Babylonian Journal of Internet of Things
Vol. 2026, **pp**. 46–56
DOI: https://doi.org/10.58496/BJIoT/2026/005; ISSN: 3006-1083
https://mesopotamian.press/journals/index.php/BJIoT

An Explainable Hybrid Deep Learning Framework for Unsupervised Network Intrusion Detection Using Isolation Forest and Bidirectional Long Short-Term Memory Autoencoder

Aerkera Kumawuese Daniel-Beston [1], Aamo Iorliam [1,*]

[1] *Department of Mathematics/Computer Science, Rev. Fr. Moses Orshio Adasu University, Makurdi (Formerly Benue State University, Makurdi), Nigeria.*

ARTICLE INFO

Article History
Received 17 Apr 2026
Revised 2 May 2026
Accepted 3 Jun 2026
Published 25 Jun 2026

Keywords
Internet of Things (IoT),
Network Intrusion Detection,
Isolation Forest,
BiLSTM Autoencoder,
Explainable Artificial Intelligence (XAI).

ABSTRACT

The increasing proliferation of Internet of Things (IoT) devices has introduced significant cybersecurity vulnerabilities due to huge data these devices generate, and the growing sophistication of cyberattacks. Traditional signature-based intrusion detection systems are often ineffective against zero-day and evolving threats, particularly in dynamic IoT environments where labeled datasets are scarce. This paper proposes an explainable hybrid unsupervised deep learning framework that integrates Isolation Forest and Bidirectional Long Short-Term Memory (BiLSTM) autoencoder models for intelligent network intrusion detection. The framework combines statistical anomaly isolation with deep temporal sequence learning to enhance the detection of abnormal network behaviors without relying on labeled data. Network traffic data obtained from the Kaggle network traffic dataset were preprocessed through data cleaning, feature engineering, normalization, and temporal sequence generation. A weighted ensemble mechanism was employed to combine anomaly scores from both models, while SHapley Additive exPlanations (SHAP) and LIME (Local Interpretable Model-Agnostic Explanations) techniques were integrated to improve interpretability and transparency of detection decisions. Experimental results demonstrated strong convergence of the BiLSTM Autoencoder with extremely low reconstruction losses and effective discrimination between normal and malicious traffic patterns. The hybrid framework successfully detected anomalous traffic bursts and suspicious communication behaviors, with "packets_per_time_unit" identified as the most influential anomaly indicator. The proposed framework provides an efficient, scalable, and explainable solution for adaptive IoT cybersecurity and intrusion detection in heterogeneous network environments.

1. INTRODUCTION

The rapid evolution of the Internet of Things (IoT) has transformed the way devices communicate, interact, and exchange data across heterogeneous environments [1]. IoT simply refers to a network of interconnected physical devices embedded with sensors, software, and communication capabilities that enable them to collect and exchange data over the internet [2]. These devices range from smart home appliances and wearable devices to industrial control systems and critical infrastructure components. As IoT continues to expand, it has found applications in diverse domains such as healthcare, transportation, agriculture, and smart cities [1], [2].

Despite its numerous benefits, IoT networks are highly susceptible to security threats due to their distributed nature, limited computational resources, and lack of standardized security mechanisms. Most traditional security approaches often rely on signature-based detection systems, which are inadequate in identifying novel or unknown attacks [3]. Network anomalies in IoT environments may arise from malicious activities such as Distributed Denial of Service (DDoS) attacks, unauthorized access, data exfiltration, or even device malfunction. These anomalies can significantly compromise the integrity, confidentiality, and availability of IoT systems [4].

Anomaly detection in IoT networks involves identifying patterns in network traffic that deviate from normal behavior. Unlike conventional intrusion detection systems, anomaly detection systems focus on detecting previously unseen threats by learning normal patterns and flagging deviations. However, the dynamic and high-dimensional nature of IoT data poses significant challenges to effective anomaly detection [4].

Advancements in Machine Learning (ML), particularly in unsupervised learning techniques, have provided new

*Corresponding author. Email: aamoiorliam@gmail.com

opportunities for addressing these challenges. Unsupervised learning algorithms such as clustering, autoencoders, and dimensionality reduction techniques can learn inherent structures in data without requiring labeled datasets. This is particularly useful in IoT environments where labeled data is scarce or unavailable. Techniques such as K-means clustering, Principal Component Analysis (PCA), and Deep Autoencoders have shown promising results in detecting anomalies in network traffic [5].

This paper focuses on developing an IoT network anomaly detection system using unsupervised learning techniques such as Isolation Forest and Bidirectional LSTM Autoencoder. The Isolation Forest isolates anomalies efficiently without needing cluster density assumptions, while the Bidirectional LSTM Autoencoder understands sequence behavior. The system analyzes network traffic data to identify abnormal patterns without prior knowledge of attack signatures. The preprocessing stage involves feature extraction and normalization of network traffic data, making it suitable for model training. The proposed system aims to enhance the detection of unknown threats and improve the overall security posture of IoT networks.

This is needed because the proliferation of IoT devices has introduced significant security challenges, primarily due to their inherent vulnerabilities and lack of robust security frameworks. IoT networks often operate in open and dynamic environments, making them attractive targets for cyber attackers. Traditional intrusion detection systems rely heavily on labeled datasets and predefined attack signatures, which limits their ability to detect zero-day exploits and evolving threats.

