SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

IoT Malware Detection and Classification using Number Theory based Secure Cryptographic Deep Learning Framework

[1]Balachandra Kumaraswamy*, [2]M Vasantha Lakshmi, [3]Vinay Joshi, [4]Ms. Preeti Hinge

[1]Professor, B.M.S. College of Engineering, Bangalore. E-mail: balachandrak.tce@bmsce.ac.in, ORCID: 0000-0002-1946-1697
[2]Associate Professor, BMS College of Engineering, Bangalore, Karnataka. E-mail: vasantha.tce@bmsce.ac.in, ORCID: 0000-0002-2114-5554
[3]Founder, Infin AIverse Pvt. Ltd., Jaipur, Rajasthan, India. E-mail: vinay.joshiii@gmail.com
[4]Assistant Professor in Mathematics, Shri Shankaracharya College of Information, Professional and Management, Raipur, Chhattisgarh, India. E-mail: preetibhoskar@gmail.com
*Corresponding mail: balachandrak.tce@bmsce.ac.in

Abstract
The rapid development of Internet of Things (IoT) Networks has made the Internet of Things devices more vulnerable to high-end malwares and thus there is a demand for the precise malware detection with efficient data security. In this research, a number-theory based secure cryptographic deep-learning system is proposed for the detection and classification of IoT Malware threats. The proposed method includes the following four process of duplication removal, treatment of missing values with median, normalization of Min-Max, Autoencoder Based Feature Denoising (AEFD), Principal Component Analysis (PCA) and Transformer-based classification. Another cryptographic mechanism based on a number theory is also embedded to provide secure communications and to optimize the management of the session keys. It can be seen from the experimental results that AEFD achieves a reconstruction loss of 0.020 in only 30 epochs, which is reduced by 76.5% and the first 6 principal components still explain 89.7% of the total variance. The Transformer's training accuracy is 97.8% and its validation accuracy is 96.9%, while the proposed classifier obtains 98.0%, 97.5%, 98.2%, 97.9% and 98.4% in terms of accuracy, Precision, Recall, F1-score and ROC-AUC, respectively. Moreover, cryptographic operations are only 3.53 ms, showing that the framework is able to provide strong malware detection and secure communication but with a small computation overhead, which makes it suitable for resource-constrained IoT environments.

Keywords: IoT Malware Detection; Deep Learning; Transformer; Auto encoder-Based Feature Denoising; Number-Theory Cryptography; PCA-Based Feature Reduction.

1. Introduction

Internet of Things (IoT) is a disrupting technology where physical things are able to collect process and communicate information using an interconnected network [1]. IoT systems can be found in smart homes, healthcare, industrial automation, transportation, agriculture, smart cities and critical infrastructure. While these systems offer greater automation and intelligent decision-making, the rapid emergence of these technologies has added a tremendous amount to the cybersecurity attack surface [2]. Due to small processing power, memory, storage space and battery life, many IoT devices are hard to secure with high computational load security methods [3]. Moreover, the weak authentication, outdated communication protocols, insufficient firmware and security updates make IoT devices easy to be attacked by malware. A compromised device can be used to form a botnet and to perform distributed denial-of-service if needed, or to steal sensitive information or disrupt critical operations [4].

Malware attacks for IoT devices have become more complex and are equipped with a wide range of features to evade traditional security tools including obfuscation, polymorphism, automated propagation and multi-stage attacks [5]. Traditional signature based detection mechanisms are limited in success to meet emerging and novel threats, because these mechanisms primarily rely on the predefined malware signatures [6]. There are some of these limitations that can be mitigated by behaviour-based methods that focus on recognizing unusual network traffic and device activity. With this in mind, one can notice that machine learning and deep learning have become popular for automatic detection of IoT malware [7]. There is potential for deep learning models to learn complex relationships looking at large scale cybersecurity data and detect malicious patterns that might not be picked up by pre-defined rules. However, the majority of the existing deep-learning based techniques mainly emphasize the accuracy of the detection and offer limited protection for sensitive data that is used during model training and analysis [8].

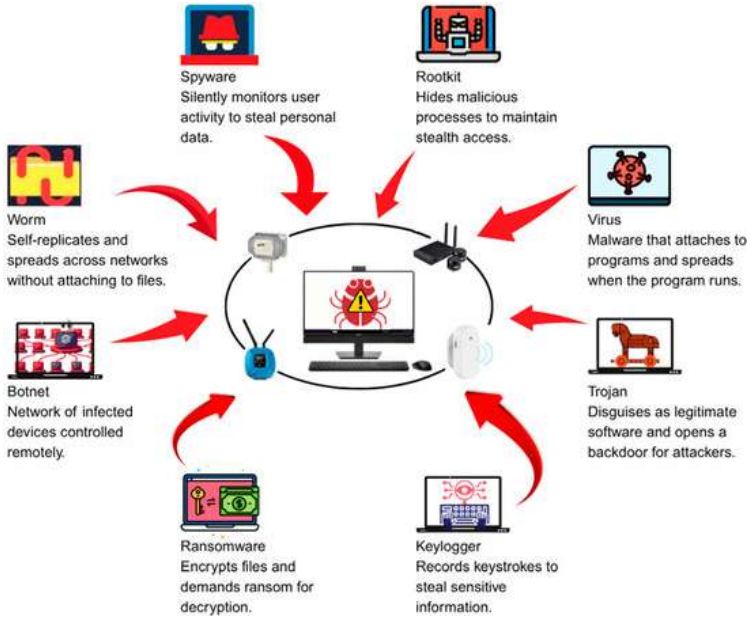

Figure 1: Major Types of IoT Malware Threats and Their Impact on Connected Devices [9]

The security of data in IoT systems is crucial as the communication flows and activities of the devices can include sensitive operational data [10]. The security system can be rendered unreliable if such information becomes available to someone else or is altered. Cryptographic methods offer a good way of safeguarding information against unauthorized access and manipulation. But traditional cryptographic algorithms may consume a lot of computing and memory resources, a characteristic that is hard to achieve on IoT devices with resource limitations [11]. Accordingly, lightweight security mechanisms are needed to offer effective security without imposing too much of a burden on system efficiency. The concepts of prime numbers, modular arithmetic, Euler's theorem and finite mathematical structures are well suited for the mathematical foundations of cryptographic design and are very applicable in number theory [12]. Therefore, the number-theory-based cryptographic mechanisms can be used to ensure secure and efficient data protection in limited resource IoT devices.

Although there is progress in study for detecting IoT malware and cryptographic security is gradually improving, these areas are generally tackled independently [13]. A deep learning model can be able to achieve high accuracy of detection but lack proper protection of sensitive data, while a cryptographic technique can be secure but be unable to identify malicious activities in an intelligent manner [14]. Additionally, IoT datasets for malware can be redundant in features, imbalanced in classes and have complex behaviours.Moreover, the IoT malware dataset can have also redundant features, classes imbalance and complex behaviours that can affect the performances of classification. Thus, a comprehensive solution is needed which integrates secure data processing with intelligent malware detection and classification [15].

