International Journal of Medical Toxicology & Forensic Medicine
2026, Volume 16, E53130

Research Paper

Modeling of Transformer-Based Autoencoder Architecture for Secure IoT-Based Intrusion Detection in Critical Healthcare Systems

Saranya Ramachandran, Murugan Ambigapathi*

Department of Data Science and Business Systems, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu - 603203, India.

Citation Saranya R, Murugan A. Modeling of Transformer-Based Autoencoder Architecture for Secure IoT-Based Intrusion Detection in Critical Healthcare Systems. International Journal of Medical Toxicology and Forensic Medicine. 2026; 16:E53130.

https://doi.org/10.22037/ijmtfm.v16.53130

Article info:

Received: 08 August, 2026

Accepted: 28 August, 2026

Published: 16 September, 2026

Keywords:

Internet of Things, Healthcare security, Intrusion detection system, Transformer-based autoencoder, Intensive care unit, Bayesian optimization

ABSTRACT

Background: The Internet of Things (IoT) is widely used in healthcare for patient monitoring and continuous data collection. In intensive care unit (ICU) environments, many connected medical devices are used to monitor patient conditions in real time. Existing intrusion detection approaches using machine learning (ML) and deep learning (DL) have shown good results, but many still struggle with high-dimensional, complex network traffic data. Therefore, a more effective intrusion detection approach is required for secure IoT-based healthcare systems. The model is designed to distinguish between normal and malicious traffic in healthcare systems using deep learning approaches.

Methods: In the proposed model, relevant features are selected using a statistical and information-theoretic approach based on mutual information, variance thresholding, and correlation analysis to remove redundant information. For classification, a transformer-based autoencoder learns meaningful representations from network traffic, and a multilayer perceptron classifies the representations into normal and malicious classes. To improve model performance, Bayesian Optimization is leveraged for hyperparameter tuning.

Results: The proposed TAE-SHID model was evaluated using a benchmark dataset. The model achieved an accuracy of 96.86%, showing good classification performance compared with existing methods.

Conclusion: The results show that the proposed TAE-SHID model is effective for classification and outperforms existing methods.

*** Corresponding Author:**

Murugan Ambigapathi, PhD

Department of Data Science and Business Systems, School of Computing, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur Campus, Tamil Nadu - 603203, India.

E-mail: *murugana@srmist.edu.in*

Introduction

The Internet of Things (IoT) is a foundational system that is the origin of many future apps in the areas of smart manufacturing, transportation, and health care systems. IoT relies on several sensors to collect data about devices, the surrounding environment, and humans [1]. This data is frequently distributed to the cloud server, and accordingly, application administrators can make many choices to enhance the efficacy of apps. Healthcare is one of the main applications of IoT, in which people are provided with wearable devices to collect data on vital signs [2]. In many scenarios, mistaken choices based on malicious information could lead to the downfall of IoT-enabled healthcare systems. To overcome these difficulties, the development of Artificial Intelligence (AI)-based methodologies has played a key part in increasing the safety of IoT. They enable an adaptive security architecture capable of detecting and responding to IoT attacks in real-world settings [3]. IoT devices can enhance security by using machine learning (ML) to learn from past attack patterns and identify abnormal activities. Hierarchical deep learning (DL) models can effectively detect cyberattacks by automatically learning useful patterns from complex healthcare network data in real-time cloud environments [4].

A brief overview of existing studies related to IoT healthcare security and intrusion detection methods is presented. Musthafa et al. [5] developed a continuous health-monitoring architecture to protect IoT-based models by incorporating lightweight cryptography with DL approaches. Kateb et al. [6] designed an improved security model for human-focused schemes that leverages DL with the jellyfish search optimization technique for IoT healthcare systems. The presented methodology enables IoT devices in healthcare domains to effectively carry clinical information and perform initial identification of health issues in the human-machine connection. Naik et al. [7] examined a new trust-aware hybrid architecture to study temporal, latent, and spatial features of physiological indications. Arthi et al. [8] introduced an intelligent SDN-empowered secure architecture for the IoT healthcare ecosystem. Mohammadi et al. [9] proposed an innovative CNN-based methodology for identifying cyber threats in MIoT environments. Unlike prior research that primarily relied on standard ML techniques or simpler DNNs, the presented method leverages CNNs' ability to efficiently analyze temporal features in network traffic data. Kumar et al. [10] proposed a Deep Gated Recurrent Unit (D-GRU) as an AI-enabled architecture for forecasting attack onset in MIoT systems in critical care units (CCUs). In [11], the combination of ML and IoT methodologies in healthcare sectors has been developed as a transformative approach.