One major issue is the difficulty in obtaining labeled IoT network traffic datasets for supervised learning approaches. Labeling large volumes of network data is time-consuming, expensive, and often impractical. Consequently, many existing detection systems fail to generalize well in real-world IoT environments. Additionally, IoT devices generate massive volumes of heterogeneous data, making it challenging to identify meaningful patterns using conventional techniques.

Another challenge is the high rate of false positives and false negatives in existing anomaly detection systems. False positives can lead to unnecessary alerts and resource wastage, while false negatives may allow malicious activities to go undetected. Furthermore, resource constraints in IoT devices, such as limited memory and processing power, restrict the deployment of complex security mechanisms.

Therefore, there is a need for an efficient and scalable anomaly detection system that can operate without labeled data and effectively identify both known and unknown threats in IoT networks. This study proposes the use of unsupervised learning techniques to address these challenges and improve the detection accuracy of IoT network anomalies.

The main contributions of this paper as summarized:

1. Develop an IoT network anomaly detection model using a hybrid Isolation Forest and BiLSTM autoencoder in Python for identifying abnormal network behavior.
2. Evaluate the performance of the developed anomaly detection system using metrics such as accuracy, precision, recall, F1-score, false positive rate, and detection efficiency.
3. Compare the effectiveness of the developed hybrid model in detecting unknown anomalies in IoT network traffic against other similar models.

2. RELATED WORK

Several researchers have explored the use of machine learning and deep learning techniques for IoT anomaly detection.

Neto et al. [6] introduced the CICIoT2023 dataset for cybersecurity, intrusion detection, and IoT. The authors proposed a comprehensive IoT attack dataset generated from a realistic topology of 105 IoT devices. A total of 33 attacks were executed and grouped into seven categories, including distributed denial of service (DDoS), denial of service (DoS), reconnaissance, web-based, brute-force, spoofing, and Mirai botnet, enhancing security research resources.

Mohammed et al. [4] investigated how machine learning can improve IoT network security by detecting DDoS attacks. The study preprocessed IoT traffic datasets through cleaning, normalization, label encoding, and feature ranking before applying machine learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest. The system analyzed traffic patterns to distinguish malicious from normal behavior. Results were evaluated using accuracy, precision, recall, and F1-score metrics. The research concluded that machine learning-based anomaly detection provided an adaptive and effective defense mechanism for protecting IoT environments against evolving DDoS threats.

Authors in [5] presented an advanced cybersecurity framework for detecting anomalous network traffic in encrypted environments. The study combined EFMS-enhanced KMeans clustering with a CNN-GRU deep learning model to improve anomaly detection accuracy and reduce false positives. Using real-world firewall traffic logs and SMOTE balancing, the proposed architecture achieved high accuracy, scalability, and efficiency in identifying malicious traffic. The paper effectively demonstrated the benefits of integrating unsupervised clustering with sequential deep learning for modern cybersecurity challenges, particularly in IoT, edge computing, and encrypted network environments.

Moamin et al. [3] comprehensively surveyed machine learning, deep learning, and ensemble-based approaches for cybersecurity applications, highlighting the effectiveness of supervised, unsupervised, and reinforcement learning techniques in detecting evolving attacks. The study examined static, dynamic, and hybrid analysis strategies alongside publicly available datasets and deployment challenges. The authors emphasized explainable AI and adaptive intrusion prevention systems as promising future research directions.

Authors in [2] reviewed the applications and security concerns of IoT systems. The authors presented a comprehensive overview of the Internet of Things (IoT) as a transformative technology that has significantly influenced modern industries and everyday life. They examined the evolution, foundational concepts, architecture, and key components of IoT systems, emphasizing the role of interconnected devices and sensors in enabling intelligent decision-making and seamless data exchange. Through real-world applications and use cases, the authors demonstrated the potential of IoT to enhance efficiency, connectivity, and innovation. Their study concluded that machine learning techniques provide effective solutions for securing heterogeneous IoT environments.

Adefemi, Mutanga, and Alimi [7] presented an effective deep learning framework for improving cybersecurity in IoT environments. The study integrated Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU) to capture both spatial and temporal patterns in network traffic. Using the IoTID20 and BoT-IoT benchmark datasets, the proposed model achieved a detection accuracy of 99.83% and 99.01%, outperforming several baseline methods such as CNN, GRU, LSTM, and FFNN. The authors demonstrated strong methodological rigor through preprocessing, SMOTE balancing, and statistical evaluation. Overall, it provided a significant contribution to intelligent IoT intrusion detection research.

Mzili et. al. [1] investigated interoperability challenges in the Internet of Things (IoT) by analyzing existing taxonomies and identifying barriers that hinder seamless integration among heterogeneous devices. Employing a systematic literature review and a mixed-method framework, the study categorized interoperability into technical, semantic, organizational, and syntactic dimensions. The findings revealed that fragmented standards, protocol heterogeneity, data incompatibility, security concerns, and scalability issues remain major obstacles to effective IoT deployment. The authors emphasized the need for standardized frameworks, collaborative governance, and the integration of emerging technologies such as blockchain and artificial intelligence to enhance interoperability.

Kamal and Mashaly [8] examined the application of intelligent anomaly detection techniques in Internet of Things (IoT) environments. The authors emphasized that the rapid growth of IoT devices has significantly increased cybersecurity vulnerabilities, thereby necessitating efficient anomaly detection systems capable of identifying abnormal network behaviors in real time. The study explored machine learning and artificial intelligence-based approaches for detecting malicious activities within IoT networks while addressing challenges such as high-dimensional data, scalability, and resource constraints. The findings revealed that deep learning and unsupervised learning techniques improve anomaly detection accuracy and adaptability in dynamic IoT environments, making them suitable for detecting evolving cyber threats and unknown attacks.