To tackle these, this study is aimed at proposing an IoT Malware Detection and Classification using Number Theory based Secure Cryptographic Deep Learning Framework. The suggested framework combines number-theory-based cryptographic protection along with deep learning for intelligent and secure analysis of malwares. The goal of the framework is to shield the sensitive information in IoT systems and facilitate precise detection and classification of malicious activities. Data preprocessing and feature optimization are integrated to enhance input data quality and secure cryptographic processing and deep-learning detection/classification are integrated. The proposed solution aims for an effective compromise between security, detection accuracy, capability of classification and computational efficiency. The aims of this work are:

1. To develop a number theory-based lightweight cryptographic mechanism for securing sensitive IoT data and communication while minimizing computational and memory overhead.
2. To design a deep learning-based malware detection model capable of distinguishing benign IoT activities from malicious behaviours using relevant network and behavioural features.
3. To develop an effective malware classification mechanism for identifying and differentiating multiple IoT malware families based on their behavioural and communication patterns.
4. To evaluate the proposed framework using accuracy, precision, recall, F1-score, detection rate, false-positive rate, computational time, memory consumption and cryptographic overhead.

It is expected that the proposed study would make significant contributions in the area of secure, intelligent and efficient cybersecurity in heterogeneous IoT environments. Incorporating mathematical cryptography and deep learning into the same framework can support data protection and malware intelligence in a single architecture, offering a basis for enhanced protection against the emerging malicious attacks in IoT.

2. Literature Review

In recent years, researchers have been integrating different technologies and software, such as AI, blockchain, federated learning and cryptography, to enhance the detection of malicious activity and secure communication in the IoT ecosystem. For image-based IoT malware detection, Sirhan et al., (2026) [16] introduced a new concept named Mal-Fedchain, which combines Federated Learning, blockchain, honeypot monitoring, IMGAN preprocessing and image-based residual capsule network. The centralized model achieved 93.52% accuracy, 92.40% precision, 93.52% recall, 92.52% F-measure, MCC of 0.9245 and AUC of 0.9976, while the federated model achieved 65.62% accuracy and 0.9840 AUC. Ahmad et al., (2026) [17] proposed a Weighted Symmetric Hashed Blockchain framework with feature weighting deep learning, AES-256-GCM encryption and permissioned blockchain. On HIKARI-2021 dataset, WSHB-DNN attained macro-F1 score of 0.6837 and on the ADFA-LD dataset, a macro-F1 score of 0.7914. Gunasridharan et al., (2026) [18] have created a quantum security framework based on VQC, QSVC, quantum neural network and BB84 QKD. The accuracy of detection at QSVC was 97.75%, which demonstrates the power of quantum cybersecurity.

Recently, lightweight and privacy preserving security for resource-constrained IoT system has also been singled out. To tackle this challenge, Mutambik et al., (2025) [19] introduced TRIM-SEC, a fusion of autoencoder denoising, PCA, Transformer-based classification and lightweight ECC optimized via PSO. The performance of the framework on accuracy of detection, number of false alarms and encryption latency outperformed the LSTM, CNN-GRU and blockchain-based methods. In the lines of this, Kumari et al., (2025) [20] presented the polar-code cryptography and quantum self-attention framework with the accuracy, precision and error of 99.5%, 96% and 5% respectively. In this way, Safeer et al., (2025) [21] have presented an RNN-Bi-LSTM model for detecting malware in IoMT, which reported an accuracy of 98.38%. Khan et al., (2024) [22] have proposed three approaches: DeepShield, BoTShield and MalwareShield that combine homomorphic encryption, secret sharing, differential privacy and federated learning. These methods aided in the secure inference, the detection of malware variants and zero days and decrease communication needs.

Previous study proved that the increasing importance of the combination of machine learning, cryptography, blockchain and intelligent intrusion detection. Baazeem et al., (2024) [23] presented a CNN-based security model with Q-WOA which gave an accuracy of 0.965. To prevent an adaptive attack, Hamarsheh et al., (2024) [24] introduced a hybrid approach of combining SDN, encrypted channels, random forests, traffic monitoring and reauthentication to develop SCAFFOLD. Devi et al., 2023 [25] presented a deep LSTM with Improved ECC and MA-BW optimization where the proposed model obtained an accuracy of 95%, precision of 92% and 5% error, with an execution time at ECC of 6.02 s. Khan et al., (2023) [26] introduced DSBEL using deep boosted CNN and ensemble learning, achieving 98.50% accuracy, 98.42% precision, 95.97% recall, 97.12% F1-score and 91.91% MCC. A fuzzy-CNN-based IDS was proposed by Santhosh Kumar et al., (2023) [27] to enhance the accuracy of recognizing DoS attacks and minimize the false positive rate. In general, the studies demonstrate great advancement in intelligent IoT security, but there were still difficulties in federated performance, computational efficiency, privacy, zero-day detection and real-time deployment.

While remarkable progress has been made in the field of detecting IoT malware, current solutions recognize that existing solutions cannot provide secure, lightweight, privacy-preserving and real-time protection in the same timeframe. Deep learning models were highly accurate in detecting but consume significant computational power, restricting its use for resource-constrained IoT devices. While Federated Learning offers benefits in terms of data privacy, it can also lead to potential performance issues and communication overhead and Blockchain can provide greater integrity and traceability, it also requires more storage and processing power. Also, a large number of the current techniques were based on attacks that were known and offer little resistance for zero-day and evolving attacks. Hence, it was important to have a lightweight framework that not only detects intelligently but also preserves privacy, ensures secure communications and facilitates efficient deployment of IoT in real-time.

3. Research Methology

This proposed framework brings together IoT traffic analysis, data preprocessing, Deep learning based Malware classification and secure Cryptography based on number-theory in a single Security Pipeline. The methodology starts by preparing network traffic data, eliminating noise and redundant information, extracting discriminative features, classifies malware and finally defenses sensitive information using optimized cryptographic key generation. Figure 2 shows the overall framework of the proposed methodology.

