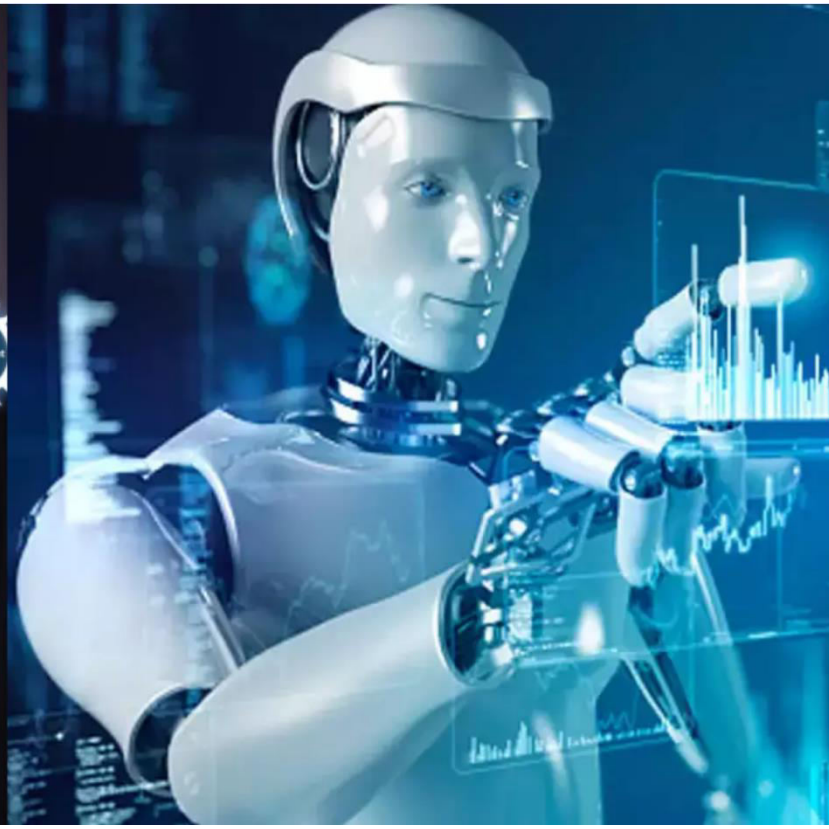
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

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.2025.1412091|

Anomaly-Based Network Intrusion Detection with Explainable Artificial Intelligence

Prof. Bhagya S G[1], Chinmayi M S[2], Nikitha Patil G M[3], Shreya B Maski[4], Vinay Laxman Attu[5]

Assistant Professor, Department of Artificial Intelligence & Machine Learning, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[1]

Student, Department of Artificial Intelligence & Machine Learning Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[2-5]

ABSTRACT: With the rapid growth of digital infrastructure, cybersecurity threats such as zero-day attacks, distributed denial-of-service (DDoS), and insider threats have become increasingly sophisticated. Traditional signature-based intrusion detection systems (IDS) fail to detect unknown attack patterns and often suffer from high false alarm rates. This research presents an Anomaly-Based Network Intrusion Detection System integrated with Explainable Artificial Intelligence

(XAI) to improve both detection accuracy and decision transparency. The system employs **Autoencoders and Isolation Forest algorithms** for anomaly detection and integrates **SHAP (SHapley Additive exPlanations)** for feature-level interpretability. The model is trained and evaluated using benchmark datasets such as **NSL-KDD and CIC-IDS 2017**. A real-time interactive dashboard is developed to visualize anomalies, alerts, and corresponding explanations. This approach enhances security analyst trust, reduces false positives, and provides actionable insights for modern cyber defense.

KEYWORDS: Intrusion Detection System, Anomaly Detection, Explainable AI, SHAP, Cybersecurity, Machine Learning, Autoencoder.

I. INTRODUCTION

Modern digital networks form the backbone of communication, business operations, healthcare, and critical infrastructure. However, this rapid digital evolution has also increased exposure to complex cyber threats such as zero-day exploits, advanced persistent threats (APT), and largescale denial-of-service attacks. Traditional IDS solutions rely on predefined signatures, making them ineffective against novel attacks.