Hussein and Répás [9] reviewed the application of machine learning-based intrusion detection methods in Internet of Things (IoT) systems. The authors emphasized that the rapid expansion of IoT devices has introduced significant cybersecurity challenges due to heterogeneous architectures, limited device resources, and evolving attack patterns. The study examined supervised, unsupervised, deep learning, and hybrid machine learning approaches for intrusion detection and anomaly detection in IoT environments. The findings revealed that machine learning techniques provide intelligent and adaptive security mechanisms capable of detecting both known and unknown attacks more effectively than traditional intrusion detection systems.

Nayak, Pawar, & BL [10] examined the application of advanced computational intelligence and machine learning techniques for solving complex engineering and data analysis problems. The authors emphasized the growing importance of intelligent algorithms in analyzing high-dimensional and dynamic datasets generated in modern digital systems. The study highlighted that machine learning and data-driven approaches improve prediction accuracy, anomaly identification, and decision-making efficiency compared to traditional analytical methods. Furthermore, the researchers discussed the challenges of computational complexity, scalability, and model optimization in real-world applications. The findings demonstrated that intelligent learning models provide effective solutions for handling complex nonlinear patterns and evolving system behaviors. The related literature reviewed in this paper is summarised in Table 1.

Several studies have also utilized Isolation Forest for anomaly detection because of its efficiency in identifying abnormal network behaviors without requiring labeled datasets. Likewise, LSTM Autoencoder models have demonstrated excellent performance in sequential anomaly detection tasks involving network traffic analysis. Despite these advancements, many existing systems suffer from high false positive rates, computational complexity, and poor generalization in real-world IoT environments. This study therefore proposes the combined use of Isolation Forest and BiLSTM autoencoder algorithms to improve anomaly detection accuracy and efficiency in IoT networks.

TABLE 1: SUMMARY OF SOME RELATED WORKS

Author(s)	Findings	Techniques Adopted
[1]	The authors identified blockchain and artificial intelligence as promising technologies for enhancing future IoT interoperability.	Comparative analysis of IoT protocols, and conceptual framework development
[2]	Reviewed IoT applications and security concerns, highlighting the effectiveness of ML for IoT security	Machine learning techniques

Author(s)	Findings	Techniques Adopted
[3]	Provided a comprehensive survey of AI-driven malware and network intrusion detection approaches. Deep learning and ensemble methods achieved superior detection performance compared with traditional techniques. The study emphasized explainable AI, AutoML, and adaptive context-aware intrusion prevention systems as promising directions for future cybersecurity research.	Machine learning, deep learning, ensemble learning, Cybersecurity analysis techniques, Network intrusion prevention systems, AI-based Malware detection frameworks.
[4]	Developed an IoT anomaly detection system for DDoS attack detection with effective adaptive defense performance	SVM, KNN, Random Forest
[5]	Proposed an anomaly detection framework for encrypted network traffic with high accuracy and reduced false positives	EFMS-KMeans clustering, CNN-GRU deep learning
[6]	Introduced the CICIoT2023 dataset for cybersecurity, intrusion detection, and IoT	IoT traffic dataset, machine learning, deep learning
[7]	Developed a deep learning framework for IoT intrusion detection with very high detection accuracy	CNN-GRU deep learning model
[8]	Examined intelligent anomaly detection methods for real-time IoT cybersecurity	Deep learning, Unsupervised learning, AI-based detection
[9]	Reviewed machine learning intrusion detection approaches for IoT systems	Supervised, Unsupervised, Deep learning, Hybrid ML
[10]	Investigated intelligent computational techniques for anomaly detection and prediction in dynamic systems	Machine learning, Computational intelligence
[11]	Investigated real-world network traffic behavior and successfully detected anomalous traffic bursts, abnormal communication diversity, and suspicious traffic intervals without labeled attack data.	Isolation Forest anomaly detection based on Wireshark traffic capture
Proposed Study	IoT network anomaly detection using unsupervised learning for adaptive cyber threat detection	Hybrid Isolation Forest and BiLSTM Autoencoder

The review of related works revealed that existing IoT anomaly detection systems face challenges such as high false positive rates, dependency on labeled datasets, and inability to detect evolving cyber threats effectively. Consequently, this study proposes an IoT network anomaly detection system using hybrid Isolation Forest algorithms and BiLSTM Autoencoder to improve anomaly detection performance and enhance IoT network security.

3. METHODOLOGY

3.1 Dataset Description

This study utilized the publicly available Kaggle network traffic dataset available from: https://www.kaggle.com/datasets/ravikumargattu/network-traffic-dataset [12] for the development and evaluation of a hybrid unsupervised anomaly detection framework for network intrusion analysis. The dataset contains heterogeneous network traffic records captured from real communication environments and includes multiple packet-level and flow-level attributes associated with network behavior.

The dataset files were downloaded programmatically using the KaggleHub library within the Python environment. Recursive file searching was employed to automatically locate all CSV files contained within the dataset directory structure. Subsequently, all discovered CSV files were merged into a unified dataframe to facilitate centralized preprocessing and analysis.

