JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 3, June 2026, hlm. 437 – 444 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i3.4652
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

FPR-CONSTRAINED HYBRID DEEP LEARNING FOR IOT ANOMALY DETECTION

Salma Nurkamila[1], Suprih Widodo[1*]
[1]Pendidikan Sistem dan Teknologi Informasi, Univeritas Pendidikan Indonesia
email: *supri@upi.edu

Abstract: Existing IoT anomaly detection studies have achieved high classification performance, but most focus on accuracy and F1-score without explicitly controlling the false positive rate (FPR). In addition, many approaches rely on a single detection perspective, limiting their operational reliability. To address this gap, this study proposes a hybrid anomaly detection framework integrating Long Short-Term Memory (LSTM), Shannon entropy, and autoencoder reconstruction error. Shannon entropy is incorporated as an additional feature, while LSTM and the autoencoder capture temporal and reconstruction characteristics. The resulting hybrid representation is processed by a constraint-based threshold selection mechanism that enforces FPR ≤ 0.10. Experiments on the TON-IoT and Edge-IIoTset datasets achieved average F1-scores of 0.9250 and 0.9934, while maintaining average FPR values of 0.0091 and 0.0714, respectively. Analysis of entropy distributions showed consistent differences between normal and anomalous traffic across both datasets, indicating that Shannon entropy provides discriminative information for anomaly detection. These results demonstrate strong detection performance with controlled false alarms, while ablation studies confirm the significant contribution of Shannon entropy to overall model performance.

Keywords: false positive rate; hybrid deep learning; Internet of Things; network anomaly detection; Shannon entropy

Abstrak: Penelitian deteksi anomali *Internet of Things* (IoT) telah menunjukkan performa klasifikasi yang tinggi, namun sebagian besar masih berfokus pada *accuracy* dan F1-score tanpa mengendalikan *false positive rate* (FPR) secara eksplisit. Selain itu, banyak pendekatan hanya memanfaatkan satu perspektif deteksi sehingga reliabilitas operasionalnya masih terbatas. Untuk mengatasi kesenjangan tersebut, penelitian ini mengusulkan kerangka deteksi anomali hybrid yang mengintegrasikan *Long Short-Term Memory* (LSTM), *Shannon entropy*, dan *autoencoder reconstruction error*. *Shannon entropy* digunakan sebagai fitur tambahan, sedangkan LSTM dan autoencoder menangkap karakteristik temporal dan deviasi rekonstruksi. Representasi *hybrid* yang dihasilkan kemudian diproses melalui mekanisme *constraint-based threshold selection* dengan batas FPR $\leq 0{,}10$. Hasil pengujian pada dataset TON-IoT dan Edge-IIoTset menghasilkan F1-score rata-rata sebesar 0,9250 dan 0,9934, dengan FPR rata-rata sebesar 0,0091 dan 0,0714. Perbedaan nilai *entropy* yang konsisten antara trafik normal dan anomali pada kedua dataset menunjukkan bahwa *Shannon entropy* menyediakan informasi diskriminatif untuk deteksi anomali. Hasil tersebut menunjukkan performa deteksi yang kuat dengan *false alar*m yang terkendali, sementara studi ablasi mengonfirmasi kontribusi signifikan *Shannon entropy* terhadap performa model.

Kata kunci: deteksi anomali jaringan; false positive rate; hybrid deep learning; Internet of Things; Shannon entropy

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 3, June 2026, hlm. 437 – 444 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i3.4652
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

The proliferation of *Internet of Things* (IoT) devices across smart cities, industrial automation, healthcare, and critical infrastructures has increased network complexity and exposure to cyberattacks such as Distributed Denial of Service (DDoS), Man-in-the-Middle (MITM), data injection attacks [1], [2], [3]. Conventional anomaly detection approaches remain limited in detecting zero-day attacks and adapting to traffic distribution shift [4]. Moreover, IoT datasets often exhibit class imbalance and distributional variations that complicate anomaly detection [5], [6], [7].

Numerous studies have applied machine learning and deep learning approaches for IoT anomaly detection. Kaya et al. and Zamanzadeh Darban et al. [8] demonstrated that deep learning models are effective for capturing temporal anomaly patterns, but their evaluations primarily focused on classification performance without explicit false alarm control. In distribution-based detection, Pandey and Mishra [9] utilized entropy-based features to identify traffic distribution irregularities, yet the approach did not integrate temporal characteristics. Similarly, autoencoder-based methods proposed by Katbi and Ksantin [10] and Salehiyan et al. [11]. Relied on reconstruction error analysis but focused on reconstruction characteristics alone.