Recent Transformer-based intrusion detection methods have demonstrated strong representation-learning capability; however, limited studies have explored their combination with autoencoder-based latent representation learning and systematic Bayesian hyperparameter optimization for high-dimensional healthcare IoT traffic. To address this gap, the proposed framework combines Transformer-based contextual feature learning, autoencoder-based compact representation, and Bayesian optimization for automated hyperparameter selection, providing an optimized approach for intrusion detection in healthcare IoT environments. This paper proposes a Transformer-Based Autoencoder for Secure Healthcare Intrusion Detection (TAE-SHID). The major contributions of this study are as follows:

- Employ a Statistical and Information-Theoretic Selection (SITS) method to select relevant features through mutual information (MI), variance thresholding, and correlation analysis.

- Integrate a transformer-based autoencoder (TAE) with a multi-layer perceptron (MLP) for capturing complex dependencies in healthcare network traffic and performing intrusion detection.

- Utilize Bayesian Optimization (BO) for hyperparameter tuning to enhance the performance of the proposed model.

Materials and Methods

Figure 1 demonstrates the overall workflow of the proposed system. The IoT healthcare dataset is primarily preprocessed using data cleaning, encoding, outlier removal, and feature scaling to enhance data quality. Relevant features are then chosen using statistical and information-based techniques to retain only the most useful attributes. The refined features are provided to a transformer-based autoencoder (TAE) for effective pattern learning. Subsequently, a multi-layer perceptron (MLP) performs classification to differentiate different traffic categories. Bayesian optimization is further used to fine-tune model parameters and enhance performance. Finally, the optimized model classifies network traffic into normal and attack categories.

Preprocessing

Data preprocessing was applied to clean, organize, and refine the data to improve model performance (Figure 1) [12].

