www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 15 Issue 01 January 2026, Page No. 28129-28137
ISSN: 2319-7242 DOI: 10.18535/ijecs/v15i01.5346

An Explainable AI Hybrid Deep Learning Model for Cloud Security: A BiLSTM-Conv1D Based Approach

Dr. Vinod Desai[1], Prof. Ramanjinamma G[2], Manoj L Bhat[3], Thrisha N S[4], Bhuvan M[5], Krishnakant Bhushan[6]

[1]Associate Professor, [2]Assistant Professor, [3,4,5,6]Student
[1,2,3,4,5,6] Department of Computer Science and Engineering Sai Vidya Institute of Technology, Bengaluru, India

Abstract:
Applications and personal data have become vulnerable to a wide range of cyber threats as cloud infrastructures continue to expand. These threats include AI-powered attacks, unknown vulnerabilities, IoT-based attacks, and constantly evolving malware. Conventional intrusion detection systems often struggle in such environments due to high false-positive rates, limited adaptability and lack of transparent decision-making mechanisms. This study proposes an Explainable AI-driven Hybrid Intrusion Detection System (XAI-IDS) for cloud security that integrates Convolutional Neural Networks (Conv1D) and Bidirectional Long Short-Term Memory (BiLSTM) networks, to learn both temporal patterns and structural characteristics from network traffic, enabling real-time anomaly detection. To improve the interpretability and trust in the system's decision, SHAP (SHapely Additive Explanations) is integrated to highlight the contribution of each feature in the prediction. Experimental results indicate that the system achieves reliable detection performance and is suitable for practical deployment in cloud environments.

Keywords: Cloud security, Intrusion detection system, Conv1D, BiLSTM, Explainable AI, SHAP.

I. Introduction

Cloud computing lets companies scale quickly while getting tech help whenever needed – this shift transformed how they handle information storage, access, and usage. Still, moving vital apps and private details online has opened doors to smart digital attacks like shape-shifting viruses, machine-learning-powered breaches, unknown vulnerabilities, and hacks through connected gadgets. Old-style security monitors often miss these fast-changing risks since their rules don't adapt easily, flag too many harmless actions, and give unclear results.

Recent studies has looked into using deep learning to spot odd network activity and tackle these issues. Although tools such as Convolutional Neural Networks pick up patterns across space, meanwhile LSTM networks handle time-based changes well - making both fit for tracking traffic over time. Yet, these models are often seen as mysterious systems that hide how they work, limiting trust in high-stakes security areas were knowing why an alert happens really matters.

The Explainable AI-driven Hybrid Intrusion Detection System proposed in this research combines SHAP-based interpretability algorithms with Conv1D and Bi-directional LSTM (BiLSTM) models. The technology is made to identify intricate and dynamic cyberthreats in real time, giving security experts clear and useful information on unusual network activity. The suggested method overcomes major drawbacks of traditional IDS and improves overall cloud security by fusing interpretability, scalability, and high detection performance.

II. Literature Survey

Recent intrusion detection research increasingly focuses on combining deep learning with explainable AI to improve both detection accuracy and the clarity behind model decisions. Shoukat et al. proposed "Trust my IDS," an explainable AI-enabled system for industrial networks that leverages SHAP-based explanations to provide interpretable and trustworthy threat detection [1]. Similarly, Serrano developed CyberAIBot, an AI-driven IDS for IoT environments, which demonstrates adaptability and lightweight detection in resource-constrained devices [2]. While these contributions improve interpretability and adaptability, their focus on industrial and IoT systems limits their applicability to large-scale cloud environments.

To detect anomalies in network traffic, deep learning models especially those based on LSTM have grown prominent. To improve detection accuracy in imbalanced datasets. Devendiran and Turukmane developed Dugat-LSTM, which was optimized using chaotic methods [3]. Vibhute and Nakum investigated anomaly identification under unbalanced situations in cloud environments, showcasing the advantages of deep learning for minority attack class recognition [4]. To attain robustness across many threat types Alhayan Et al. extended this line of study by creating an ensemble of deep models optimized with a beluga whale method, but at a considerable computing cost [5]. These studies show the potential of optimization and ensemble approaches, but they also draw focus on persistent issues with interpretability and scalability.