Consequently, existing studies generally employ a single detection perspective and rarely incorporate explicit FPR constraints, limiting their operational reliability in normal-traffic-dominated IoT environments. As a result, it remains unclear whether temporal, distributional, and reconstruction characteristics can be effectively integrated while simultaneously maintaining false alarm rates within an operationally acceptable range.

Despite these studies showing high detection performance, evaluations are still focused on classification metrics such as accuracy, precision, recall, and F1-score, without explicitly imposing FPR as an operational constraint. In addition, existing approaches generally leverage only a single detection perspective, such as temporal characteristics in LSTM-based models [4], [8], distributional characteristics in entropy-based approaches [9], [12], or reconstruction characteristics in autoencoder-based approaches [10], [11]. As a result, the complex nature of IoT traffic anomalies is not yet comprehensively represented [2], [8]. Furthermore, models with high classification performance do not necessarily achieve acceptable false alarm rates in practical IoT environments dominated by normal traffic [13], [14]. Given these limitations, there remains a need to develop anomaly detection frameworks that simultaneously capture temporal, distributional, and reconstruction characteristics while explicitly controlling false alarms through a structured threshold selection mechanism.

To address this gap, this study proposes a hybrid anomaly detection framework combining LSTM, Shannon entropy, and autoencoder reconstruction error. LSTM captures temporal dependencies, Shannon entropy represents local distributional changes, and the autoencoder quantifies reconstruction deviation [8], [9], [10]. Their integration produces a unified hybrid anomaly score that is subsequently processed by a constraint-based threshold selection mechanism as the final decision layer.

Unlike previous studies that focused on optimizing accuracy or F1-score, this study applied a constraint-based threshold selection mechanism

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 3, June 2026, hlm. 437 – 444 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i3.4652
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

with an explicit objective of an operational FPR ≤ 0.10. The FPR ≤ 0.10 constraint was applied during validation to limit false alarms while maintaining a balance between detection capability and operational reliability [14]. This threshold was selected based on the base-rate fallacy principle proposed by Axelsson [13], which states that a small increase in FPR can result in a very large number of false alarms in environments dominated by normal traffic. Therefore, more relaxed constraints, such as FPR ≤ 0.50 or FPR ≤ 0.70, were not considered because they could substantially reduce the practical reliability of the detection system [13], [14]. This approach is also aligned with the need to control the trade-off between detection capability and false alarm, as discussed by Liu et al. [14] and Sørbø and Ruocco [15]. The framework was evaluated on the TON-IoT [5] and Edge-IIoTset [7] datasets through 30 independent multi-seed experiments to assess model stability under random initialization.

Based on the identified research gap, this study makes three contributions. First, it proposes a hybrid anomaly detection framework integrating temporal, distributional, and reconstruction characteristics through LSTM, Shannon entropy, and an autoencoder. Second, it incorporates a constraint-based threshold selection mechanism with an operational FPR limit of 0.10. Third, it evaluates model stability and component contributions through multi seed experiments, ablation studies, and Wilcoxon Signed Rank Tests on the TON IoT and Edge IIoTset datasets.

METHOD

This research is an experimental, computation-based study aimed at designing and evaluating a hybrid anomaly detection framework for IoT networks. This study uses IoT network traffic data obtained from two public benchmark datasets: TON-IoT [5] and Edge-IIoTset [7]. The TON-IoT dataset, developed by the University of New South Wales, consists of 211,043 samples with 44 features covering normal traffic and 7 attack categories, while the Edge-IIoTset dataset consists of 257,800 samples with 63 features and 14 attack categories. Both datasets were selected because they represent different IoT traffic distribution characteristics, allowing for cross-dataset generalizability evaluation [6], [8].

Experiments were conducted using Python with TensorFlow, Keras, and scikit-learn. Independent variables include original features, Shannon entropy, and autoencoder reconstruction error, while the dependent variable is the binary classification label. All preprocessing was performed exclusively on the training set to avoid data leakage, including the removal of irrelevant features, handling missing values with zeros, label encoding for categorical features, and Min-Max normalization to the range [0,1] [5], [7]. The dataset was divided using stratified sampling into a 70:20:10 split: 70% training, 20% validation, and 10% test. Sliding windows and entropy calculations were carried out after separation to avoid temporal leakage.

