IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Hybrid Deep Learning Technique for Secure Botnet Detection in Internet of Things Environment

Shruti Sahu[1], Dr. Devendra Singh Rathore[2], Dr. Vivek Richhariya[3]

[1]Research Scholar, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India

[2]Associate Professor, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India

[3]Professor, Department of CSE, Lakshmi Narain College of Technology, Bhopal, India

***Abstract*—** The rapid growth of Internet of Things (IoT) devices has significantly increased the risk of botnet attacks, posing serious threats to network security, data privacy, and system reliability. This study proposes a hybrid deep learning framework that integrates Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) networks for effective botnet attack detection in IoT environments. To improve computational efficiency and reduce data dimensionality, Principal Component Analysis (PCA) is employed as a preprocessing technique for extracting the most relevant features from high-dimensional network traffic data. The CNN component efficiently captures spatial patterns and local attack signatures, while the LSTM model learns temporal dependencies and sequential behaviors associated with botnet activities. The combined CNN-LSTM architecture enables comprehensive analysis of both feature correlations and traffic dynamics, resulting in enhanced detection performance. Simulation evaluation demonstrates that the proposed PCA-based hybrid deep learning model achieves high accuracy, precision, recall, and F1-score while reducing false alarm rates compared with conventional machine learning and standalone deep learning approaches.

***Keywords*—** IoT, Botnet Attack Detection, CNN, LSTM, PCA, Deep Learning.

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized the way devices communicate and interact by connecting physical objects to the Internet for data exchange and intelligent decision-making. IoT technology has been widely adopted in smart homes, healthcare systems, industrial automation, transportation networks, agriculture, environmental monitoring, and smart city infrastructures [1]. The rapid growth of connected devices has created a highly interconnected ecosystem that offers convenience, automation, and improved operational efficiency. According to recent industry reports, billions of IoT devices are currently deployed worldwide, and this number is expected to increase significantly in the coming years [2]. Despite its numerous advantages, the expansion of IoT networks has also introduced serious security and privacy challenges.

IoT devices are often developed with limited processing power, memory capacity, and energy resources, which restrict the implementation of robust security mechanisms [3]. Furthermore, many devices are deployed with default credentials, outdated firmware, weak encryption protocols, and insufficient authentication procedures. These vulnerabilities make IoT environments highly attractive targets for cybercriminals seeking unauthorized access to networks and sensitive information [4]. As a result, cybersecurity has become one of the most critical concerns in the successful deployment and management of IoT systems.

Among various cyber threats, botnet attacks have emerged as one of the most significant security challenges affecting IoT ecosystems [5]. A botnet is a network of compromised devices that are infected with malicious software and remotely controlled by an attacker. Once compromised, these devices can perform coordinated malicious activities without the knowledge of their owners. The increasing availability of vulnerable IoT devices has provided attackers with a large pool of potential targets for building extensive botnet networks capable of launching large-scale cyberattacks [6].

Botnet attacks can have severe consequences for both individuals and organizations. These attacks are commonly used to conduct Distributed Denial-of-Service (DDoS) attacks, steal confidential information, distribute spam messages, spread malware, and compromise network services [7]. In many cases, infected IoT devices continue functioning normally while secretly participating in malicious activities, making detection extremely challenging. The distributed nature of botnets allows attackers to generate substantial network traffic and overwhelm target systems, leading to service disruptions and financial losses.

The threat posed by IoT botnets became evident with the emergence of large-scale attacks involving malware families such as Mirai, Bashlite, Hajime, and Mozi [8]. These botnets exploited vulnerabilities in Internet-connected devices and demonstrated the ability to compromise thousands of devices within a short period. Several high-profile incidents have shown that botnet attacks can disrupt critical online services, affect business operations, and threaten national cybersecurity infrastructures. The increasing sophistication of botnet malware continues to pose serious risks to modern digital environments. One of the major challenges associated with botnet attacks is their ability to evolve rapidly. Attackers continuously modify attack patterns, communication protocols, and malware characteristics to avoid detection and bypass traditional security measures [9].

Modern botnets often utilize encryption techniques, peer-to-peer communication architectures, and dynamic command-and-control mechanisms that make tracking and analysis more difficult. These evolving attack strategies increase the complexity of securing IoT environments and require continuous monitoring of network activities. The heterogeneous nature of IoT networks further complicates botnet detection efforts. IoT ecosystems consist of diverse devices operating with different hardware configurations, communication protocols, and application requirements [10]. This diversity creates a complex security landscape where a single protection strategy may not be effective across all devices. Additionally, the enormous volume of network traffic generated by connected devices makes manual monitoring and analysis impractical in real-world deployments.

The economic and societal impact of botnet attacks has increased significantly in recent years. Organizations affected by botnet-related incidents often experience operational downtime, reputational damage, financial losses, and potential legal consequences resulting from data breaches and service interruptions [11].