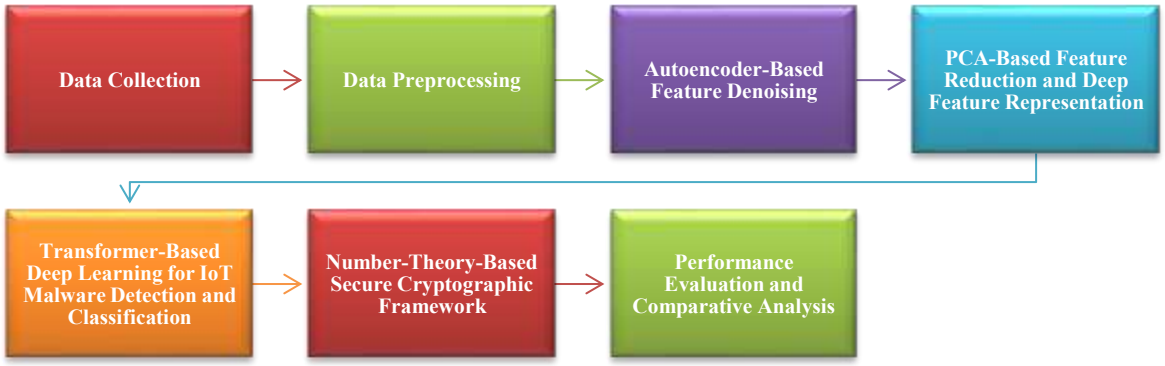

Figure 2: Framework of methodology

3.1Data Collection

The proposed study adopts the IoT Botnets Attack Detection Dataset [28] which is obtained from the N-BaIoT dataset for the detection and classification of IoT malware. The data were from five IoT devices: Danmini Doorbell, Ecobee Thermostat, Philips Baby Monitor, Provision Security Camera and Samsung Webcam. Malicious traffic corresponding to different botnet behaviors is incorporated into the dataset. The acquired data were displayed in the shape of

$$D = \{(X_i, Y_i)\}_{i=1}^{N}, \quad (1)$$

where N denotes the total traffic records, X_i represents the feature vector and Yi denotes the corresponding class label. Each record contains network traffic characteristics used for malware classification.

3.2 Data Preprocessing

Network traffic data from the raw IoT network may contain redundant data, missing values and features can have different numerical scales. Hence, there is a need to improve the data quality by preprocessing it before extracting features and classifies those using deep learning methods. Firstly duplicate observations were eliminated to prevent bias. Empty numeric cells were filled in with the median value of the feature. Lastly, Min-Max normalization is used to normalise the numerical features to a standard range. The purpose of these preprocessing steps is to improve the consistency of data and to avoid the features having a high numerical value overtaking the deep-learning model.

- **Removal of Duplicate Records**

Traffic observations were duplicated to avoid having duplicate samples that might compromise the training. Records were called duplicates if feature values were the same. This step can help minimize data redundancy and enhance the reliability of doing future malware detection and classification.

$$X_i = X_j, \quad i \neq j \quad (2)$$

where X_i and X_j represent two identical traffic records.

- **Missing-Value Treatment**

In the event that the numerical values of one of the features were missing, the negative values were replaced with the median from which the positive values were detected. This results in a preserved central tendency of the traffic data with a reduced impact of extreme observations. The resulting dataset may then be used successfully for further normalization and deep learning computation.

$$X_{ij}^{*} = \begin{cases} X_{ij}, & X_{ij} \neq NULL \\ Median(X_j), & X_{ij} = NULL \end{cases} \quad (3)$$

where X_j represents the j^{th} feature.

- **Min-Max Normalization**

Normalization is a technique to convert numerical features into a standard range of 0 to 1 using the Min Max technique. This helps to maintain consistent feature scales and avoid having network properties without much magnitude information drowning the deep-learning model. The normalized features give stable inputs to the AEFD and subsequent Transformer based classification stages.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

The normalized feature satisfies:

$$0 \leq X' \leq 1 \quad (5)$$

This normalization ensures comparable feature scales and prevents high-magnitude attributes from dominating the deep-learning model.

3.3Autoencoder-Based Feature Denoising

- **Encoder Architecture**

The features of the preprocessed IoT network traffic were further compressed to a small latent representation using the Encoder of the Autoencoder-Based Feature Denoising (AEFD) module. Assume X ∈ R^(N×d) represents the normalized input feature matrix with N number of traffic samples and d number of input features. The encoder successively performs nonlinear transformations so as to learn the most informative features of IoT traffic, while eliminating irrelevant variations and noise. The encoded representation is represented as:

$$Z = f_\theta(X) \quad (6)$$

where Z denotes the latent feature representation and f_θ represents the encoder parameterized by θ. The resulting latent features were subsequently used for dimensionality reduction and deep-learning-based malware classification.

- **Decoder Architecture**

The decoder restores the transformer's initial IoT traffic characteristics from the generated latent representation.The decoder recovers the original IoT traffic features from the restored latent representation from the transformer's training. Using this reconstruction process, the AEFD model is able to learn the network traffic structure and minimize the impact of irrelevant and noisy data. The decoder can be represented as:

$$\hat{X} = g_\emptyset(Z) \quad (7)$$

where Z represents the encoded latent features, $g_\emptyset$ denotes the decoder function, ∅ represents the decoder parameters and $\hat{X}$ represents the reconstructed input. During training, the decoder attempts to reproduce the important characteristics of the original feature vector.

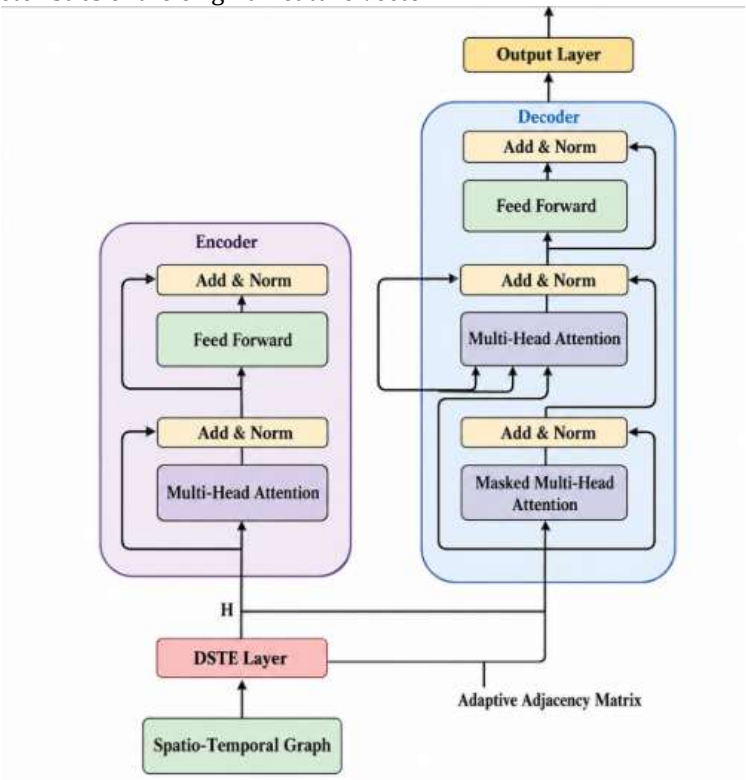

Figure 3: Autoencoder-Based Feature Denoising for IoT Malware Detection [29]

- **Reconstruction Loss**

The reconstruction loss is the difference between the original IoT traffic features and the features reconstructed by the decoder. In the proposed AEFD module, The Mean Squared Error (MSE) can be used for training the encoder-decoder network. The loss from reconstruction is calculated as follows:

$$L_{AE} = \frac{1}{N}\sum_{i=1}^{N}\left\|X_i - \hat{X}_i\right\|^2 \quad (8)$$

where X_i represents the original feature vector of the i^{th} IoT traffic sample and $\hat{X}_i$ represents its reconstructed feature vector. Minimizing L_{AE} encourages the AEFD model to preserve important traffic characteristics while suppressing noise and redundant variations.