Anomaly-based IDS offers a powerful alternative by learning normal network behavior and detecting deviations. However, most machine learning-based systems function as black boxes, providing limited explanations of their decisions. This lack of transparency reduces analyst trust and delays incident response. To overcome this limitation, this research integrates **Explainable Artificial Intelligence (XAI)** with anomaly detection. By using SHAP explanations, security analysts can understand why a specific network flow is classified as malicious. This combination improves trust, accountability, and regulatory compliance for IDS deployments in sensitive environments.

II. LITERATURE SURVEY

- Tavallaee et al. introduced the **NSL-KDD dataset**, addressing redundancy and imbalance issues in the original KDD99 dataset for IDS evaluation.
- Iglesias and Zseby applied **Autoencoder-based anomaly detection** in unsupervised environments for network traffic monitoring.
- Ribeiro et al. proposed **LIME**, while Lundberg and Lee introduced **SHAP**, both enabling transparency in black-box ML models.

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.2025.1412091|

- Recent research combines **deep learning and XAI**, proving that SHAP-based explanations significantly improve IDS reliability and analyst trust.

The reviewed studies confirm that while anomaly detection improves threat discovery, **explainability is essential for operational deployment**, which directly motivates the proposed system.

III. PROBLEM STATEMENT

Current intrusion detection systems struggle to identify evolving cyber threats and generate a high number of false alarms. Moreover, most deep learning-based IDS models lack explainability, making it difficult for security analysts to understand detection decisions. This research addresses the problem by developing an **interpretable anomaly-based IDS using XAI** to ensure both accurate detection and transparent reasoning.

IV. OBJECTIVES

- To design an anomaly-based IDS using machine learning techniques.
- To integrate SHAP for interpretable intrusion detection.
- To detect zero-day attacks and unknown threats efficiently.
- To visualize real-time intrusion alerts using a web-based dashboard.
- To enhance trust and reliability in automated cybersecurity systems.

V. PROPOSED SOLUTION

The proposed system utilizes **Autoencoders and Isolation Forest algorithms** to detect anomalies in network traffic. **SHAP explainability** is applied to highlight key network features influencing each decision. The system supports real-time traffic analysis and displays alerts through a user friendly dashboard. This allows cybersecurity professionals to understand and respond to threats effectively.

VI. SYSTEM DESIGN

6.1 System Architecture

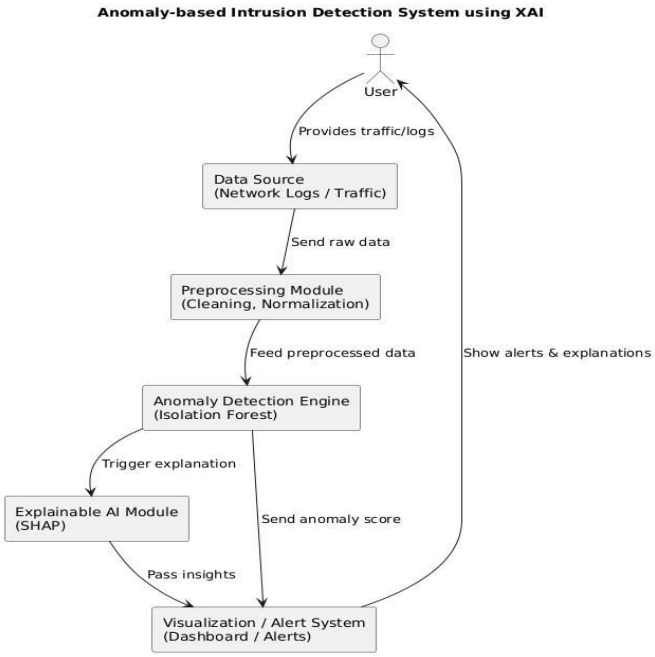


Fig 6.1: System Architecture

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.2025.1412091|

The system consists of the following modules:

- Traffic Capture Module
- Data Preprocessing Module
- Anomaly Detection Engine
- XAI Engine (SHAP)
- Visualization Dashboard
- Database Storage

Raw traffic flows are preprocessed and passed to the anomaly detection engine. Detected anomalies are explained using SHAP and visualized in the dashboard for analyst inspection.