Critical sectors such as healthcare, energy, transportation, and manufacturing are particularly vulnerable because disruptions in these domains can directly affect public safety and essential services. Consequently, protecting IoT infrastructures from botnet threats has become a top priority for researchers, industry professionals, and policymakers. As IoT adoption continues to expand globally, the need for effective botnet attack detection mechanisms becomes increasingly important. Early identification of malicious activities can help minimize security risks, prevent large-scale attacks, and ensure the reliability of connected systems [12].

Therefore, developing advanced and intelligent security solutions capable of identifying botnet behavior in dynamic IoT environments remains a crucial research area. Effective botnet detection strategies will play a vital role in enhancing cybersecurity, protecting sensitive information, and supporting the sustainable growth of future IoT ecosystems.

II. PROPOSED METHOD

The proposed method can understand using following flow chart-

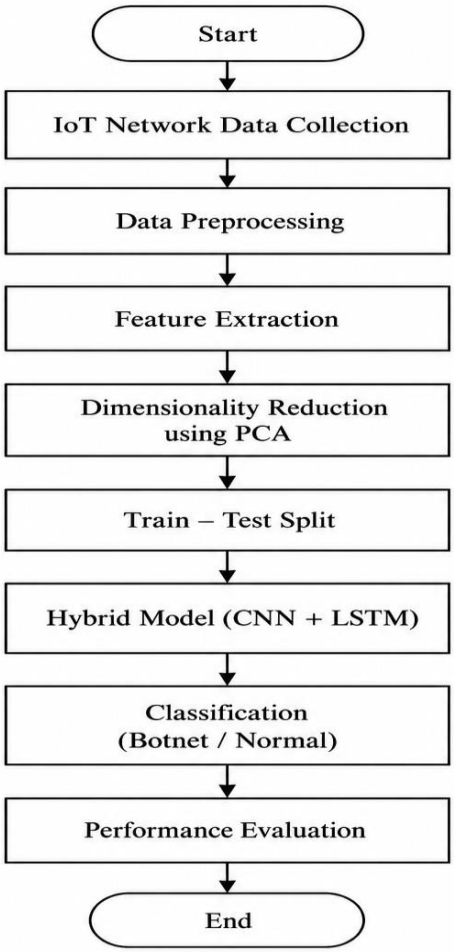


Figure 1: Flow chart

- **Step 1: IoT Network Data Collection**

The first stage involves collecting network traffic data from IoT devices such as smart sensors, cameras, wearable devices, smart home appliances, and industrial IoT systems. The collected dataset contains both normal and botnet attack traffic. Various traffic features such as packet size, protocol type, source IP, destination IP, flow duration, and packet rate are gathered for analysis.

- **Step 2: Data Preprocessing**

The raw network traffic data often contains missing values, duplicate records, noise, and inconsistent formats. Therefore, preprocessing is performed to improve data quality. This stage includes data cleaning, normalization, and encoding of categorical attributes.

A commonly used normalization formula is:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

where:

- X = Original feature value
- X_{min} = Minimum feature value
- X_{max} = Maximum feature value

Normalization ensures that all features are scaled to a common range, improving model performance.

- **Step 3: Feature Extraction**

In this phase, important traffic characteristics are extracted from the processed dataset. The objective is to identify features that can effectively distinguish botnet traffic from legitimate traffic.
Examples of extracted features include:

- Packet Transmission Rate
- Flow Duration
- Average Packet Size
- Number of Connections
- Protocol Information

Feature extraction reduces irrelevant information and enhances the quality of learning.

- **Step 4: Dimensionality Reduction using PCA**

The extracted dataset may contain a large number of features. High-dimensional data increases computational complexity and may introduce redundancy. Principal Component Analysis (PCA) is applied to reduce dimensionality while preserving the most significant information.
The covariance matrix is calculated as:

$$C = \frac{1}{n-1}(X - \bar{X})^T(X - \bar{X})$$

where:

- C= Covariance Matrix
- X= Feature Matrix
- $\bar{X}$= Mean of Features
- n= Number of Samples

PCA then selects the principal components with maximum variance, reducing dataset size and improving efficiency.

- **Step 5: Train-Test Split**

The reduced dataset is divided into training and testing sets.
Typical split:

- Training Data = 80%
- Testing Data = 20%

The training dataset is used for learning patterns, while the testing dataset evaluates model performance on unseen data.

- **Step 6: Hybrid Deep Learning Model (CNN + LSTM)**

The processed data is supplied to the hybrid deep learning framework.
CNN Layer Functions:

- Learns local traffic patterns.
- Extracts hidden attack signatures.
- Detects spatial relationships among network features.