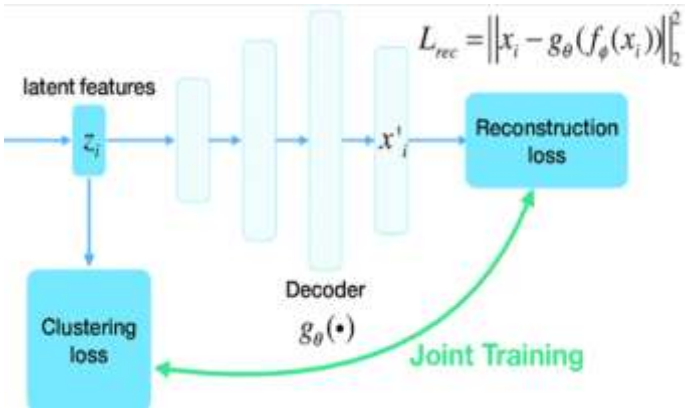

Figure 4: Reconstruction Loss for IoT Malware Feature Learning [30, 31]

3.4PCA-Based Feature Reduction and Deep Feature Representation

Based on AEFD, dimensionality reduction is performed using Principal Component Analysis (PCA), to remove redundant information and to keep the most important aspects of IoT network traffic. PCA maps the denoised feature space to a smaller dimensional space of uncorrelated principal components. The mean of each feature is first computed and the data after denoising is centred by subtracting the mean. This process would ensure the dominant patterns and informative variations in the IoT traffic data were properly captured in the subsequent covariance analysis.

$X_c = X_{denoised} - \mu$ (9)

The covariance matrix is then calculated as:

$\Sigma = \frac{1}{N-1} X_c^T X_c$ (10)

The eigenvalues and eigenvectors were obtained from:

$\Sigma_{v_i} = \lambda_i v_i$ (11)

The eigenvectors corresponding to the largest k eigenvalues were selected to form the projection matrix:

$W_k = [v_1, v_2, \dots, v_k]$ (12)

The reduced representation is obtained as:

$Z = X_c W_k$ (13)

Finally, the reduced features were transformed into a compact deep representation for subsequent Transformer-based malware classification.

3.5Transformer-Based Deep Learning for IoT Malware Detection and Classification

The features derived from the denoised and PCA-reduced IoT traffic were fed into the Transformer-based deep-learning module for malicious activity detection and classification.The features of the denoised and PCA-reduced IoT traffic were then input to the Transformer-based deep-learning module to detect and classify malicious activities. The input feature representation Z is converted to query, key and value matrices:

$Q = ZW_Q$ (14)

$K = ZW_K$ (15)

$V = ZW_V$ (16)

The self-attention mechanism evaluates relationships among traffic features:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (17)$$

Multiple attention heads were combined to capture diverse feature relationships:

$MHA = Concat(head_1, \dots, head_h)W^0$ (18)

The resulting representation is passed through a feed-forward network, followed by Softmax classification:

$P_c = \frac{e^{Sc}}{\sum_{j=1}^{C} e^{S_j}}$ (19)

The classification loss is calculated using categorical cross-entropy:

$L_{cls} = -\sum_{c=1}^{C} y_c \log(P_c)$ (20)

This Transformer architecture captures contextual dependencies within IoT traffic and enables accurate differentiation among benign and diverse malware classes.

3.6Number-Theory-Based Secure Cryptographic Framework

The proposed number-theory-based cryptographic framework secures IoT communication in terms of the generation of prime numbers, the use of modular arithmetic, key generation, encryption and optimized session-key management. Two prime numbers p and q were chosen to form the modulus:

$N = pq$ (21)

The Euler totient is calculated as:

$\phi(N) = (p-1)(q-1)$ (22)

A public exponent e is selected such that:

$$\gcd(e, \phi(N)) = 1 \quad (23)$$

The private key is obtained as:

$$d = e^{-1}(\mathrm{mod}\phi(N)) \quad (24)$$

For plaintext M, encryption and decryption were performed as:

$$C = M^e \mathrm{mod}\ N \quad (25)$$

$$M = C^d \mathrm{mod}\ N \quad (26)$$

A session key K_s is generated for secure communication, while PSO optimizes cryptographic parameters using:

$$v_i^{t+1} = Wv_i^t + C_1 r_1 (pbest_i - x_i^t) + C_2 r_2 (gbest - x_i^t) \quad (27)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (28)$$

3.7 Performance Evaluation and Comparative Analysis

- **Accuracy**

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (29)$$

- **Precision**

$$\text{Precision} = \frac{TP}{TP+FP} \quad (30)$$

- **Recall**

$$\text{Recall} = \frac{TP}{TP+FN} \quad (31)$$

- **F1-Score**

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (32)$$

- **ROC-AUC**

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (33)$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, TPR = True Positive Rate (Recall) and FPR = False Positive Rate.

4. Results and Discussion

For five representative devices, the figure5 shows the distribution of IoT traffic by device, with a comparatively balanced communication pattern. The top traffic sources were Danmini Doorbell with around 21.4%, Samsung Webcam with ~20.5% traffic and Philips Baby Monitor, with ~20.3% traffic. Provision Camera accounts for approximately 19.1% of the contribution, with Ecobee Thermostat taking the lowest amount with 18.7%. The traffic imbalance between the top and bottom in the ranking is about 2.7 percent, which suggests that there is not too much imbalance in traffic usage between different devices. Balanced device-level traffic is crucial for effective malware detection: The proposed secure cryptographic deep learning framework needs to differentiate between malicious and normal behaviors in IoT communication.

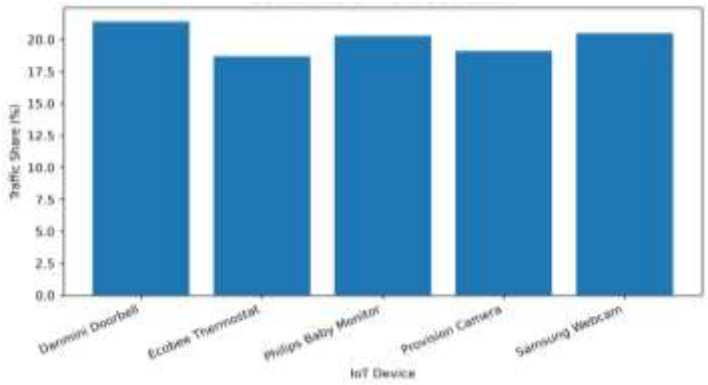

Figure 5: Device-Wise Distribution of IoT Traffic [32, 33]

The number of traffic records kept at various preprocessing stages for the IoT malware detection dataset is shown in figure 6. The raw data were about 100,000 records as the primary data source for developing the model. Once duplicates were removed, that leaves approximately 98,000 documents, meaning that almost 2,000 duplicate occurrences have been detected and removed. After excluding the missing values, about 98,000 instances were left, indicating that the pre-processing process retains nearly all good data. The small decrease across each stage shows that there was little loss of information during data cleansing. This fair logbook enables secure evaluation and training of the proposed secure cryptographic deep learning framework.