VII. RESULTS

7.1 Dashboard Output

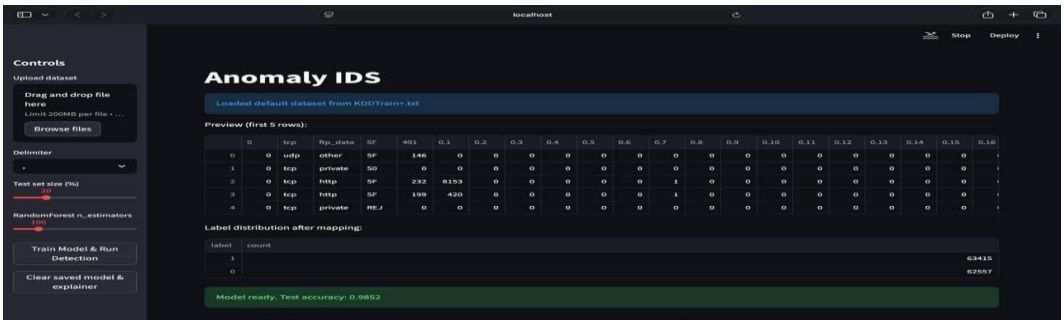


Fig 7.1: Dashboard

The proposed Intrusion Detection System provides an interactive control panel that enables users to configure and execute the detection process. Users can upload network traffic datasets in CSV or TXT format; if no dataset is provided, the system automatically loads the default **KDDTrain+** dataset. Support for both comma-separated and whitespace-separated data ensures compatibility with diverse dataset formats. Adjustable parameters include the test-train split ratio and the number of estimators in the Random Forest classifier, allowing users to balance accuracy and computational cost. The system also includes options to initiate model training and detection, as well as to clear the trained model and SHAP explainer for fresh retraining.

A dataset preview displays the initial records to provide insight into feature structure and network attributes. After preprocessing, the system presents the label distribution to verify correct mapping of normal and attack traffic. Model performance is reported through test accuracy, achieving high detection effectiveness, and a detailed classification report presenting precision, recall, and F1score for both classes. This interface ensures transparency, configurability, and reliable evaluation of the intrusion detection process.

7.2 Confusion Matrix Analysis

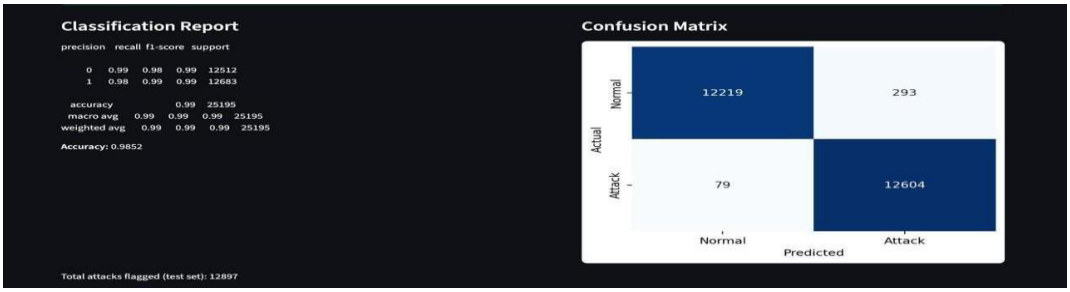


Fig 7.2: Confusion Matrix

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.2025.1412091|

The confusion matrix visualizes true vs. predicted classifications
Analysis Predicted → Normal Attack Actual Normal True Normal (TN) values False Alarm (FP) values Actual Attack Missed Attack (FN) values True Attack (TP) values **Interpretation:**

- High TP and TN values → Strong classification
- Very low FP → System avoids falsely marking normal traffic as malicious
- Low FN → Actual attacks are not missed The overall distribution shows the Random Forest classifier generalizes well.