Feature Selection Method

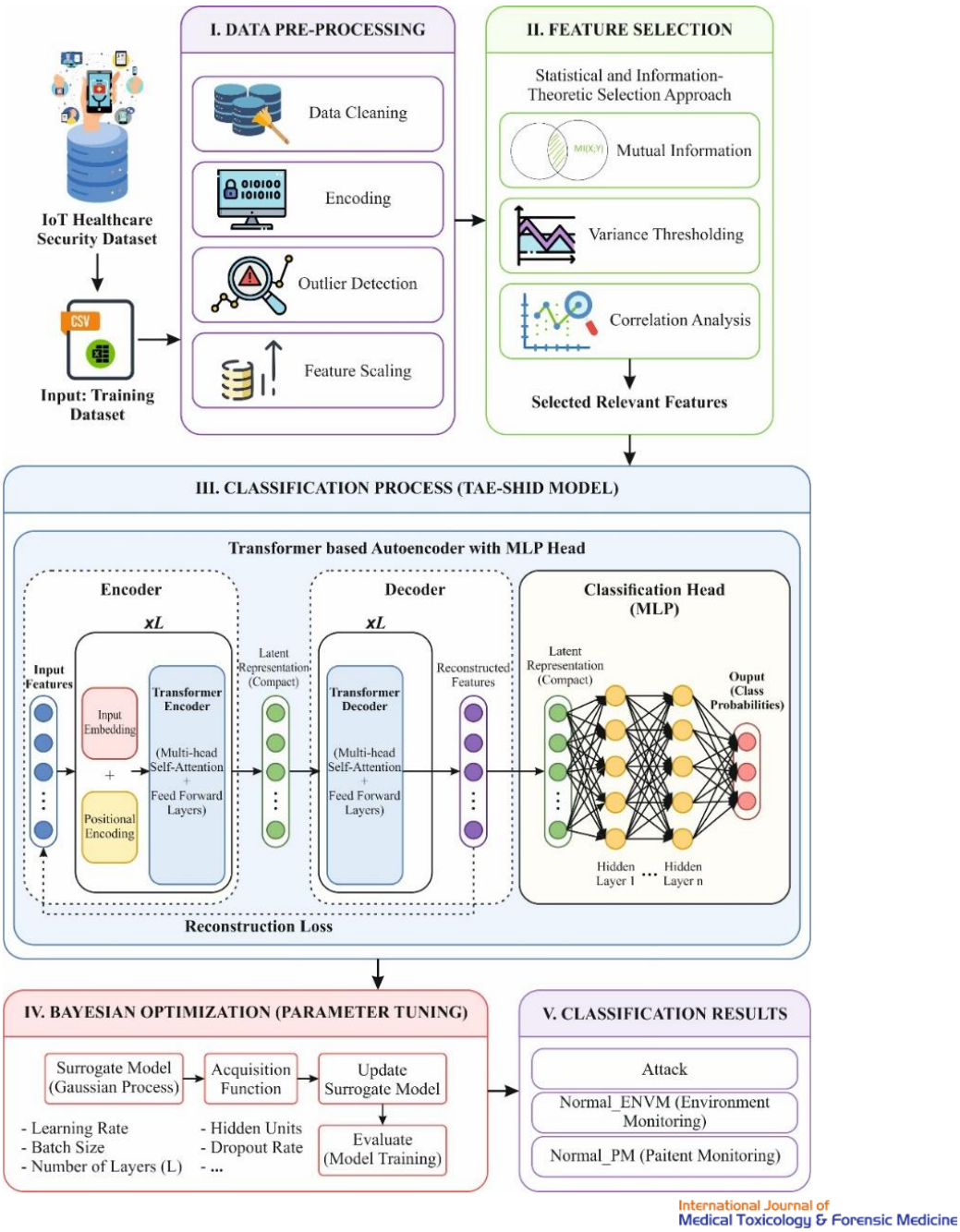
The FS procedure plays an important role in attack detection by selecting relevant attributes from the data, thereby enhancing accuracy and performance. This proposed model utilizes a SITS, ensuring strong and effective FS through different methods [13]. The SITS method uses three techniques: MI, variance thresholding, and correlation analysis.

MI

An initial phase includes computing the MI between every variable and the dependent feature. It evaluates the data obtained about one feature across another. In X are the features, and the Y implies a dependable feature; the MI $I(X;\ Y)$ has been computed.

$$I(X;Y) = \sum x \in x \sum p\,(x,y) \log\,(\frac{p(x,y)}{p(x)p(y)}) \quad (1)$$

Here, $p(x,y)$infers X and Y joint probability, while $p(x)$and $p(y)$ implies their individual probability



International Journal of
Medical Toxicology & Forensic Medicine

Figure 1. Overall workflow of the proposed model.

distributions. Features with high MI values are retained because they have a stronger relationship with the target feature.

Variance thresholding

Afterward, this method is used to eliminate features with low variance that provide little information for distinguishing among classes. The σ^2 implies that the variance of a feature X has been represented in Eq. (2).

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (2)$$

Let N denote the count of instances, x_i implies the feature value for the i-th instance, and μ Signifies the feature mean. We dropped variables with low variance.

Correlation analysis

It is used to classify the extremely correlated attributes. The coefficient of Pearson correlation r among dual features X and Y is computed in Eq. (3).

$$r(X,Y) = \frac{\sum_{i=1}^{N}(x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_X)^2 \sum_{i=1}^{N}(y_i - \mu_Y)^2}} \quad (3)$$

Hybrid Deep Learning-Based Classification Approach

For the attack classification in an IoT-based ICU healthcare network, a transformer-based autoencoder (TAE) is combined with MLP classification heads. The transformer encoder uses multi-head self-attention to learn non-linear relations among network traffic features and preserves meaningful details via layer normalization and residual connections [14].