Additionally, explainability in intrusion detection is becoming more and more important and IDS for IOT data streams that incorporates explainable deep learning to provide interpretable, low-latency alerts in real-time was proposed by Prashanth et al. [6]. In order to identify Advanced Persistent Threats (APTs), Ahmed et al. developed an explainable deep learning system that captures long term temporal patterns and aids forensic investigations [7]. In the same manner Chen et al. created a scalable hybrid IDS that learns both temporal and spatial aspects from massive network traffic using CNN and BiLSTM models [8]. Although these hybrid system increases accuracy and scalability, they still lack the integrated real-time interpretability needed for real world application.

To lower false positives and improve reliability, hybrid machine-learning and deep-learning techniques have also been investigated. To improve interpretability and performance, Sajid et al. suggested an ML-DL hybrid approach for cloud-based IDS that strikes a balance between conventional models and deep feature representations [9]. Transparent and adaptive AI algorithms that can could be used for intrusion detection, particularly for dynamic and complex threats, were presented by Agomuo et al. with a focus on fraud detection [10]. These studies demonstrate the potential of explainable and hybrid system, but their applicability is still restricted because of inadequate cloud-specific evaluation and insufficient defense against emerging AI-driven threats.

Despite significant progress, existing IDS solutions still face major challenges. Real-time interpretability for deep models such as LSTM and Bi-LSTM remains limited, and most systems struggle to detect AI-powered, polymorphic, and zero-day attacks. High false-positive rates continue to hinder large-scale deployment, and important cloud-specific challenges like multi-tenant traffic behavior and API-based anomalies, are often overlooked.

III. Proposed Methodology

The suggested approach uses a mix of deep learning parts - specifically BiLSTM and 1D Conv layers working together. Instead of just moving forward through data, the BiLSTM reads sequences both ways, catching timing links across time. Meanwhile, the Conv layers pull out key features by spotting small-scale patterns in traffic flow. Because it handles space-like structures and time-based changes at once, this combo works well against tricky threats. Even new or evolving risks - including unseen exploits or smart attack tools - struggle to bypass its detection.

Before reaching the model, network traffic undergoes a careful preprocessing pipeline where raw data is cleaned, normalized, and encoded. This step removes noise, fixes missing or inconsistent values, and scales features appropriately, ensuring high-quality input for reliable predictions. After preprocessing, the data flows through the Conv1D and BiLSTM layers and then into fully connected and softmax layers that classify traffic

as benign or malicious. SHAP-based explainable AI is incorporated to increase transparency, stressing the key elements underlying each prediction and offering valuable information to security experts.

The overall system is designed for cloud environments, supporting scalability, multi-tenant workloads, and real-time operation with minimal latency. As shown in Fig. 1., the architecture consists of sequential stages, beginning with the preprocessing pipeline and continuing through hybrid Conv1D-BiLSTM layers, classification, explainability, and cloud deployment. This design ensures reliable and interpretable intrusion detection suitable for practical cloud environments.

3.1 Data Collection

The first stage involves dataset collection, where diverse network traffic datasets such as NSL-KDD, CIC-IDS, and ToN-IoT are gathered. These datasets contain a mixture of benign and malicious traffic, providing a balanced foundation for building a robust model capable of detecting a wide spectrum of attack patterns.

3.2 Data Preprocessing

The next step's about prepping the data so it works well with deep learning models. Cleaning up gaps or weird values comes first, then turning labels into numbers instead of words. After that, number ranges get adjusted so nothing throws off the system later. By sorting out messiness early, this phase helps the model focus on real trends, not random glitches. It smooths things out before training even starts.