3.2 Data Acquisition and Integration

The dataset acquisition process was automated using the KaggleHub API. The downloaded dataset directory was scanned recursively using the Glob library to identify all available CSV files. Each CSV file was loaded independently using the Pandas data analysis library.

Datasets containing attack-related columns such as label, attack, category, or subcategory were prioritized during

concatenation. Multiple datasets were subsequently merged into a single consolidated dataframe using vertical concatenation with index reallocation.

3.3 Data Preprocessing

Data preprocessing was conducted to improve data quality, eliminate redundancy, and prepare the network traffic data for machine learning analysis. Duplicate traffic records were removed to reduce repetitive patterns that could bias anomaly detection.

Missing numerical values were handled using mean-value imputation. Numerical columns containing null values were automatically identified and replaced with the mean of their corresponding features. Low-variance features that contributed minimal informational value were eliminated using variance threshold feature selection and the Variance Threshold is set to 0.0001.

3.4 Feature Engineering

To enhance temporal anomaly representation, a new derived feature named "packets_per_time_unit" was introduced. The feature was computed based on the time interval between consecutive packets.

First, the temporal difference between adjacent packet timestamps was calculated as shown in Equation 1:

$$\text{Time Delta} = \text{Time}_i - \text{Time}_{i-1} \quad \text{Eq (1)}$$

Subsequently, packet transmission intensity was estimated using the inverse temporal interval as shown in Equation 2:

$$\text{Packets\_Per\_Time\_Unit} = 1 / (\text{Time Delta} + 10^{-9}) \quad \text{Eq (2)}$$

A small epsilon constant was added to avoid division-by-zero instability. This engineered feature enabled the framework to capture burst-based traffic anomalies commonly associated with denial-of-service attacks, flooding attacks, and suspicious traffic spikes.

3.5 Data Normalization

Feature scaling was performed using the Min-Max normalization technique as shown in Equation 3.

$$X_{scaled} = (X - X_{min}) / (X_{max} - X_{min}) \quad \text{Eq (3)}$$

Normalization ensured consistent numerical ranges across all traffic features.

3.6 Isolation Forest-Based Anomaly Detection

The first stage of the proposed hybrid framework employed the Isolation Forest algorithm for unsupervised anomaly detection and is shown in Equations 4 and 5:

$$IF(X) = \text{Decision Function}(X) \quad \text{Eq (4)}$$

$$\text{Isolation Score} = -\text{Decision Function}(X) \quad \text{Eq (5)}$$

The Isolation Forest component effectively captured statistical irregularities within the network traffic data.

3.7 Bidirectional LSTM Autoencoder

The second stage of the framework utilized a deep Bidirectional Long Short-Term Memory (BiLSTM) Autoencoder for temporal anomaly detection. Network traffic sequences were generated using a sliding temporal window of 10 timesteps.

The encoder architecture consisted of stacked Bidirectional LSTM layers with dropout regularization and a bottleneck dense layer for compressed latent representation generation.

The model optimization process employed the Adaptive Moment Estimation (Adam) optimizer with Mean Squared Error (MSE) loss as shown in Equation 6.

$$\text{MSE} = (1/\text{n}) \sum_{i=1}^{n} (x_i - \hat{x}_i\, 1\,)^2 \quad \text{Eq (6)}$$

3.8 Reconstruction Error Computation

Following model training, the LSTM Autoencoder reconstructed the testing sequences as shown in Equation 7.

$$\text{Reconstruction Error} = (1/n)\, \Sigma\, (X - \hat{X})^2 \quad \text{Eq (7)}$$

Higher reconstruction errors indicated abnormal traffic patterns that deviated significantly from learned normal traffic behavior.

3.9 Hybrid Ensemble Anomaly Detection

To improve anomaly detection robustness, the anomaly scores generated by the Isolation Forest and the LSTM Autoencoder were combined using a weighted ensemble mechanism as shown in Equation 8.

$$\text{Hybrid Score} = 0.4(\text{IF Score}) + 0.6(\text{LSTM Score}) \quad \text{Eq (8)}$$

Anomalies were identified using the 95th percentile threshold of the hybrid score distribution as shown in Equation 9.

$$\text{Threshold} = P_{95}(\text{Hybrid Scores}) \quad \text{Eq (9)}$$

Traffic samples exceeding the threshold were classified as anomalous.

3.10 Model Evaluation and Visualization

Since the framework adopted an unsupervised learning approach, evaluation focused primarily on anomaly score analysis, traffic distribution visualization, and anomaly clustering inspection.

Visualization techniques employed include histogram analysis, scatter plots, temporal anomaly progression analysis, protocol distribution analysis, and packet length comparison between normal and anomalous traffic.

3.11 Explainable Artificial Intelligence (XAI)

To enhance interpretability and transparency of the proposed hybrid framework, SHapley Additive exPlanations (SHAP) and LIME (Local Interpretable Model-Agnostic Explanations) techniques were integrated into the anomaly analysis pipeline.

SHAP was employed to estimate the contribution of each traffic feature toward the hybrid anomaly score, while LIME generated local surrogate explanations for individual anomalous traffic instances. The interpretability framework enabled identification of the most influential network traffic attributes contributing to anomaly detection decisions.