7.3 SHAP Explainability Results
7.3.1 Global SHAP Summary

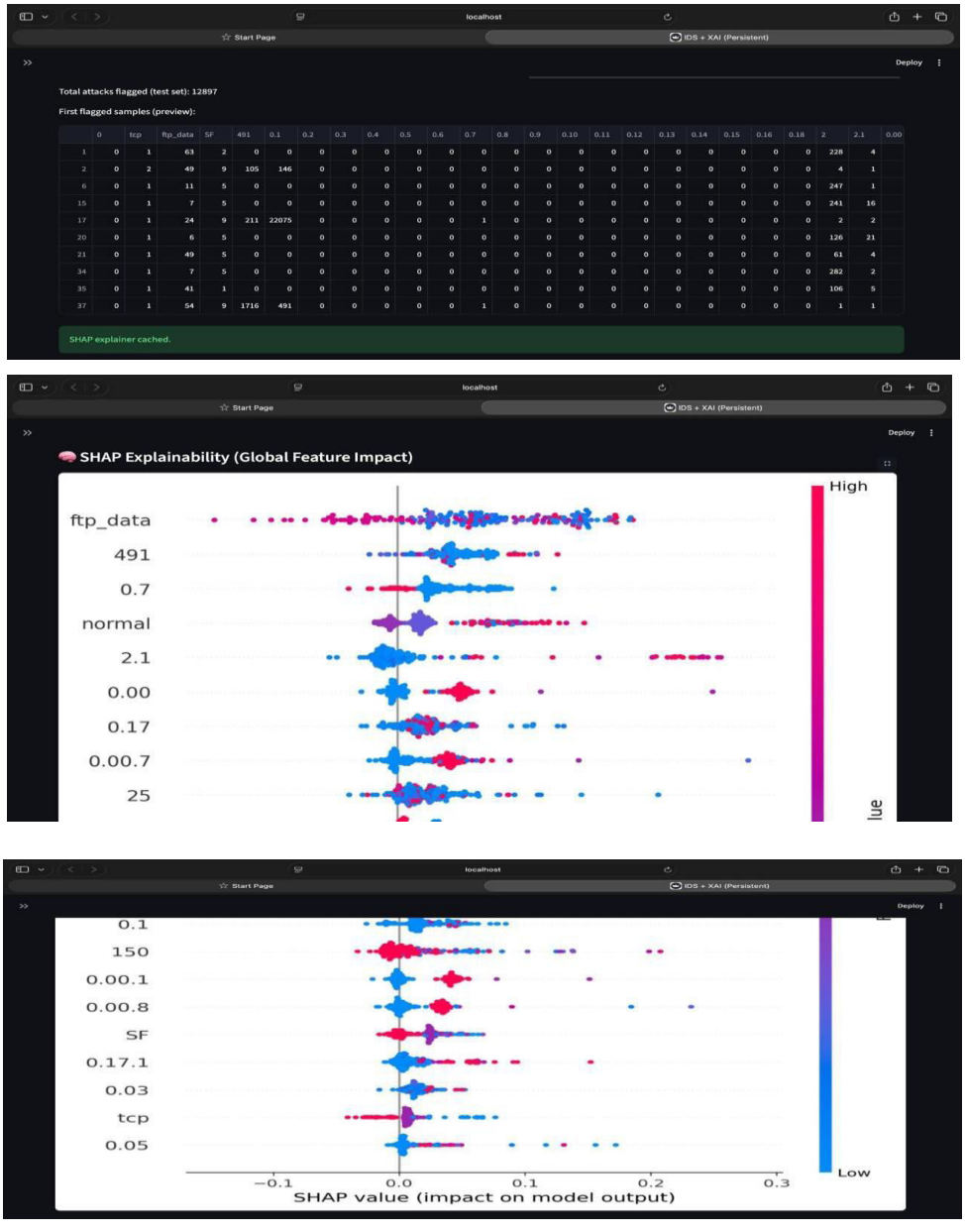


Fig 7.3.1: Global SHAP Summary

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.2025.1412091|

The SHAP summary plot highlights the top features influencing attack predictions. Common highly influential features include:

- service
- src_bytes
- dst_bytes
- serror_rate
- same_srv_rate
- dst_host_count

These features show the strongest contribution towards model decisions.