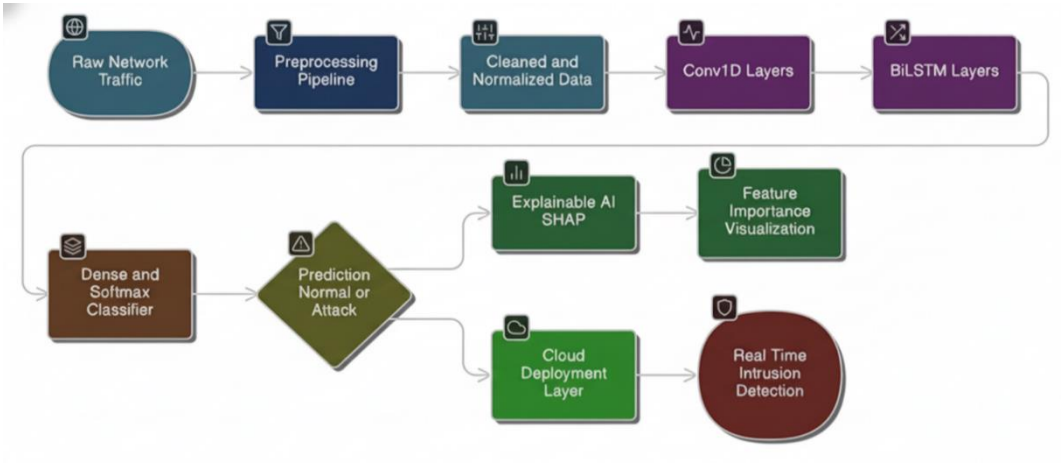

Fig. 1. System Architecture of Hybrid LSTM-based IDS

Alt Text: Block diagram illustrating the hybrid Conv1D–BiLSTM intrusion detection system architecture, including preprocessing, feature extraction, classification, explainability, and cloud deployment stages.

3.3 Hybrid Model Training

Once the data is prepared, the third stage, hybrid model training, is carried out. The proposed architecture integrates Conv1D layers to capture spatial dependencies among network features and BiLSTM layers to learn temporal dependencies in traffic sequences. Together, these layers enable the system to detect both localized anomalies and long-term behavioural patterns in attacks. Our model is trained on pre-processed dataset, with the final classification performed using dense and SoftMax layers.

3.4 Model Evaluation

In step four, we check how well the trained model works by looking at key measures - accuracy, precision, recall, or F-score. These numbers show whether it can spot bad activity without flagging too many harmless cases. After that review, we try it out on unknown traffic data so we can see if it adapts well to new patterns. This helps ensure it stays strong when facing fresh or changing threats.

3.5 Explainability Integration

To provide transparency, the fifth stage incorporates explainability with SHAP, which highlights the most influential features contributing to each detection decision. This allows analysts to understand why the system classified a particular instance as benign or malicious, thereby improving trust in the system.

3.6 Deployment

Finally, in the deployment stage, the trained and explainable model is integrated into a cloud environment for real-time intrusion detection, ensuring scalability, adaptability, and interpretability in operational scenarios.

IV. Eresearch-educationxperimental Results

The hybrid Conv1D–BiLSTM Intrusion Detection System under consideration was assessed through three classic datasets:NSL-KDD, CICIDS-2017 V2 and TON-IOT, which capture traditional, modern cyber intursions, and IoT-based attacks, correspondingly, thereby addressing the range of potential attacks relevant to environments leveraging cloud services. The datasets were subjected to preprocessing (e.g., cleaning, normalization, encoding), so they would maintain a high-quality level across training and evaluation.

The hybrid architecture delivered strong results across all datasets, demonstrating both its robustness and adaptability. It achieved 98.77% accuracy for NSL-KDD dataset and 98.63% accuracy on CIC-IDS 2017 V2. Even on the more challenging ToN-IoT dataset known for its highly varied network traffic our model performed remarkably, achieving 98.77% accuracy. These results illustrate the capacity of the hybrid Conv1D–BiLSTM architecture the ability detect both spatial and temporal patterns of adversarial attacks, which was demonstrated through accurate identification of malicious activities across different contexts.