4. RESULTS AND DISCUSSION

The experimental results obtained from the hybrid unsupervised anomaly detection system demonstrate the effectiveness of integrating a BiLSTM autoencoder with weighted ensemble anomaly scoring for identifying abnormal IoT network traffic patterns. The training performance of the BiLSTM autoencoder indicates that the model successfully learned the normal behavioural structure of the network traffic data through progressive reduction in reconstruction loss across the training epochs.

At the beginning of the training process, the model recorded a training loss of 0.0170 and a validation loss of 0.0033 during Epoch 1. These relatively higher values are expected at the early learning stage because the autoencoder had not yet fully captured the complex temporal relationships present within the network traffic sequences. However, as training progressed, both the training and validation losses decreased consistently, demonstrating stable convergence and effective feature learning. By Epoch 4, the validation loss had significantly reduced to 2.6774e-04, while the training loss also declined to 8.7552e-04. The best validation performance was achieved around epochs 8 and 9, where validation losses of 1.6704e-04 and 1.6253e-04 respectively were obtained. These extremely low loss values indicate that the model developed a strong capability for reconstructing legitimate network traffic patterns with minimal error. A final training loss of 3.9938e-04 confirms strong model generalisation and effective temporal representation learning. The training outcome therefore validates the suitability of BiLSTM autoencoders for modelling sequential IoT network traffic in unsupervised anomaly detection tasks.

After training, reconstruction errors were computed and transformed into BiLSTM anomaly scores, which were subsequently combined using a weighted ensemble strategy to produce the final hybrid anomaly scores. Figure 1 provides important insight into the statistical behaviour of the anomaly detection system. The histogram exhibits a heavily right-skewed distribution, where the majority of the network traffic samples possess low anomaly scores concentrated between 0.0 and 0.15. This indicates that most traffic instances were recognised as normal by the hybrid framework. However, the long tail extending towards anomaly scores above 0.4 and approaching 0.8 reveals the presence of significantly abnormal traffic behaviours. The sparse distribution of these high anomaly scores demonstrates that anomalous events are relatively rare but highly distinguishable from legitimate traffic patterns. This clear separation between normal and abnormal score distributions confirms the robustness of the hybrid anomaly scoring mechanism as shown in Figure 1.

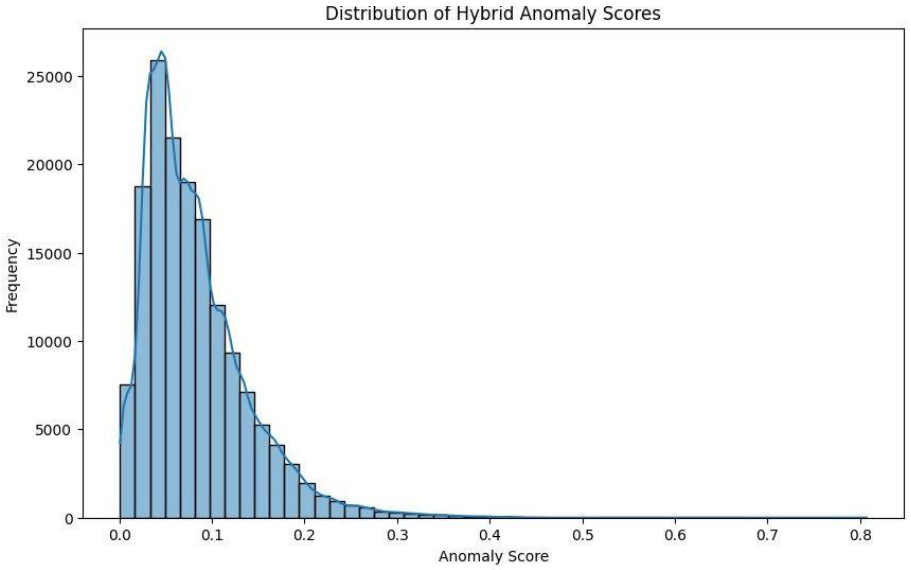


Fig. 1. Distribution of Hybrid Anomaly Scores.

Figure 2 further validates the discriminative ability of the proposed framework. Most data points are densely clustered around lower anomaly score regions represented by darker colours, indicating normal traffic behaviour. In contrast, several isolated points displayed brighter green and yellow colours corresponding to higher anomaly scores. These outlier points are concentrated mainly around scaled feature values between 0.5 and 0.7, with some anomaly scores exceeding 0.8. The visual separation between normal clusters and anomalous observations confirms that the hybrid framework successfully captured abnormal deviations in the feature space as shown in Figure 2. To classify anomalies automatically, the system applied a percentile-based thresholding approach using the 95th percentile of the hybrid anomaly score distribution. The calculated anomaly threshold value of 0.1891 implies that any network instance with a score exceeding this threshold was considered anomalous. Using this threshold, the framework detected 7,883 anomalies, representing approximately 5.00% of the total network traffic samples. This result aligns accurately with the selected thresholding strategy and confirms the consistency of the unsupervised anomaly detection process.