Encoder

The encoder receives the selected feature set and converts it into a compact latent module through acquiring the relationships between features by self-attention mechanisms.

Decoder

The decoder reconstructs the actual input features from the latent module. The variation between the original input and the reconstructed output is measured by a reconstruction loss, which helps the model learn informative feature representations. The reconstruction loss is calculated as:

$$L_{rec} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2 \quad (4)$$

whereas N is the input feature counts. The latent representation ZIt is then forwarded to the MLP classifier to predict the traffic category. The classification loss has been computed using the cross-entropy function:

$$L_{cls} = -y\log\sum(\hat{y}) \quad (5)$$

here y implies the true label and $\hat{y}$is the predicted label. Finally, both losses are combined during training:

$$L_{total} = L_{cls} + \lambda L_{rec} \quad (6)$$

MLP classification head

The latent representation produced by the transformer-based autoencoder is passed into an MLP head for final classification. The MLP head consists of two dense layers with ReLU activation, followed by a softmax output layer for multi-class attack prediction. This module maps the learned latent embeddings into various attack and benign classes in the IoT-based ICU healthcare dataset. The latent representation z produced by the transformer-based autoencoder is passed to the MLP classification head.

$$h_1 = \text{ReLU}(W_1 z + b_1) \quad (7)$$

$$h_2 = \text{ReLU}(W_2 h_1 + b_2) \quad (8)$$

$$\hat{y} = \text{Softmax}(W_3 h_2 + b_3) \quad (9)$$

Here, the latent representation obtained from the encoder is denoted as z, the predicted probability distribution over multiple attack classes is represented as $\hat{y}$, the trainable weight parameters are W_1, $W2, W_3$, and the bias vectors are b_1, b_2, b_3.

The ReLU activation function enables non-linearity, whereas the softmax layer performs the final step of multi-class classification.

Hyperparameter Tuning using Bayesian Optimization

Hyperparameter tuning is employed to determine which model performs better on a validation set: the Transformer-based autoencoder or an MLP classifier [15].

$$x^* = \arg\min_{c\epsilon x} f(x) \quad (10)$$

Where $f(x)$ defines the objective function to minimize, for instance, the validation set error rate, $x^{\star}$ refers to the optimal set of hyperparameters that attains the best model performance, and c indicates any value in the domain of x. Employing constant hyperparameter optimization, including grid and random search, improves model performance compared to physical inspection.

Results

This section investigates the performance of the proposed TAE-SHID method on the IoT Healthcare Security Dataset [16] from the Kaggle repository (https://www.kaggle.com/datasets/faisalmalik/iot-healthcare-security-dataset). This dataset simulates an ICU environment in which medical IoT devices generate network traffic data. The dataset comprises 50

Table 1. Dataset Details.

Class	No. of Samples
Attack	80126
Normal_ENVM (Environment Monitoring)	31758
Normal_PM (Patient Monitoring)	76810
Total Samples	**188694**

International Journal of
Medical Toxicology & Forensic Medicine

Table 2. Attack Detection Result Of The TAE-SHID approach with a 70:30 split.

Classes	$Accur_y$ (%)	$Preci_n$ (%)	$Recal_l$ (%)	F_{score} (%)	MCC
TRAP (70%)					
Attack	95.52	94.67	94.79	94.73	90.84
Normal_ENVM	98.47	94.60	96.39	95.49	94.58
Normal_PM	95.97	95.47	94.61	95.04	91.66
Average	**96.66**	**94.92**	**95.26**	**95.09**	**92.36**
TESP (30%)					
Attack	95.84	94.96	95.26	95.11	91.49
Normal_ENVM	98.60	94.99	96.84	95.91	95.07
Normal_PM	96.15	95.78	94.69	95.23	92.00
Average	**96.86**	**95.24**	**95.60**	**95.42**	**92.86**

International Journal of
Medical Toxicology & Forensic Medicine

features; only 24 must be selected for further analysis. Table 1 shows that the dataset comprises three classes, totaling 188,694 samples. Table 2 presents the attack detection results of the TAE-SHID approach using a 70:30 data split.