In addition to accuracy, the performance of the model was evaluated using precision, recall and F1-score to assess the model's effectiveness. The detailed performance metrics obtained across all benchmark datasets are summarized in Table 1, providing comparative evaluation with existing approaches. Furthermore, SHAP adds clarity by showing which features mattered most during predictions. It points out key traffic patterns behind each result, helping experts check if outputs make sense. Because users can see how decisions form, trust grows naturally. This clear approach keeps high performance while making tech easier to grasp and use in live cyber defense setups.

4.1. Performance Metrics

The proposed IDS's performance was assessed using a collection of commonly used metrics that capture various facets of model behavior. These metrics demonstrate the system's capacity to properly identify genuine threats while reducing false detections, in addition to capturing how well it separates attack from regular traffic. When taken as a whole, these metrics offer a clear and thorough insight of the systems efficiency across all assessed datasets.

4.4.1. Accuracy (ACC)

Accuracy represents overall effectiveness by measuring how may predictions are correct out of all the predictions made. It gives a general idea of how good the system performs across both attack and normal samples. This relationship is expressed using Eq. (4.1):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (4.1)$$

4.4.2. Precision (PRE)

Precision shows how trustworthy the model's attack predictions are. Instead, it checks what share of flagged cases were actual attacks. That tells you how frequently the IDS gets it right when sounding the alarm. You can find this value using Eq. (4.2):

$$PRE = \frac{TP}{TP + FP} \times 100 \qquad (4.2)$$

Table 1: Performance comparison of proposed model with existing approaches on benchmark datasets

Sl No.	*Existing System*	*Dataset*	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*
1.	Muhammad Sajid et al. [5], 2024	NSL-KDD	88.41	-	-	98.64
2.	Nazreen Banu A et al. [11], 2025	TON-IOT	95.8	94.3	93.7	94.0
3.	Shifa Shoukat et al. [1], 2025	CICIDS 2017 - V2	98.57	95.12	83.02	84.52
4.	Proposed System-BILSTM-CONV1D	NSL-KDD	98.77	99.13	98.63	99.11
5.	Proposed System-BILSTM-CONV1D	TON-IOT	98.63	99.13	98.27	98.44
6.	Proposed System-BILSTM-CONV1D	CICIDS 2017-V2	98.77	98.66	98.16	98.03

4.4.3. Recall (REC)

Recall gives the model's ability to detect all relevant attack instances. It measures how many of the actual attacks were successfully identified by the system, making it a critical metric for intrusion detection. The calculation is shown in Eq. (4.3):

$$REC = \frac{TP}{TP + FN} \times 100 \qquad (4.3)$$

4.4.4. F1-score (F1)

F-score links recall with precision through one number. It works well when handling uneven data, especially if one group outnumbers another. You get the F-score by following Formula, Eq. (4.4):

$$F1 = 2 \times \frac{PRE\ X\ REC}{PRE + REC} \times 100 \qquad (4.4)$$

The results from the experiments demonstrate that the proposed system offers a substantial enhancement over traditional IDS approaches in terms of accuracy, robustness, and provide explainability, making the system a worthy candidate for deploying as a platform for securing an evolving cloud-based security model.

The comparative performance of the proposed system across different datasets is shown in Fig. 2, highlighting accuracy, precision, recall and F1-Score metrics.