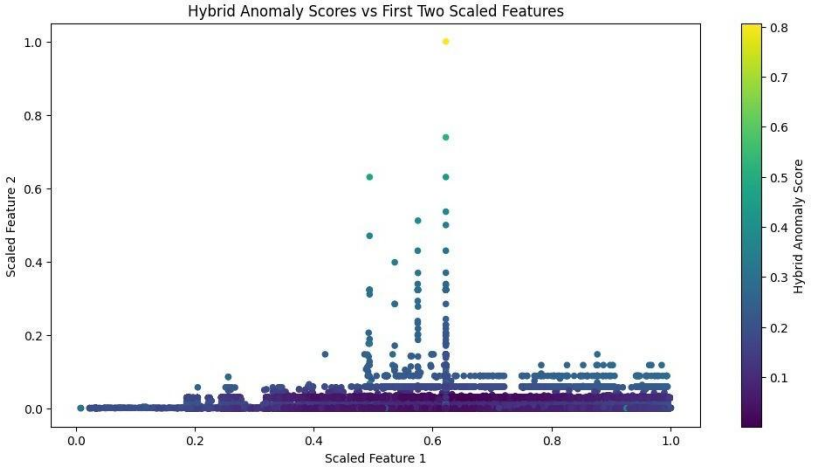


Fig. 2. Discrimination of Data Points

Similarly, Figure 3 demonstrates the temporal anomaly detection capability of the framework. The blue curve represents the variation of hybrid anomaly scores over time, while the red markers indicate detected anomalies. Most traffic samples maintained relatively low anomaly scores throughout the observation period, reflecting stable network behaviour.

Nevertheless, several distinct spikes were observed across different time intervals, particularly around the middle and later stages of the timeline, where anomaly scores increased beyond 0.6 and approached 0.8. These spikes indicate sudden abnormal network activities or suspicious traffic bursts. The repeated occurrence of detected anomalies across multiple

periods suggests that the framework is capable of identifying both isolated attacks and recurring malicious traffic behaviours as shown in Figure 3.

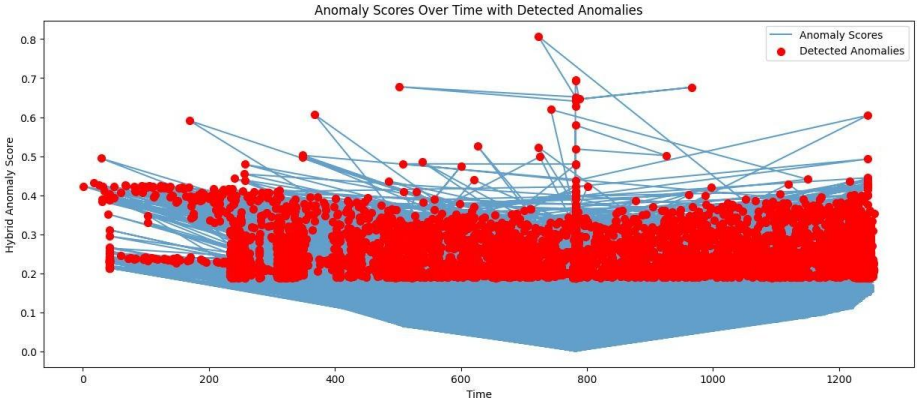


Fig. 3 Anomaly Scores with Detected Anomalies

To further justify the anomalies detected by the Isolation Forest score, BiLSTM autoencoder score, and the Hybrid Anomaly Score, Table 2 shows these scores. The top 10 Hybrid Anomaly Score values, which range from approximately 0.64 to 0.81, indicate varying degrees of abnormal behavior across the network traffic samples. Entries with high Isolation Forest scores, LSTM Autoencoder scores, and Hybrid Anomaly scores were successfully classified as anomalies, as shown by the "Detected_Anomaly" value of 1. The model effectively captured unusual packet transmission patterns, irregular packet timing, and suspicious communication between source and destination IP addresses. These findings suggest that combining Isolation Forest with LSTM Autoencoder improves anomaly detection capability by leveraging both statistical isolation and deep temporal pattern learning.

TABLE 2: TOP 10 DETECTED ANOMALOUS DATA POINTS (HYBRID SCORE)

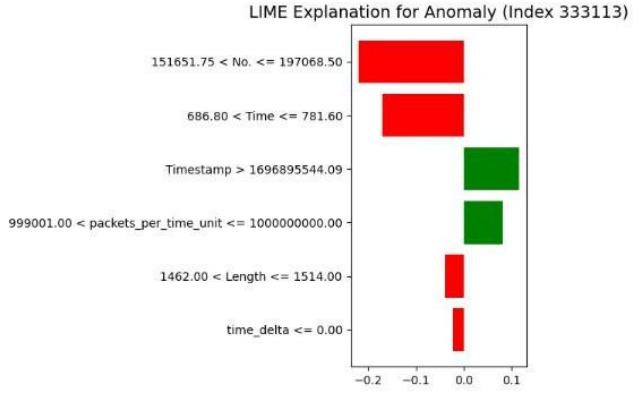
	Timestamp	Length	Time	No.	time_delta	packets_per_time_unit	Isolation_Forest_Score	LSTM_Autoencoder_Score	Hybrid_Anomaly_Score	Detected_Anomaly	Source	Destination
333113	1.696896e+09	1494.0	781.599378	197068.5	0.000000	1.000000e+09	0.517664	1.000000	0.807066	1	NaN	NaN
140595	1.696895e+09	1514.0	781.599378	197068.5	0.000000	1.000000e+09	0.255340	0.992917	0.697886	1	NaN	NaN
340986	1.696896e+09	1514.0	781.599378	197068.5	0.000000	1.000000e+09	0.252654	0.991475	0.695947	1	NaN	NaN
732220	1.696896e+09	1494.0	1106.669322	338085.0	0.000002	4.997501e+05	0.216441	0.987703	0.679199	1	188.156.172.107	192.167.7.162
781387	1.696896e+09	1514.0	1240.197733	387252.0	0.000002	4.997501e+05	0.197231	0.992366	0.674312	1	104.91.166.113	192.167.7.162
312744	1.696896e+09	54.0	781.599378	197068.5	0.000000	1.000000e+09	0.959990	0.454533	0.656716	1	NaN	NaN
70419	1.696895e+09	1514.0	781.599378	197068.5	0.000000	1.000000e+09	0.149615	0.986627	0.651822	1	NaN	NaN
282354	1.696896e+09	1514.0	781.599378	197068.5	0.000000	1.000000e+09	0.126232	0.993952	0.646864	1	NaN	NaN
559313	1.696896e+09	49745.0	722.691208	165178.0	0.000062	1.612877e+04	0.118019	0.992276	0.642573	1	192.167.7.162	35.186.194.58
65151	1.696895e+09	1370.0	781.599378	197068.5	0.000000	1.000000e+09	0.121740	0.984814	0.639584	1	NaN	NaN