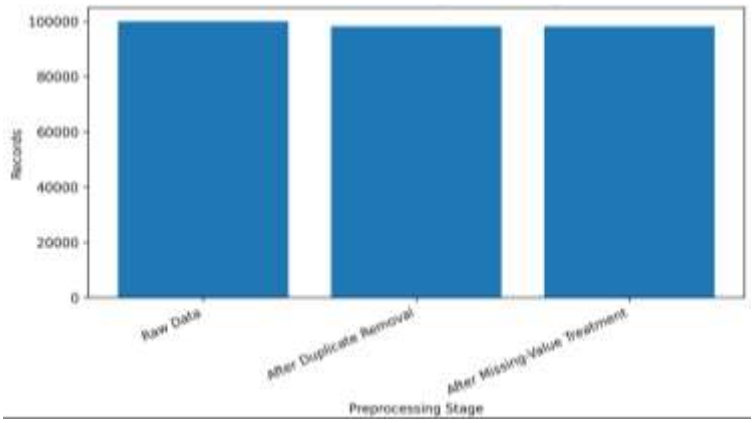

Figure 6: Traffic Records during Preprocessing [34]

Figure 7 shows the non-normalized and normalized (after the normalization process) representative values of features. The feature values in the raw data range widely in numbers from minimum of close to 0, to maximum of about 920, with the median being around 45. This variation can lead to dominant deep learning optimizations due to the larger values of the features. With normalization, the min, median and maximum values appear to have equal height (i.e. have been scaled to about 0 to 1, which is almost flat - the bars were very short relative to the original scale). This scaling would enhance the numerical spread, ensure computational stability and facilitate coherently oriented education. Therefore, in the proposed IoT malware detection framework, the features should be normalized form an appropriate input.

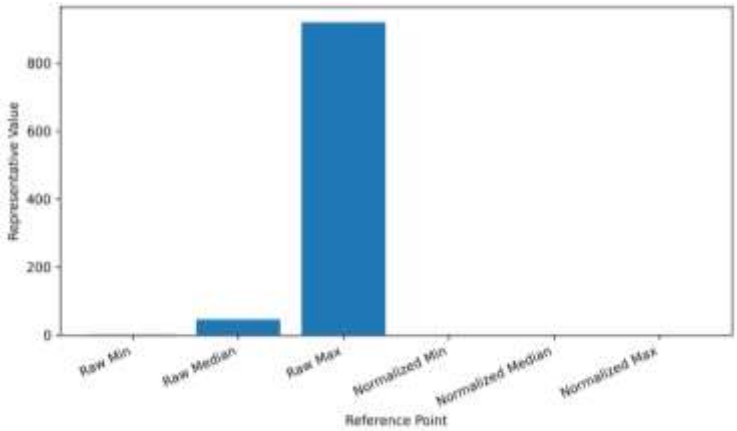

Figure 7: IoT Feature Value Normalization

Figure 8 shows how closely the AEFD model can reconstruct the data and its convergence behavior after 30 epochs. It can be seen that the mean squared error (MSE) is steadily reduced from about 0.085 at epoch 1 to 0.061 at epoch 5, 0.044 at epoch 10 and 0.033 at epoch 15. If further training is applied, the loss decreases to 0.027 at epoch 20, 0.023 at epoch 25 and almost 0.020 at epoch 30. In total, the reconstruction error reduces by about 76.5%, which proves that the reconstructed traffic patterns were similar to the underlying patterns with good learning. The convergence is slowly being reduced in epoch 20, indicating convergence is stable, which further elucidates the suitability of the AEFD model for the detection and classification of IoT malware.

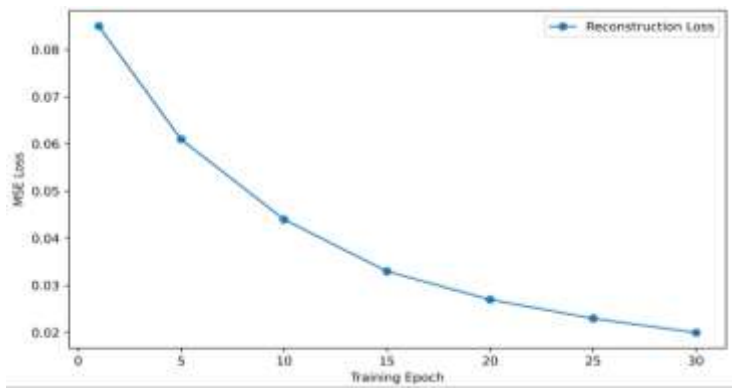

Figure 8: AEFD Reconstruction Loss during Training

The variance explained by the first six principal components following the PCA-based feature transformation is shown in figure 9. PC1 explains approximately 31.2% of the total variance, followed by PC2 with 19.5%, PC3 with 13.8%, PC4 with 10.7%, PC5 with 8.4% and PC6 with 6.1%. As a whole, 89.7% of the total variance is captured by these 6 components, which demonstrate that a lot of information can be retained from a smaller set of features. The PC1 and PC2 contributions show that there is significant dimensionality reduction potential. This compact and summarize representation can help in reducing computational complexity while preserving crucial patterns needed for IoT malware detection and classification.

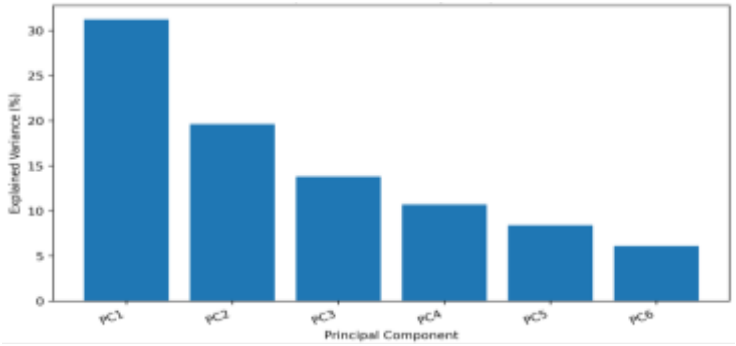

Figure 9: PCA Explained Variance across Components

Figure 10 shows the variations that were retained when more principal components were added to the transformed IoT traffic data. PC1 accounts for around 31% of the overall variance, PC2 accounts for 51% and PC3 accounts for 65% of the total variance. Adding PC4 increases the retained variance to ~75% and adding PC5 to it results in ~83% retained variance. When there were six principal components (PC6) about 89% of the original variance is retained. The results show that the compact PCA representation can represent most of the information in the original feature space. This dimensionality reduction may boost the computational efficiency while preserving informative patterns for malware classification in IoT.