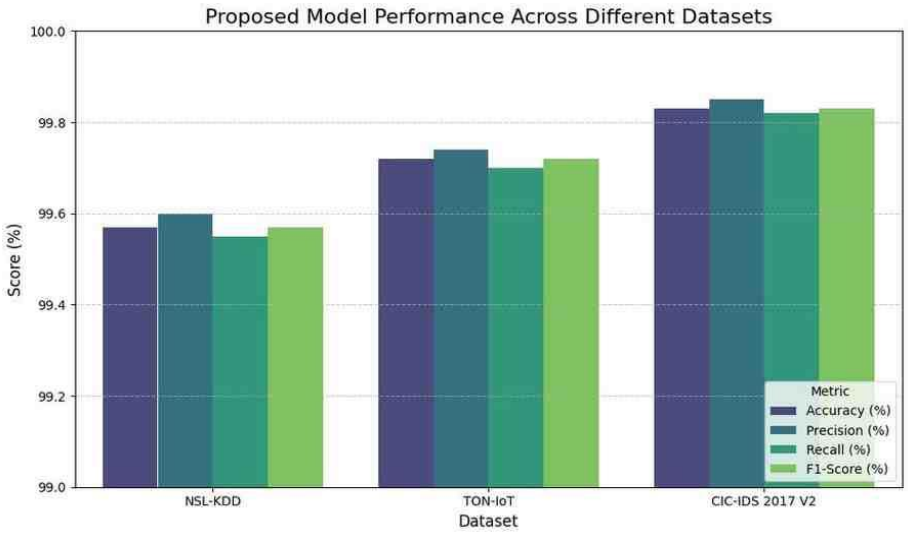

Fig. 2. Model Performance Scores

Alt Text: Bar chart showing accuracy, precision, recall, and F1-score of the proposed intrusion detection system across benchmark datasets.

4.2 Explainability Results using SHAP

SHAP (SHapely Additive exPlainations) helped check what each input does in the model's choices, so we can see why the mix of ConvD and BiLSTM picks certain outcomes. Instead of guessing, this tool reveals which parts drive results for every prediction. It uncovers overall trends plus individual triggers behind each class decision. The summary chart from NSL-KDD data (see Fig. 3) points out top traits shaping attack spotting. Values like count, login status, service type, along with errors, heavily shift the outcome. How far these numbers go up or down affects whether something looks suspicious. This range tells us how the model views normal vs odd traffic behavior.

For the CICIDS-2017 dataset, the SHAP bar plot (Fig. 4) presents the global importance of the most impactful traffic features. Elements like PSH_Flag_Count, Idle_Mean, Subflow_Fwd_Bytes, and Flow_Duration stand out as major contributors. These finding imply that the model successfully captures temporal trends and protocol-level characteristics that differentiates normal behavior from malicious activity in modern cloud environments.

Overall, SHAP-based explainability adds a valuable layer of insight by highlighting the logic behind the model's decisions. This level of interpretability not only improves the system's credibility but also supports its practical deployment in real cloud security settings, where understanding the cause of alerts is as important as detecting them.

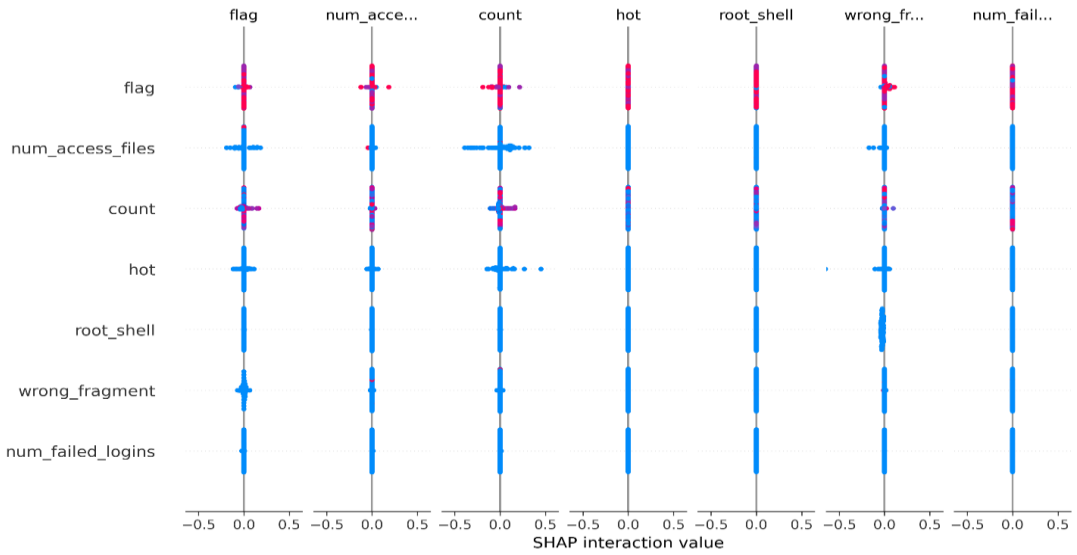

Fig. 3. SHAP Summary Plot for NSL-KDD Dataset

Alt Text: SHAP summary plot illustrating the influence of key network traffic features on intrusion detection decisions for the NSL-KDD dataset.

