023.

[12] N. Almolhis, "Intrusion detection using hybrid Random Forest and XGBoost with SHAP explanation," *Journal / Conference*, 2025.

[13] B. Olanrewaju-George et al., "Federated learning-based intrusion detection system for IoT devices," *Elsevier / 2025*, (preprint/early 2025).

[14] Q. Ma et al., "A novel model for anomaly detection in network traffic," *Elsevier*, 2021 (cited by many 2023+ works) — include for methodological context and feature engineering references.

[15] Research Case Study: "A Case Study with CICIDS2017 on the robustness of ML-based IDS against adversarial attacks," *ACM / 2023–2024 proceedings*.