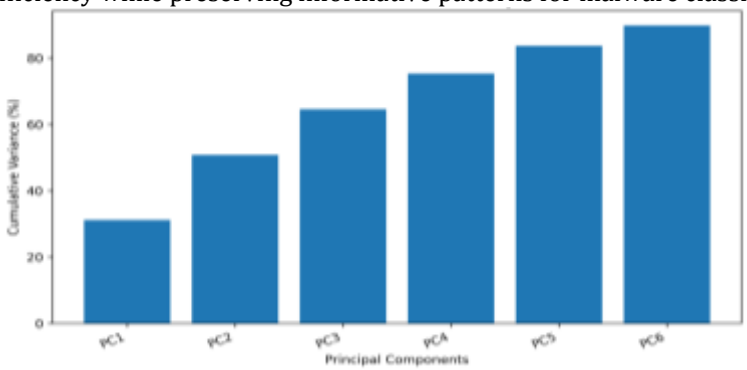

Figure 10: Cumulative Variance Preserved Through PCA

The accuracy of training and validation of the Transformer model for 30 training epochs is depicted in Figure 11. Training accuracy increases from 82.0% at epoch 1 to 89.0% at epoch 5, 93.0% at epoch 10 and 95.0% at epoch 15. It further improves to 96.5% at epoch 20, 97.3% at epoch 25 and approximately 97.8% at epoch 30. Validation accuracy follows a similar trend, increasing from 80.0% to 87.0%, 91.5%, 94.0%, 95.7%, 96.4% and 96.9%, respectively. IoT malware classification has good generalization and stable convergence with the small accuracy gap, about 0.9 percentage points.

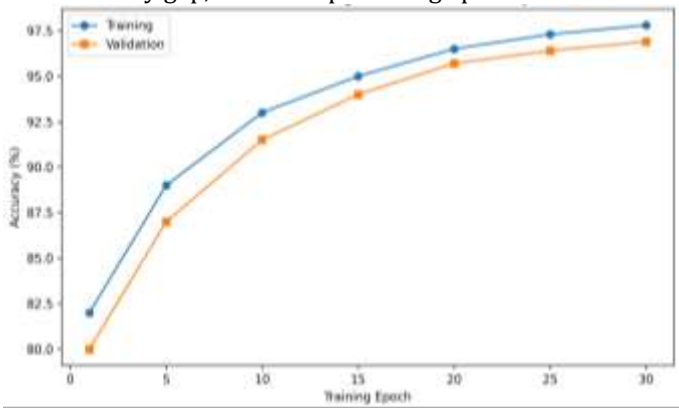

Figure 11: Transformer Training and Validation Accuracy

The cross-entropy loss for training and validation data of Transformer model is plotted over 30 epochs in Figure 12. The training loss is reduced by around 0.55 at epoch 1, 0.35 at epoch 5, 0.23 at epoch 10 and 0.16 at epoch 15. It then falls to 0.12 by epoch 20, 0.095 by epoch 25 and about 0.08 by epoch 30. Validation loss follows a similar pattern, decreasing from 0.59 to 0.39, 0.27, 0.20, 0.15, 0.13 and approximately 0.115, respectively. The rapid decrease suggests that the model is being optimized steadily and the small gap between the final loss ends and the loss during training (around 0.035) suggests that there was only limited overfitting and good generalization.

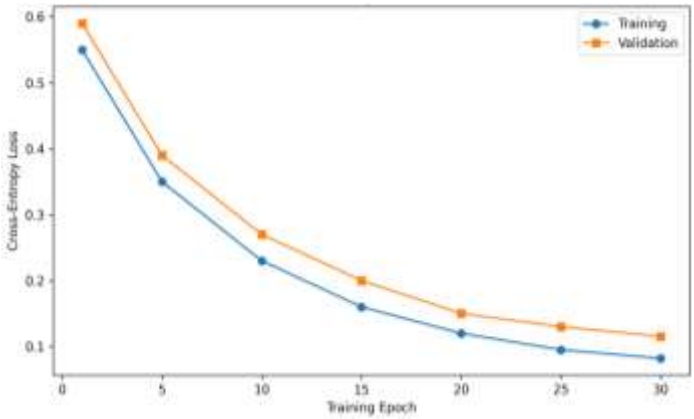

Figure 12: Transformer Training and Validation Loss

The accuracy of the models in classification the different types of malware is compared in Figure 13. The proposed AEFD-Transformer has achieved the highest accuracy of 97.8%, which is superior to LSTM (96.5%), CNN (95.5%) and RF (94.8%). The proposed model performs around 2.3, 1.3 and 3.0 percentage points better than CNN, LSTM and RF, respectively, compared to the former. The results highlight the benefit of using AEFD-based feature representation in conjunction with Transformer learning for learning complex IoT traffic patterns. The superior classification accuracy shows that the proposed method has better malware discrimination power than the traditional deep learning and machine learning methods compared in this study.

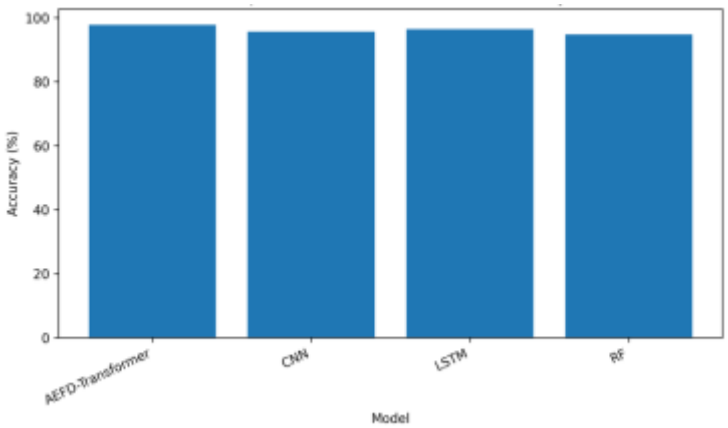

Figure 13: Comparative Malware Classification Accuracy

The performance of the proposed malware classifier is depicted in figure 14 in five evaluation metrics. The accuracy can be around 98.0% and the precision can be around 97.5%, which means that there is a high level of accuracy for the classification of malicious and benign traffic. Recall has the capability to detect actual malware instances with nearly 98.2% accuracy. The F1 score is about 97.9%, indicating a balance between precision (recall). The ROC-AUC value of 98.4% shows high performance criteria for all classification thresholds for distinguishing between malware and legitimate traffic. Overall, the performance over the narrow range of the metric, between 97.5-98.4%, validates the effectiveness of the proposed AEFD-Transformer framework for accurate classification and detection of IoT malware in general.