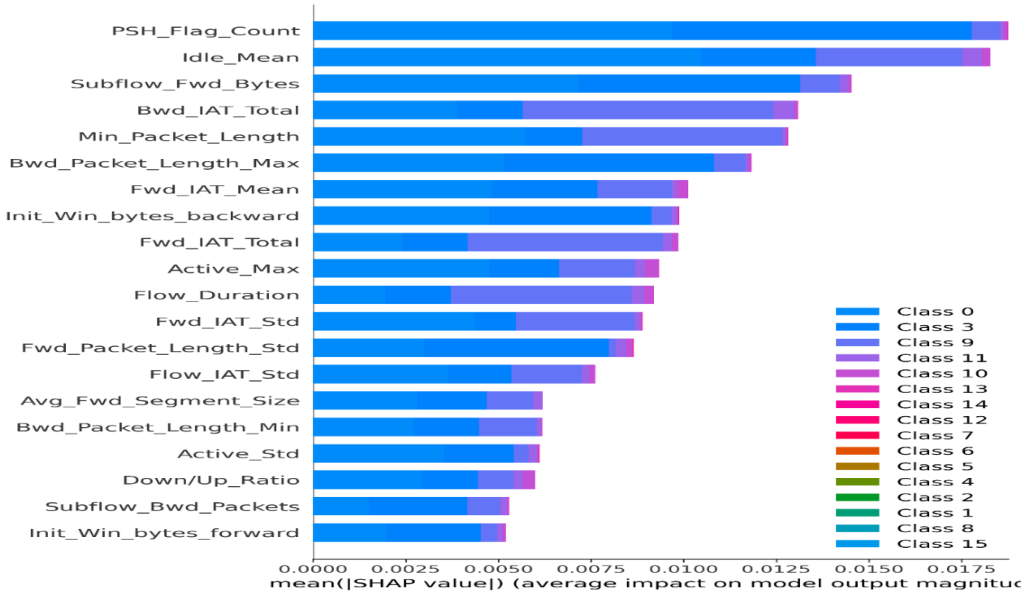

Fig. 4. SHAP feature-importance bar-plot for CICIDS-2017 Dataset

Alt Text: SHAP feature-importance bar plot displaying the most influential features contributing to intrusion detection outcomes for the CICIDS-2017 dataset.

V. Conclusion and Future Work

The proposed hybrid Conv1D–BiLSTM Intrusion Detection System, integrated with SHAP-based Explainable AI, shows excellent capability in detecting a diverse range of modern and evolving cyber threats across cloud-based datasets. The model continuously attains high accuracy, precision, recall and F1-scores,

by learning both spatial and temporal aspects of network traffic. SHAP further improves the system by providing intuitive visual explanations that highlight which features most influence each prediction, helping analysts understand and trust the model's decisions. Although some attack categories in the datasets remain imbalanced and SHAP introduces a small amount of additional computation, the system still offers an effective balance of performance, transparency, and real-world practicality for cloud security environments.

The experimental results confirm that the proposed IDS improves significantly over traditional approaches in-terms of robustness, detection quality, and interpretability. Its potential to generate clear, actionable insights helps security analysts respond to threats more efficiently. Looking ahead, the system can be expanded to support multi-cloud deployments, leverage federated learning for privacy-preserving distributed training, and incorporate automated response mechanisms. Further improvements—such as creating lightweight versions for resource-constrained environments, enhancing resistance to adversarial attacks, and integrating richer telemetry sources—will continue to strengthen the system's scalability, adaptability, and usefulness for real-time cybersecurity challenges.