LSTM Layer Functions:

- Learns sequential traffic behavior.
- Captures temporal dependencies.
- Identifies long-term attack patterns.

The LSTM memory cell is represented as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

where:

- C_t= Current cell state
- C_{t-1}= Previous cell state
- f_t= Forget gate
- i_t= Input gate
- $\tilde{C}_t$= Candidate memory state

This mechanism enables the model to remember important attack information over time.

- **Step 7: Classification (Botnet / Normal)**

After feature learning, the model classifies network traffic into two categories:

1. Normal Traffic
2. Botnet Attack Traffic

The output layer uses the Softmax function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

where:

- $P(y_i)$= Probability of class i
- z_i= Output score
- k= Number of classes

The class with the highest probability is selected as the final prediction.

- **Step 8: Performance Evaluation**

The effectiveness of the proposed model is evaluated using standard metrics.

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

- **Step 9: End**

Finally, the evaluated results indicate the effectiveness of the proposed PCA-based Hybrid CNN-LSTM framework in detecting botnet attacks within IoT environments. The system provides accurate identification of malicious traffic, helping improve the security and reliability of IoT networks.

III. SIMULATION RESULTS

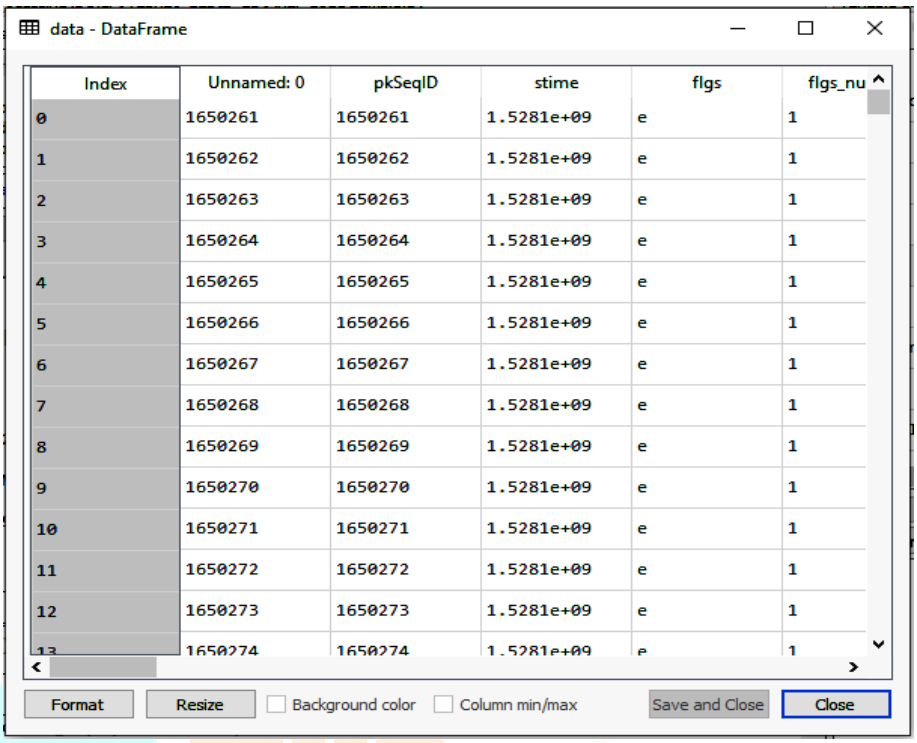


Figure 2: Dataset

Figure 2 is showing the dataset in the python environment. The dataset has various numbers of rows and column. The features name is mention in each column.