For the interpretability of the hybrid model, the SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are used as explained. Across all evaluated instances, the feature "packets_per_time_unit" consistently recorded the highest SHAP importance values as shown in Figure 4. This indicates that packet transmission intensity is the most significant contributor to the model's anomaly detection decisions. In network security analysis, unusually high packet transmission rates are commonly associated with malicious activities such as Distributed Denial-of-Service (DDoS) attacks, flooding, or abnormal traffic bursts [13]. The strong positive contribution of this feature suggests that the model effectively captures suspicious traffic behaviour patterns and uses them as primary indicators of anomalies.

The features "No." and "Time" also demonstrated notable influence, although their SHAP values were negative.

Fig. 4. Explanations to SHAP importance values

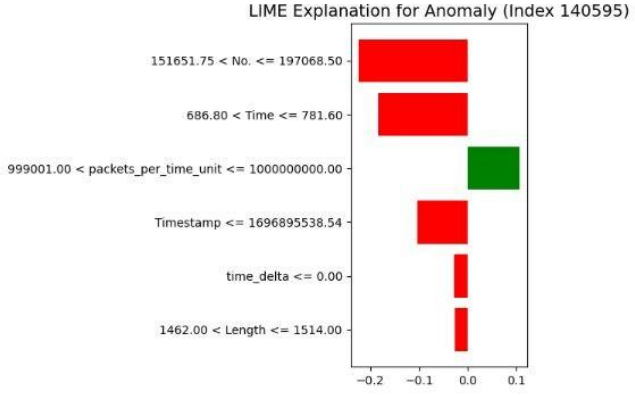
Furthermore, LIME identified "packets_per_time_unit" as a strong positive contributor toward anomaly detection. Specifically, high "packet-per-time-unit" ranges contributed positively to anomalous classifications, confirming the model's sensitivity to traffic spikes. Additionally, LIME revealed that specific ranges of "No." and "Time" strongly contributed negatively to predictions, reinforcing the conclusion that these ranges are associated with benign traffic behavior as shown in Figures 5, 6 and 7.




```
--- Feature Importance Comparison for Instance (Index 333113) ---

SHAP Feature Importance (Absolute Value, Top Features):
packets_per_time_unit    0.092927
No.                     -0.065056
Time                    -0.041444
time_delta               0.001602
Length                   0.000318
Timestamp                0.000000
```

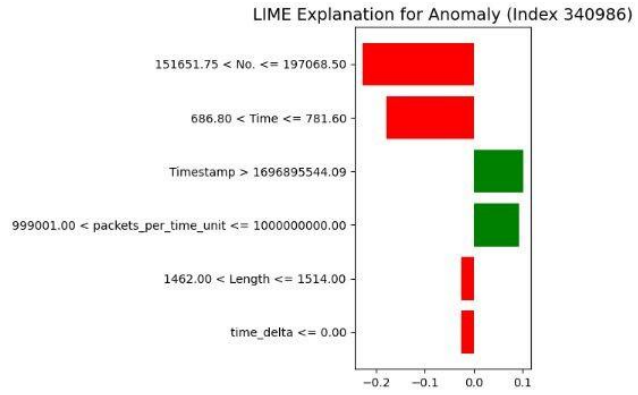
Fig. 5. LIME Explanation for Anomalous Instance (Index 333113)




```
--- Feature Importance Comparison for Instance (Index 140595) ---

SHAP Feature Importance (Absolute Value, Top Features):
packets_per_time_unit    0.084361
No.                     -0.069877
Time                    -0.046398
Length                  -0.017439
time_delta              -0.003461
Timestamp                0.000000
```

Fig. 6. LIME Explanation for Anomalous Instance (Index 140595)




```
--- Feature Importance Comparison for Instance (Index 340986) ---

SHAP Feature Importance (Absolute Value, Top Features):
packets_per_time_unit    0.092206
No.                     -0.062031
Time                    -0.038553
Length                  -0.009594
time_delta               0.004384
Timestamp                0.000000
```

Fig. 7. LIME Explanation for Anomalous Instance (Index 340986)

LIME explains predictions using interpretable feature intervals such as "999001.00 < packets_per_time_unit <=1000000000.00." These interval-based explanations are particularly valuable in cybersecurity applications because they provide operationally actionable insights for security analysts. Such thresholds can be incorporated into real-time intrusion detection rules or monitoring systems to improve early anomaly identification.