DATA AVAILABILITY STATEMENT

The datasets used in this study are publicly available benchmark datasets and can be shared upon reasonable request from the corresponding author.

Competing Interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

References

1. Shifa Shoukat, Tianhan Gao, Danish Javeed, Muhammad Shahid Saeed, Muhammad Adil, "Trust my IDS: An explainable AI integrated deep learning-based transparent threat detection system for industrial networks", Computers & Security, Volume 149, Article Number 104191, 2025.
2. Serrano, W. "CyberAIBot: Artificial Intelligence in an intrusion detection system for CyberSecurity in the IoT". Future Generation Computer Systems, 166, 107543, 2025.
3. Devendiran, R., & Turukmane, A. V." Dugat-LSTM: Deep learning based network intrusion detection system using chaotic optimization strategy". Expert Systems with Applications, 245, 123027, 2024.
4. Vibhute, A. D., & Nakum, V." Deep learning-based network anomaly detection and classification in an imbalanced cloud environment". Procedia Computer Science, 232, 1636-1645, 2024.
5. Alhayan, F., Alruwais, N., Alamgeer, M., Alashjaee, A. M., Abdullah, M., Khadidos, A. O., ... & Alshareef, A." Design of advanced intrusion detection in cybersecurity using ensemble of deep learning models with an improved beluga whale optimization algorithm". Alexandria Engineering Journal, 121, 90-102, 2025.
6. Muhammad Sajid, Kaleem Razzaq Malik, Ahmad Almogren, Tauqeer Safdar Malik, Ali Haider Khan, Jawad Tanveer, Ateeq Ur Rehman, "Enhancing intursion detection: a hybrid machine and deep learning", Journal of Cloud Computing: Advances, Systems and Applications, Volume 13, Article Number 123, 2024
7. Prasanth, B. Karthikeyan, S. R. Gupta, "An Intrusion Detection System over IoT Data Streams Using Explainable Deep Learning", Sensors, vol. 25, no. 3, pp. 847, 2025.
8. Ahmed, J. Lee, M. Hassan, "Explainable Deep Learning Approach for Advanced Persistent Threats Detection", Artificial Intelligence Review, vol. 37, no. 2, pp. 211-228, 2024.
9. Kanumalli, S. S., Lavanya, K., Rajeswari, A., Samyuktha, P., & Tejaswi, M." A scalable network intrusion detection system using bi-lstm and cnn". In 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 1-6). IEEE, 2023.
10. Agomuo, O. C., Uzoma, A. K., Khan, Z., Otuomasirichi, A. I., & Muzamal, J. H." Transparent AI for Adaptive Fraud Detection". In 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-6). IEEE, 2025.

11. Nazreen Banu A and S.K.B. Sangeetha, "Intrumer: A Multi Module Distributed Explainable IDS/IPS for Securing Cloud Environment", Computers, Materials & Continua, Volume 82, Article Number 059805, 2025.
12. Utsav Upadhyay, Alok Kumar, Umashankar Rawat, Satyabrata Roy, Sandeep Chaurasia, "Defending the Cloud: Understanding the role of Explainable AI in Intrusion Detection Systems", 2023 16th International Conference on Securityof Information and Networks (SIN), Article Number 10475080, 2023.
13. Diogo Gaspar, Paulo Silva, and Catarina Silva, “Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron”, IEEE Access, Volume 12, Article Number 3368377, 2024
14. Jisna P, Jarin T, Praveen P N, “Advanced Intrusion Detection Using Deep Learning-LSTM Network On Cloud Environment”, 2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS), Article Number 9673607, 2021
15. Iqbal H. Sarker, Helge Janicke, Ahmad Mohsin, Asif Gill, Leandros Maglaras, “Explainable AI for cybersecurity automation, intelligence and trustworthiness in digital twin: Methods, taxonomy, challenges and prospects”, ICT Express, Volume 10, pp. 935-958, 2024.