Discussion

Table 3 presents a comparison of the TAE-SHID method with existing approaches across different measures [17-19]. The comparative study against these models provides a comprehensive evaluation across complexity, computational efficiency, feature extraction capabilities, and sequence-learning mechanisms, enabling a robust assessment of the proposed method's performance. Based on CT, the proposed TAE-SHID system achieves a lower CT value of 2.85sec and a higher accuracy of 96.86%. Although the remaining models, such as DNN, CNN + LSTM, Two Stage-LSTM, BiLSTM, Deep LSTM, DexCRNN, SqueezeNet, ShuffleNet, InceptionV3, VGG-16, and ResNet-50, achieved reasonable performance, they did not surpass the proposed TAE-SHID technique in classification accuracy or computational efficiency (Table 3).

Table 3. Comparison result of the TAE-SHID approach with other systems.

Model	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	CT (sec)
DNN [19]	93.74	93.71	93.82	93.47	6.3
CNN + LSTM [19]	92.90	90.59	91.10	91.55	4.8
Two Stage-LSTM [19]	91.27	91.65	90.05	90.24	5.3
BiLSTM [17]	92.50	92.00	93.00	92.00	5.3
Deep LSTM [17]	95.00	93.00	94.00	93.00	6.1
DexCRNN [17]	95.80	95.00	95.00	95.00	7.2
SqueezeNet [18]	93.47	94.34	78.26	86.76	4.4
ShuffleNet [18]	94.72	94.08	82.68	89.30	6.9
Inceptionv3 [18]	94.89	94.17	82.28	89.11	3.7
VGG-16 [18]	95.38	93.86	84.29	90.14	5.5
ResNet-50 [18]	95.62	93.64	85.70	90.84	3.1
TAE-SHID	96.86	95.24	95.60	95.42	2.8

International Journal of
Medical Toxicology & Forensic Medicine

The Transformer-based autoencoder can provide a more effective representation of network traffic than conventional deep learning models because its self-attention mechanism captures long-range dependencies and relationships among high-dimensional network features. The autoencoder further learns compact latent representations by reconstructing the input data, thereby reducing redundancy and preserving informative patterns. This combination enables TAE-SHID to learn more discriminative representations of normal and malicious traffic, while Bayesian optimization further improves the model by identifying suitable hyperparameter configurations.

Conclusion

In this paper, the TAE-SHID model is presented to improve attack detection performance in IoT-enabled healthcare environments. Initially, data preprocessing was performed through data cleaning, encoding, outlier detection, and feature scaling to improve the dataset's

quality and consistency. Additionally, a SITS approach was used to select relevant features using MI, variance thresholding, and correlation analysis, thereby reducing redundancy and retaining informative attributes. For classification, the proposed TAE-SHID model employed a transformer-based autoencoder to capture complex dependencies among network traffic features and generate compact latent representations. Moreover, an MLP was incorporated as the classification head. Bayesian Optimization was integrated for hyperparameter tuning to enhance model performance. Experimental evaluation on the benchmark dataset shows that the proposed TAE-SHID approach achieves superior classification performance compared to existing techniques. In the future, the model can be tested on larger real-world healthcare datasets to assess its robustness and practical performance. Explainable AI and advanced DL techniques can also be examined to improve prediction accuracy and model interpretability.

Ethical Approval

None.

Acknowledgment

None.

Funding

No external funding was received for this research.

Conflicts of Interest

The authors report there are no competing interests to declare.