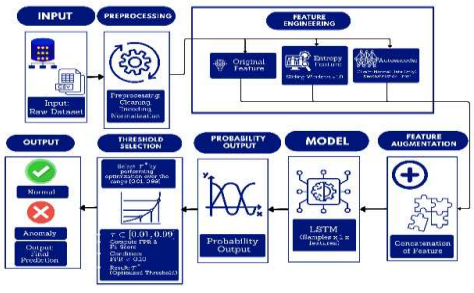

Figure 1. Proposed Hybrid IoT Anomaly Detection Framework

The feature engineering approach

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 3, June 2026, hlm. 437 – 444 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i3.4652
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

in this study involves two complementary stages of additional feature extraction. Shannon entropy is calculated to quantify the uncertainty of network traffic distribution locally within sliding windows of size 10 [9], [12], as given by equation (1).

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \quad (1)$$

The value $p(x_i)$ represents the probability of a feature occurrence in a given window. The Shannon entropy value quantifies the uncertainty in the traffic distribution and is used to distinguish between normal traffic and anomalous traffic [9]. Autoencoder reconstruction error is calculated using the Mean Squared Error (MSE) between the input and the autoencoder's output. A symmetric fully connected autoencoder was used to compute reconstruction error. As shown in equation (2) below:

$$RE(x) = \frac{1}{d}\sum_{j=1}^{d}(x_j - \hat{x}_j)^2 \quad (2)$$

The variable d denotes the input feature dimension [10]. The entropy and autoencoder reconstruction error features are then combined with the original features through feature augmentation. The final input dimension of the model consists of 21 features for the TON-IoT and 13 features for Edge-IIoTset. Sequence representation employed a sliding window of size 10, producing LSTM inputs of shape $(N \times 10 \times F)$, where N denotes the number of samples and F the number of features [8].

The model architecture consists of two LSTM layers [4]. The first layer consists of 64 LSTM units, followed by Batch Normalization and Dropout for regularization, a second LSTM layer with 32 units, and a Dense layer with sigmoid activation for binary classification. Training uses the Binary Cross-Entropy loss function and the Adam optimizer. Class imbalance is handled using class weights based on the class distribution in the training set, while early stopping based on validation loss is used to prevent overfitting. The ablation studies were conducted using three configurations: w/o Entropy, w/o Autoencoder, and w/o Constraint. Each configuration was evaluated through 30 multi-seed experiments with identical data splits to ensure evaluation consistency [8].

Statistical significance was assessed using the Wilcoxon Signed-Rank Test to compare the F1-score distributions across 30 multi-seed ablation experiments. This nonparametric test was chosen because it does not assume normality in deep learning model performance [8]. Constraint-based threshold selection was applied to control FPR [13], [14]. Threshold candidates τ were evaluated on the validation set over the range $[0.01 - 0.99]$ with an interval of 0.01. Thresholds satisfying FPR ≤ 0.10 were retained, and the optimal threshold τ^* that maximized the F1-score was selected, as formulated in equation (3) below:

$$\tau^* = argmax_{\{\tau \in \mathcal{T}, FPR(\tau) \leq 0.10\}} F1(\tau) \quad (3)$$

The optimal threshold τ^* obtained from the validation set is consistently applied to the test set for final evaluation to avoid data leakage. In addition, performance evaluation was conducted using four main metrics: F1-score, precision, recall, and FPR, as formulated in equations (4), (5), (6), (7) below [10], [15]:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 3, June 2026, hlm. 437 – 444 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i3.4652
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (7)$$

TP, FP, FN, and TN represent true positives, false positives, false negatives, and true negatives, respectively [10], [13]. Model stability against random initialization variation was evaluated across 30 experiments using identical data and different seeds. Evaluation results are reported as mean ± standard deviation to provide a more statistically reliable performance estimate [6].