7.3.2 Local SHAP Explanation

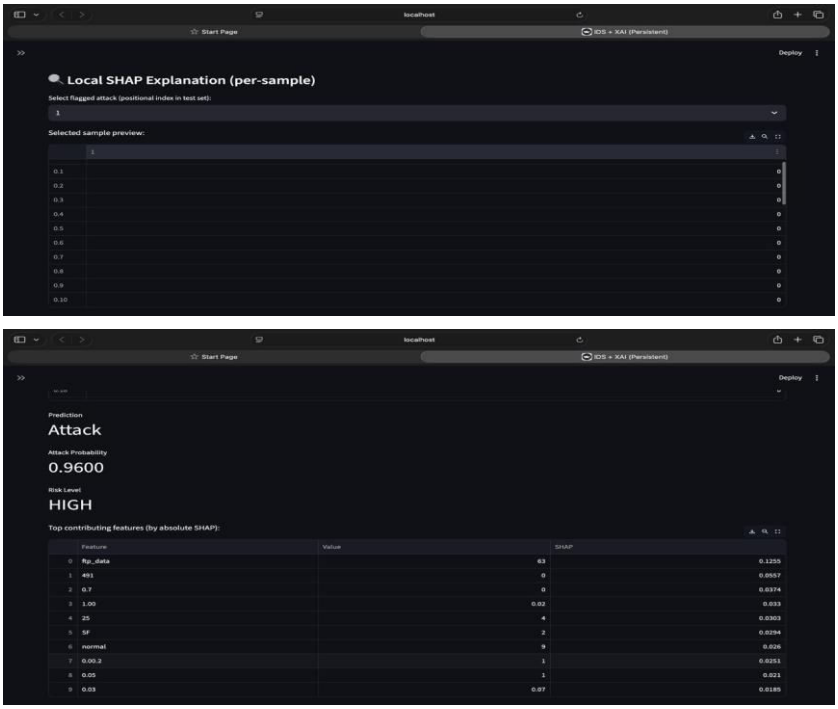


Fig 7.3.2: Local SHAP Exaplanation

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.2025.1412091|

For any specific attack sample:

- The waterfall plot reveals how individual features push the prediction toward attack or normal.
- Feature contributions clearly show which network behaviors triggered the alert.

This enables deeper analysis for security teams and justifies predictions.

VIII. CONCLUSION

This paper presents an anomaly-based Intrusion Detection System (IDS) that combines Machine Learning with Explainable Artificial Intelligence (XAI) to achieve accurate and interpretable attack detection. A Random Forest classifier is used to identify malicious network traffic, while SHAP provides transparent, feature-level explanations to address black-box limitations. The system follows an end-to-end pipeline including data preprocessing, model training, detection, evaluation, and visualization, achieving high detection accuracy on benchmark datasets. An interactive Streamlit dashboard improves usability by presenting detection results and explanations in an intuitive manner. Future work includes real-time intrusion detection, integration of deep learning models, cloud deployment, multi-class attack classification, and the use of advanced datasets to enhance scalability and real-world applicability.

REFERENCES

[1] R. Ahmad, F. Ahmad, M. Alazab and A. Jolfaei, "Explainable AI for Cybersecurity: State-of the-Art, Challenges, and Future Directions," IEEE Access, vol. 11, pp. 42836–42856, 2023.

[2] Y. Ren, Y. Zhang and W. Zhou, "XAI-Based Intrusion Detection Framework for Smart Networks," IEEE Internet of Things Journal, vol. 10, no. 3, pp. 2310–2320, Feb. 2023.

[3] M. Nasir, I. U. Haq and M. Aslam, "A Deep Learning-Based Anomaly Detection System with SHAP Interpretability," in Proc. 2022 Int. Conf. on Cyber Security and Protection of Digital Services (Cyber Security), pp. 1–8.

[4] A. Muniyandi, R. R. Jino Ramson and M. Dhanasekaran, "Anomaly Detection using Hybrid ML Models for Network Security," Computers & Security, vol. 126, pp. 102986, 2023.

[5] B. M. Mendoza and D. Z. Rodríguez, "A Comprehensive Survey on Explainable Artificial Intelligence for Intrusion Detection," Journal of Network and Computer Applications, vol. 223, pp. 103648, 2023.

[6] A. Sadeghzadeh, A. A. Ghorbani and M. Debbabi, "Explainable AI in Intrusion Detection Systems: Trends, Challenges, and Opportunities," IEEE Transactions on Dependable and Secure Computing, early access, 2024.

[7] N. Alshahrani and S. E. S. Abugharsa, "Explainable AI Using SHAP for Detecting Cybersecurity Threats in IoT," in Proc. 2023 Int. Conf. on AI and Smart Systems (ICAIS), pp. 78 83.

[8] D. Patel and A. K. Sharma, "Deep Autoencoder and SHAP for Interpretable Anomaly Detection," in Proc. 2022 IEEE Int. Conf. on Cyber Security and Digital Forensics (CyberSec), pp. 243–249.

[9] A. Hussain et al., "Explainable IDS for Cyber-Physical Systems Using SHAP and LSTM," IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 214–224, Jan. 2023.

[10] R. Saxena and P. Joshi, "An Interpretable Deep Learning-Based Intrusion Detection Framework," in Proc. 2022 IEEE Global Communications Conference (GLOBECOM), pp. 1–6.

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