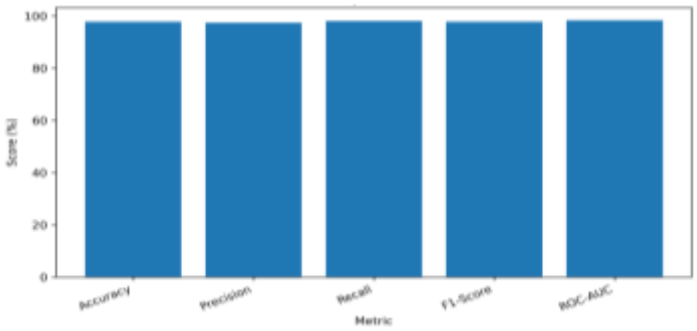

Figure 14: Proposed Malware Classifier Performance

The classification accuracy of the proposed model is studied for four classes: Benign, Mirai, Gafgyt and Other as shown in the confusion matrix of Figure 15. The model is able to predict the correct class for 19,600 of 20,000 test samples, indicating an overall accuracy of 98.00%. The Benign class has 5920 correct predictions and 30, 25 and 25 samples misclassified as Mirai, Gafgyt and Other, respectively. Mirai, 4,930 samples correctly identified, while 25, 20 and 15 were misclassified as Benign, Gafgyt and Other. Analogously, 4,920 Gafgyt samples and 5,910 Other samples were accurately classified. The high precision of 97.50%, recall of 98.20%, F1-score of 97.90% and ROC-AUC of 98.40% also indicates that the model shows a high and consistent classification performance in IoT malware.

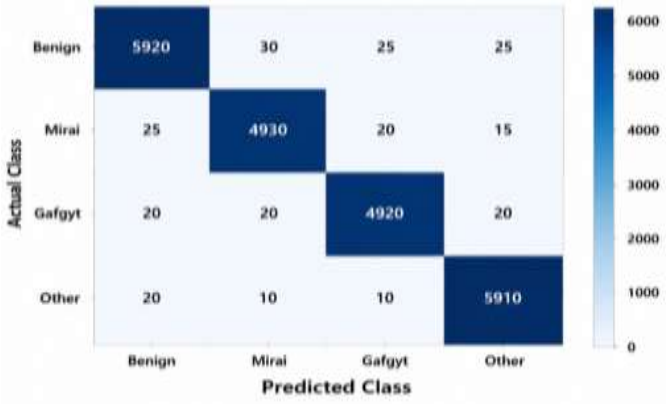

Figure 15: Confusion Matrix of Proposed Malware Classifier

The computational time taken by some of the main cryptographic tasks with the proposed secure framework is shown as shown in Figure. 16. The encryption process takes about 0.83ms and the key generation process takes about 1.42ms, which has the highest computational costs. The decryption operation takes about 0.91ms, which is below the overhead of the key generation operation. However, the processing cost is lowest in the case of session-key updating, with a delay of only about 0.37ms. Combining the time for all four operations is about 3.53ms, which represents a lightweight cryptographic process. Through these results, it can be seen that the number-theory-based security mechanism is capable of being implemented with minimum computational overhead to deliver cryptographic protection, making it well suited to efficiently deploying such security mechanism to both resource-limited IoT environments and malware detection applications.

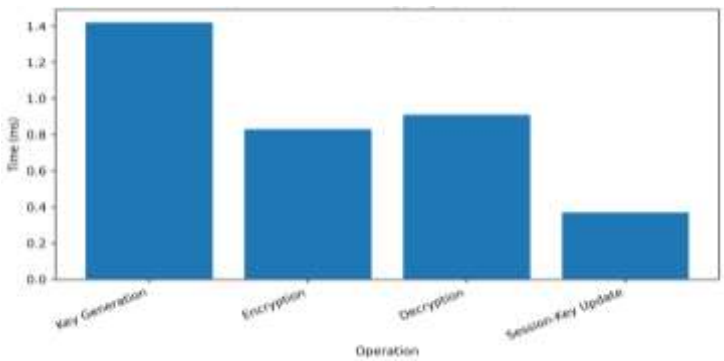

Figure 16: Computational Cost of Cryptographic Operations

5. Conclusion

This study proposed an integrated framework for IoT malware detection and classification based on AEFD, PCA, Transformer-based deep learning and cryptographic protection based on number theory. Results

were shown, revealing the ability to address both malware classification and feature redundancy, all while operating in communication security. AEFD successfully achieved informative traffic representations and decreased the reconstruction loss from 0.085 to 0.020 after 30 epochs, which is equivalent to 76.5% reduction. PCA further reduced the dimensionality of the features, while retaining around 89.7% of the original variance within the 6 principal components. The Transformer's learning was stable, with the validation accuracy being 96.9% and the training accuracy being 97.8%, representing only a 0.9-percentage point difference at the end. On the other hand, comparative evaluation was performed and the accuracy of the proposed AEFD-Transformer was found to be 97.8%, outperforming LSTM, CNN and Random Forest with 96.5, 95.5 and 94.8% respectively. The overall classifier achieved 98.0% accuracy, 98.2% recall, 97.9% F1-score and 98.4% ROC-AUC. Moreover, the cryptographic time needed was 3.53 ms in key generation, encryption, decryption and session-key updating. The results show that the framework is effective in detecting IoT malware in resource constrained environments, with both accuracy and security and is computationally efficient.