RESULT AND DISCUSSION

Table 1. Single Experiment Evaluation Results on Both Datasets

Metric	TON-IoT	Edge-IIoTset
F1-score	0.9469	0.9964
Recall	0.9015	1.0000
Precision	0.9971	0.9929
FPR	0.0084	0.0395
Optimal threshold ($\tau*$)	0.63	0.48

Edge-IIoTset produced a higher F1-score and recall, whereas TON-IoT produced lower FPR and higher precision. These differences reflect the distinct distribution characteristics of the two datasets [5], [7]. Both datasets satisfied the FPR ≤ 0.10 constraint, confirming the effectiveness of the proposed threshold selection mechanism in controlling false alarms while preserving anomaly detection capability [14]. [13].

Table 2. Distribution of Shannon Entropy and Reconstruction Error

Parameter	TON-IoT	Edge-IIoTset
Normal Entropy	1.8240	0.5275
Anomaly Entropy	1.5045	0.3818
Entropy Difference (Δ)	0.3194	0.1457
Normal Error	0.000013	0.010370
Anomaly Error	0.000120	0.010960
Anomaly/ Normal Ratio	9.2×	1.06×

In both datasets, normal traffic exhibits higher Shannon entropy than anomalous traffic [9], [12]. The larger entropy difference in TON-IoT ($\Delta = 0.3194$ compared to $\Delta = 0.1457$) indicates strong discriminative power, consistent with the findings of Pandey and Mishra [9]. Meanwhile, the separability of the autoencoder reconstruction error is much higher in TON-IoT with a ratio of 9.2×. In contrast, the ratio in Edge-IIoTset is 1.06×, indicating that the autoencoder contributes more actively to anomaly discrimination in TON-IoT, whereas its contribution in Edge-IIoTset is relatively limited owing to the smaller separation between normal and anomalous reconstruction errors [10]. These consistent entropy differences across both datasets indicate that Shannon entropy provides discriminative information for separating normal and anomalous traffic distributions. To further evaluate model robustness, threshold stability, and the contribution of individual components, additional analyses were conducted through multi-seed experiments and ablation studies.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. XII No 3, June 2026, hlm. 437 – 444 ISSN 2550-0201 (Online)
DOI: http://dx.doi.org/ 10.33330/jurteksi.v12i3.4652
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 3. Threshold, Ablation, and Multi-Seed Stability Characteristics

Parameter	TON-IoT	Edge-IIoTset
Mean F1 ± Std	0.9250 ± 0.0141	0.9934 ± 0.0077
Mean Recall	0.8633	0.9996
Mean FPR	0.0091	0.0714
Seeds with FPR > 0.10	0/30	5/30
Highest FPR	0.0468 (Seed 8)	0.4591 (Seed 8)
Lowest FPR	0.0020 (Seed 28)	0.0370 (Seed 13)
Threshold Range	0.65 – 0.93	0.20 – 0.85
F1 w/o Entropy	0.8277	0.9879
FPR w/o Entropy	0.0227	0.1323
F1 w/o Autoencoder	0.9198	0.9931
FPR w/o Autoencoder	0.0275	0.0671
F1 w/o Constraint	0.8380	0.9954
FPR w/o Constraint	0.9139	0.0464

The multi-seed results show that TON-IoT has better constraint stability than Edge-IIoTset. All experiments on TON-IoT met the FPR ≤ 0.10 constraint, ranging from $0.0020 - 0.0468$, while Edge-IIoTset exhibits greater variability, with five seeds violating the constraint and one extreme case reaching an FPR of 0.4591. The ablation study confirms that the entropy feature provides the most consistent contribution to model performance on both datasets. Removing entropy caused the largest F1-score reduction and increased variability, confirming its contribution to both performance and stability. Removing the constraint mechanism resulted in substantially higher FPR, whereas removing the autoencoder caused only a moderate performance decline [10].

Wilcoxon Signed-Rank Test results showed that removing entropy significantly reduced F1-score on TON-IoT ($p = 1.9 \times 10^{-9}$) and Edge-IIoTset ($p = 0.0006$). In contrast, removing the autoencoder did not produce significant differences on either dataset. Although removing the constraint mechanism did not significantly affect F1-score on Edge-IIoTset ($p = 0.1718$), it eliminated the operational FPR boundary, potentially reducing practical reliability under varying traffic distributions [13], [14].

Multi-seed analysis also shows the trade-off between recall and FPR across datasets. TON-IoT produces a higher threshold in the range of $0.65 - 0.93$ with a low recall of 0.8633 and tightly controlled FPR of 0.0091, while Edge-IIoTset achieves a recall of 0.9996 but with more variable FPR up to 0.459. These findings indicate that constraint effectiveness is influenced by dataset distribution characteristics [6], [15].