The proposed hybrid anomaly detection framework has significant practical importance in cybersecurity, particularly for securing IoT environments against sophisticated cyberattacks. The study revealed that the feature "packets_per_time_unit" was the most influential indicator of anomalous behaviour, demonstrating that unusual packet transmission intensity strongly correlates with attacks such as DDoS flooding, malicious traffic bursts, and unauthorized communications.

Features such as "No." and "Time" also contributed to distinguishing legitimate from suspicious activities. By combining Isolation Forest and BiLSTM autoencoder techniques with SHAP and LIME explainability, the framework improves early threat detection, enhances analyst trust, supports real-time intrusion prevention, and strengthens intelligent network defence systems.

5. CONCLUSION

This study presented a hybrid unsupervised anomaly detection framework that integrates Isolation Forest and BiLSTM autoencoder techniques with explainable artificial intelligence methods for IoT network traffic analysis. The experimental results demonstrated that the proposed framework effectively learned the normal behavioural patterns of network traffic through progressive reduction in reconstruction loss during training. The low validation and training losses obtained confirmed the capability of the BiLSTM autoencoder to model complex temporal relationships within IoT traffic data. Furthermore, the weighted hybrid anomaly scoring mechanism successfully distinguished normal and abnormal traffic patterns, producing a clear separation between legitimate network behaviour and anomalous activities.

The anomaly detection framework effectively identified suspicious traffic instances, including abnormal packet transmission patterns and irregular communication behaviours. The percentile-based thresholding strategy also provided consistent anomaly classification, enabling the detection of significant malicious traffic activities within the dataset.

Additionally, the integration of SHAP and LIME enhanced the interpretability and transparency of the model by identifying "packets_per_time_unit" as the most influential feature contributing to anomaly detection. This finding highlights the importance of packet transmission intensity as a major indicator of cyber threats such as Distributed Denial-of-Service attacks and abnormal traffic flooding in IoT environments.

For future work, the framework can be extended by incorporating advanced deep learning architectures such as Transformer networks, Graph Neural Networks, or attention-based models to further improve temporal and spatial anomaly detection performance. Future studies may also evaluate the framework using more IoT datasets containing diverse attack categories and heterogeneous device behaviours.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgment

None.

References

[1] T. Mzili, M. Mzili, S. I. Bouderba, A. Abatal, W. Aribowo, and A. K. Arya, "Interoperability in Internet of Things: Taxonomies and open challenges," *Babylonian Journal of Internet of Things*, vol. 2025, pp. 101–112, 2025.

[2] S. Das and S. Namasudra, "Introducing the Internet of Things: Fundamentals, challenges, and applications," *Advances in Computers*, vol. 137, pp. 1–36, 2025.

[3] S. A. Moamin, M. K. Abdulhameed, R. M. Al-Amri, A. D. Radhi, R. K. Naser, and L. G. Pheng, "Artificial intelligence in malware and network intrusion detection: A comprehensive survey of techniques, datasets, challenges, and future directions," *Babylonian Journal of Artificial Intelligence*, vol. 2025, pp. 77–98, 2025.

[4] B. H. Mohammed, H. Sallehudin, N. S. M. Satar, H. D. Murhg, S. A. Mohamed, F. A. Alaba, *et al*., "Anomaly detection of distributed denial of service (DDoS) in IoT network using machine learning," in *Digital Technologies and Transformation in Business, Industry and Organizations*, vol. 3. Cham, Switzerland: Springer Nature Switzerland, 2025, pp. 41–64.

[5] D. Quirumbay Yagual, D. Fernández Iglesias, and F. J. Nóvoa, "A hybrid deep learning-based architecture for network traffic anomaly detection via EFMS-enhanced KMeans clustering and CNN-GRU models," *Applied Sciences*, vol. 15, no. 20, p. 10889, 2025.

[6] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," *Sensors*, vol. 23, no. 13, p. 5941, 2023.

[7] K. O. Adefemi, M. B. Mutanga, and O. A. Alimi, "A hybrid CNN–GRU deep learning model for IoT network intrusion detection," *Journal of Sensor and Actuator Networks*, vol. 14, no. 5, p. 96, 2025.

[8] H. Kamal and M. Mashaly, "Shared autoencoder-based unified intrusion detection across heterogeneous datasets for binary and multi-class classification using a hybrid CNN–DNN model," *Machine Learning and Knowledge Extraction*, vol. 8, no. 2, p. 53, 2026.

[9] S. A. Hussein and S. R. Répás, "A hybrid intrusion detection framework using deep autoencoder and machine learning models," *AI*, vol. 7, no. 2, p. 39, 2026.

[10] V. Nayak, S. Pawar, and S. K. BL, "An empirical study of Hy-BiLSTM-Tr and AE-GRU deep learning models for intrusion detection," *Computers and Electrical Engineering*, vol. 134, p. 111108, 2026.

[11] A. S. Alhumaima, O. S. Hameed, and H. Alkattan, "Comprehensive analysis and anomaly detection of network traffic using Isolation Forest modeling," *Babylonian Journal of Internet of Things*, vol. 2026, pp. 25–33, 2026.

[12] R. K. Gattu, "Network traffic dataset," *Kaggle*, 2020. [Online]. Available: https://www.kaggle.com/datasets/ravikumargattu/network-traffic-dataset. Accessed: May 2026.

[13] Q. Yan, F. R. Yu, Q. Gong, and J. Li, "Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 602–622, 2015.