References

1. Bangash, Shah Hussain, Daud Khan, Atif Ishtiaq, Muhammad Imad, Mohsin Tahir, Waqas Ahmad, Ghassan Husnain, and Latif Jan. "Integrating machine learning and deep learning approaches for efficient malware detection in IoT-Based smart cities." Journal of Computing & Biomedical Informatics 5, no. 02 (2023): 280-299.
2. Riaz, Sharjeel, Shahzad Latif, Syed Muhammad Usman, Syed Sajid Ullah, Abeer D. Algarni, Amanullah Yasin, Aamir Anwar, Hela Elmannai, and Saddam Hussain. "Malware detection in internet of things (IoT) devices using deep learning." Sensors 22, no. 23 (2022): 9305.
3. Hua, Binjie, and Haiyan Xi. "A privacy preserving intrusion detection framework for IIoT in 6G networks using homomorphic encryption and graph neural networks." Scientific Reports 16, no. 1 (2025): 2297.
4. Akwaronwu, Bright, Innocent U. Akwaronwu, Oluwabamise J. Adeniyi, and Ayodeji G. Abiodun. "Assessing Machine Learning Techniques for Cryptographic Attack Detection: A Systematic Review and Meta-Analysis." Indonesian Journal of Data and Science 6, no. 2 (2025): 184-202.
5. Rahmati, Milad, and Antonino Pagano. "Federated learning-driven cybersecurity framework for IoT networks with privacy preserving and real-time threat detection capabilities." In Informatics, vol. 12, no. 3, p. 62. MDPI, 2025.
6. Asiri, Mohammed, Maher A. Khemakhem, Reemah M. Alhebshi, Bassma S. Alsulami, and Fathy E. Eassa. "Rpfl: A reliable and privacy-preserving framework for federated learning-based iot malware detection." Electronics 14, no. 6 (2025): 1089.
7. Chinbat, Tserendorj, Samaneh Madanian, David Airehrour, and Farkhondeh Hassandoust. "Machine learning cryptography methods for IoT in healthcare." BMC Medical Informatics and Decision Making 24, no. 1 (2024): 153.
8. Almotiri, Sultan H. "AI driven IOMT security framework for advanced malware and ransomware detection in SDN." Journal of Cloud Computing 14, no. 1 (2025): 19.
9. Maghanaki, Mazdak, Soraya Keramati, F. Frank Chen, and Mohammad Shahin. "Generation of a Multi-Class IoT Malware Dataset for Cybersecurity." Electronics 14, no. 21 (2025): 4196.
10. Taşcı, Burak. "Deep-learning-based approach for IoT attack and malware detection." Applied Sciences 14, no. 18 (2024): 8505.
11. Maghanaki, Mazdak, Soraya Keramati, F. Frank Chen, and Mohammad Shahin. "Systematic evaluation of machine learning and deep learning models for IoT malware detection across ransomware, rootkit, spyware, trojan, botnet, worm, virus, and keylogger." Sensors 26, no. 6 (2026): 1750.
12. Kim, Ho-myung, and Kyung-ho Lee. "IIoT malware detection using edge computing and deep learning for cybersecurity in smart factories." Applied Sciences 12, no. 15 (2022): 7679.
13. Ijaz, Aqsa, Ammar Ahmad Khan, Muhammad Arslan, Ashir Tanzil, Alina Javed, Muhammad Asad Ullah Khalid, and Shouzab Khan. "Innovative machine learning techniques for malware detection." Journal of Computing & Biomedical Informatics 7, no. 01 (2024): 403-424.
14. Abid, Yawar Abbas, Jinsong Wu, Muhammad Farhan, and Tariq Ahmad. "ECMT framework for internet of things: an integrative approach employing In-Memory attribute examination and sophisticated neural network architectures in conjunction with hybridized machine learning methodologies." IEEE Internet of Things Journal 11, no. 4 (2023): 5867-5886.
15. El-Ghamry, Amir, Tarek Gaber, Kamel K. Mohammed, and Aboul Ella Hassanien. "Optimized and efficient image-based IoT malware detection method." Electronics 12, no. 3 (2023): 708.
16. Sirhan, Najem N., Riyad Alrousan, and Hussam N. Fakhouri. "A Blockchain and Federated Learning Framework for Image-Based IoT Malware Detection and Prevention." IoT 7, no. 3 (2026): 56.
17. Ahmad, Naveed, Yue Cao, and William Liu. "A Blockchain-Enabled Security Framework for Cloud-Based Sensor Systems with Deep Learning-Driven Attack Classification." Sensors 26, no. 16 (2026): 5198.

18. Gunasridharan, Ramasubramaniyan, Ali Altalbe, Bharathi Mohan Gurusamy, Gundala Pallavi, and Prasanna Kumar Rangarajan. "A hybrid framework integrating QML, AI, and quantum-safe cryptography for cybersecurity." Journal of Computational and Cognitive Engineering 5, no. 1 (2026): 16-29.
19. Mutambik, Ibrahim. "AI-driven cybersecurity in IoT: Adaptive malware detection and lightweight encryption via TRIM-SEC framework." Sensors 25, no. 22 (2025): 7072.
20. Kumari, Swati. "Next-Gen IoT Security using Polar Codes-based Cryptography for malware defence through quantum self-attention neural network." Knowledge-Based Systems 321 (2025): 113716.
21. Safeer, Ehtesham, Sidra Tahir, Asif Nawaz, Mamoona Humayun, Momina Shaheen, and Maqbool Khan. "Advanced hybrid malware identification framework for the Internet of Medical Things, driven by deep learning." Security and privacy 8, no. 1 (2025): e454.
22. Khan, Sabbir Ahmed. "Privacy-Preserving Deep Learning Framework for IoT Malware Detection." PhD diss., Old Dominion University, 2024.
23. Baazeem, Rami. "Multilayered framework for enhancing data confidentiality, integrity, and threat detection through blockchain, advanced cryptography, and machine learning." International Journal of Computer Networks and Applications (2024).
24. Hamarsheh, Ala. "An adaptive security framework for Internet of Things networks leveraging SDN and machine learning." Applied Sciences 14, no. 11 (2024): 4530.
25. Devi, R. Aiyshwariya, and A. R. Arunachalam. "Enhancement of IoT device security using an Improved Elliptic Curve Cryptography algorithm and malware detection utilizing deep LSTM." High-Confidence Computing 3, no. 2 (2023): 100117.
26. Hashi, Abdinasir Ismael, and Mr Osman Abdullahi Jama. "Evaluating the Performance of AI-Based Software Tools in Intelligent Decision-Making Systems." International Journal of Computing and Engineering 7, no. 22 (2025): 1-20.
27. Shanker, Ravi. "Edge-Accelerated Analytics for Encrypted Network Traffic using Optimized Machine Learning." In 2025 International Conference on Computing Technologies & Data Communication (ICCTDC), pp. 1-7. IEEE, 2025
28. Hashi, Abdinasir Ismael, and Mr Osman Abdullahi Jama. "Evaluating the Performance of AI-Based Software Tools in Intelligent Decision-Making Systems." International Journal of Computing and Engineering 7, no. 22 (2025): 1-20.
29. Khan, Saddam Hussain, Tahani Jaser Alahmadi, Wasi Ullah, Javed Iqbal, Azizur Rahim, Hend Khalid Alkahtani, Wajdi Alghamdi, and Alaa Omran Almagrabi. "A new deep boosted CNN and ensemble learning based IoT malware detection." Computers & Security 133 (2023): 103385.
30. Santhosh Kumar, S. V. N., M. Selvi, and A. Kannan. "A comprehensive survey on machine learning-based intrusion detection systems for secure communication in internet of things." Computational intelligence and neuroscience 2023, no. 1 (2023): 8981988.
https://www.kaggle.com/datasets/saurabhshahane/anomaly-detection-using-deep-learning
31. Sun, Wei, Rongzhang Cheng, Yingqi Jiao, Junbo Gao, Zhedian Zheng, and Nan Lu. "Transformer network with decoupled spatial–temporal embedding for traffic flow forecasting." Applied Intelligence 53, no. 24 (2023): 30148-30168.
32. Min, Erxue, Xifeng Guo, Qiang Liu, Gen Zhang, Jianjing Cui, and Jun Long. "A survey of clustering with deep learning: From the perspective of network architecture." IEEE access 6 (2018): 39501-39514.
33. Hashi, Abdinasir Ismael, Mr Osman Abdullahi Jama, and Mr Abdirizak Mohamed Hashi. "A Quantum-Resilient Decentralized Swarm Intelligence Framework for Secure Real-Time Synchronization in Industrial Metaverse Environments." Journal of Technology and Systems 8, no. 2 (2026): 61-82.