CONCLUSION

This study successfully demonstrated that constraint-based threshold selection can maintain false alarm rates within a predefined operational boundary while preserving strong anomaly detection performance across different IoT traffic distributions. The findings further confirm that integrating temporal, distributional, and reconstruction characteristics improves detection reliability, with Shannon entropy contributing significantly to both performance and stability. Future work could develop adaptive thresholding mechanisms that dynamically adjust FPR limits to account for distribu-

tional shifts and evaluate model performance in more complex real-time IoT environments.

BIBLIOGRAPHY

[1] B. Rathi *et al.*, "Realizing the potential of Internet of Things (IoT) in Industrial applications," Dec. 01, 2025, *Springer Nature*. doi: 10.1007/s43926-025-00141-5.

[2] Z. A. Haider *et al.*, "A Survey on anomaly detection in IoT: Techniques, challenges, and opportunities with the integration of 6G," Oct. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.comnet.2025.111484.

[3] A. Amara Korba, A. Diaf, M. A. Bouchiha, and Y. Ghamri-Doudane, "Mitigating IoT botnet attacks: An early-stage explainable network-based anomaly detection approach," *Comput. Commun.*, vol. 241, Sep. 2025, doi: 10.1016/j.comcom.2025.108270.

[4] M. O. Kaya, M. Ozdem, and R. Das, "A new hybrid approach combining GCN and LSTM for real-time anomaly detection from dynamic computer network data," *Computer Networks*, vol. 268, Aug. 2025, doi: 10.1016/j.comnet.2025.111372.

[5] A. Alsaedi, N. Moustafa, Z. Tari, A. Mahmood, and Adna N Anwar, "TON-IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems," *IEEE Access*, vol. 8, pp. 165130–165150, 2020, doi: 10.1109/ACCESS.2020.3022862.

[6] S. Ismail, S. Dandan, and A. Qushou, "Intrusion Detection in IoT and IIoT: Comparing Lightweight Machine Learning Techniques Using TON_IoT, WUSTL-IIOT-2021, and EdgeIIoTset Datasets," *IEEE Access*, vol. 13, pp. 73468–73485, 2025, doi: 10.1109/ACCESS.2025.3554083.

[7] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning," *IEEE Access*, vol. 10, pp. 40281–40306, 2022, doi: 10.1109/ACCESS.2022.3165809.

[8] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep Learning for Time Series Anomaly Detection: A Survey," *ACM Comput. Surv.*, vol. 57, no. 1, Oct. 2024, doi: 10.1145/3691338.

[9] N. Pandey and P. K. Mishra, "Conditional entropy-based hybrid DDoS detection model for IoT networks," *Comput. Secur.*, vol. 150, Mar. 2025, doi: 10.1016/j.cose.2024.104199.

[10] A. Katbi and R. Ksantini, "One-class IoT anomaly detection system using an improved interpolated deep SVDD autoencoder with adversarial regularizer," *Digital Signal Processing: A Review Journal*, vol. 162, Jul. 2025, doi: 10.1016/j.dsp.2025.105153.

[11] A. Salehiyan, P. S. Moghaddam, and M. Kaveh, "An Optimized Transformer–GAN–AE for Intrusion Detection in Edge and IIoT Systems: Experimental Insights from WUSTL-IIoT-2021, EdgeIIoTset, and TON_IoT Datasets," *Future Internet*, vol. 17, no. 7, Jul. 2025, doi: 10.3390/fi17070279.

[12] C. E. Shannon, "A Mathematical Theory of Communication," 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

[13] S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," 2000. doi: 10.1145/357830.357849.

[14] B. Liu, Z. Zhang, S. Hu, S. Sun, D. Liu, and Z. Qiu, "A Security Trade-Off Scheme of Anomaly Detection System in IoT to Defend against Data-Tampering Attacks," *Computers, Materials and Continua*, vol. 78, no. 3, pp. 4049–4069, 2024, doi: 10.32604/cmc.2024.048099.

[15] S. Sørbø and M. Ruocco, "Navigating the metric maze: a taxonomy of evaluation metrics for anomaly detection in time series," *Data Min. Knowl. Discov.*, vol. 38, no. 3, pp. 1027–1068, May 2024, doi: 10.1007/s10618-023-00988-8.