References

[1] Hezam M, Ismail MM, Ali AM, Sallam K, Abdel-Basset M. DP-DL-ZT: a zero-trust-enhanced differential privacy framework with CNN-LSTM for cyber threat detection in IoT healthcare. Neural Comput Appl. 2026;38(8):289. [DOI: 10.1007/s00521-026-11890-x]

[2] Priyadarshani PP, Nayak J, Dash PB. Optimizer influenced deep neural network for IoT healthcare security. SN Comput Sci. 2026;7(1):118. [DOI: 10.1007/s42979-025-04657-z]

[3] Alfakeeh S. A blockchain-enabled IoT framework for secure attack detection and advanced feature selection in smart healthcare. Eng Technol Appl Sci Res. 2025;15(5):28219-23. [DOI: 10.48084/etasr.13349]

[4] Ali TE, Zoltan AD. Hierarchical deep learning for robust cybersecurity in multi-cloud healthcare infrastructures. Eng Technol Appl Sci Res. 2025;15(1):20358-66. [DOI: 10.48084/etasr.8918]

[5] Musthafa S, Jenifa G, Murugan L, Moorthy U. IoT based security enhanced continuous health monitoring framework using lightweight cryptography and deep learning techniques. Int J Comput Intell Syst. 2026. [DOI: 10.1007/s44196-026-01237-8]

[6] Kateb F, Ragab M, Abukhodair F, Abdulkader OA, Maghrabi LA, Binyamin SS, et al. Improved security for IoT-based remote healthcare systems using deep learning with jellyfish search optimization algorithm. Sci Rep. 2025;15(1):13223. [DOI: 10.1038/s41598-025-97065-5]

[7] Naik N, Surendranath N, Raju SAB, Madduri C, Dasari N, Shukla VK, et al. Hybrid deep learning-enabled framework for enhancing security, data integrity, and operational performance in Healthcare Internet of Things (H-IoT) environments. Sci Rep. 2025;15(1):31039. [DOI: 10.1038/s41598-025-15292-2]

[8] Arthi R, Krishnaveni S, Zeadally S. An intelligent SDN-IoT enabled intrusion detection system for healthcare systems using a hybrid deep learning and machine learning approach. China Commun. 2024;21(10):1-21. [DOI: 10.23919/JCC.ja.2022-0681]

[9] Mohammadi H, Ghahramani S, Asghari SA, Aminian M. Securing healthcare with deep learning: a CNN-based model for medical IoT threat detection. In: 2024 19th Iranian Conference on Intelligent Systems (ICIS). IEEE; 2024. p. 168-73. [DOI: 10.1109/ICIS64839.2024.10887510]

[10] Kumar PM, Kavin BP, Jagathpally A, Shahwar T. Transforming the cybersecurity space of healthcare IoT devices using deep learning. In: 2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC). IEEE; 2025. p. 1-6. [DOI: 10.1109/ICAIC63015.2025.10849305]

[11] Das D, Ghosh U, Shetty S, Mohanty S.

Machine learning-enabled IoT applications for smart healthcare monitoring systems. Adv Comput. 2025;137:175-218. [DOI: 10.1016/bs.adcom.2024.06.007]

[12] García S, Ramírez-Gallego S, Luengo J, Benítez JM, Herrera F. Big data preprocessing: methods and prospects. Big Data Anal. 2016;1(1):9. [DOI: 10.1186/s41044-016-0014-0]

[13] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003;3:1157-82. [Link]

[14] Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE; 2013. p. 6645-7. [DOI: 10.1109/ICASSP.2013.6638947]

[15] Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst. 2012;25.

[16] Faisal Malik. IoT healthcare security dataset [Internet]. Kaggle. [Link]

[17] Darwish R, Abdelsalam M, Khorsandroo S. Deep learning-based XIoT malware analysis: a comprehensive survey, taxonomy, and research challenges. J Netw Comput Appl. 2025;242:104258. [DOI: 10.1016/j.jnca.2025.104258]

[18] Khan SH, Alahmadi TJ, Ullah W, Iqbal J, Rahim A, Alkahtani HK, et al. A new deep boosted CNN and ensemble learning based IoT malware detection. Comput Secur. 2023;133:103385. [DOI: 10.1016/j.cose.2023.103385]

[19] Hossain MA. Deep learning-based intrusion detection for IoT networks: a scalable and efficient approach. EURASIP J Inf Secur. 2025;2025(1):28. [DOI: 10.1186/s13635-025-00202-w]