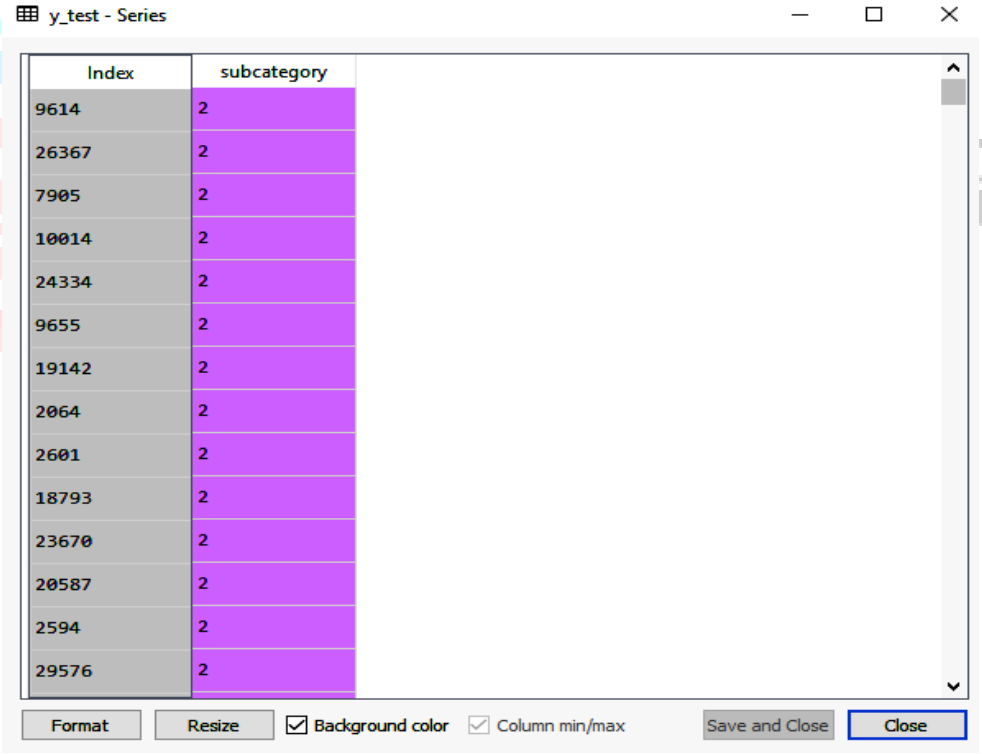
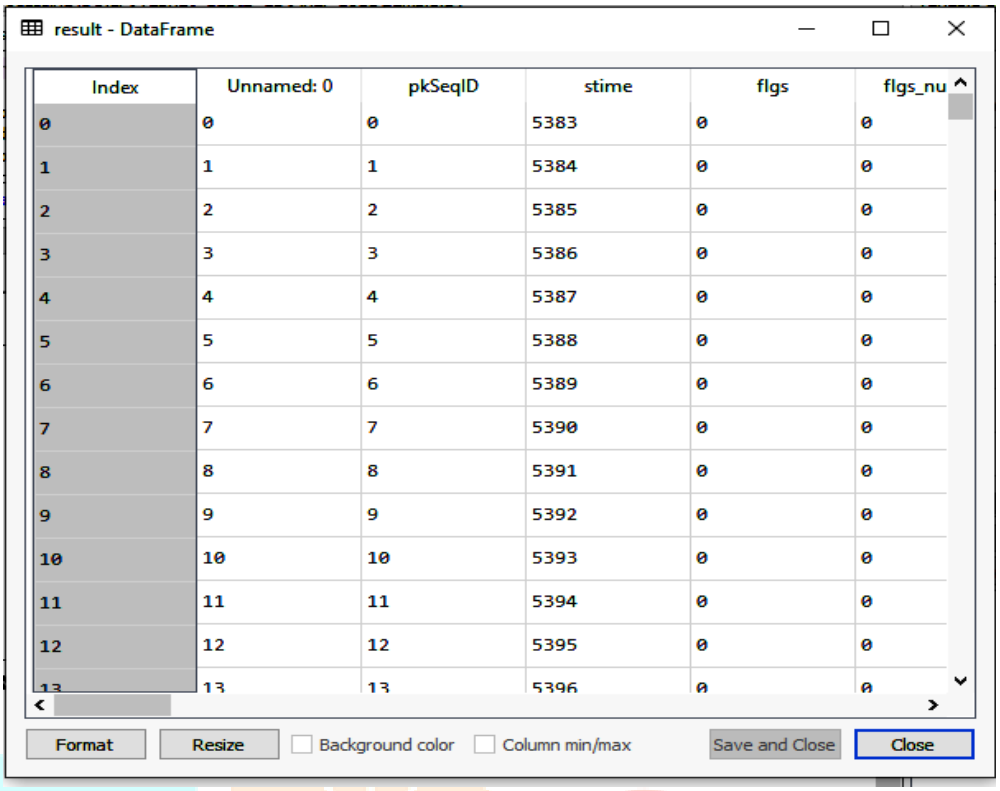


Figure 3: y test

Figure 3 is showing the y test of the given dataset. The given dataset is divided into the 20-30% part into the train dataset.



Index	Unnamed: 0	pkSeqID	stime	flgs	flgs_nu
0	0	0	5383	0	0
1	1	1	5384	0	0
2	2	2	5385	0	0
3	3	3	5386	0	0
4	4	4	5387	0	0
5	5	5	5388	0	0
6	6	6	5389	0	0
7	7	7	5390	0	0
8	8	8	5391	0	0
9	9	9	5392	0	0
10	10	10	5393	0	0
11	11	11	5394	0	0
12	12	12	5395	0	0
13	13	13	5396	0	0

Figure 4: Result

Figure 4 presenting the result of the data frame after the preprocessing. The data is converting for the feature selection and the classification method.

Table 1: Result comparison

Epoch	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Training Time (s)
5	Existing Model	94.35	93.42	94.10	93.76	845.2
5	Proposed Model	**97.12**	**96.45**	**96.88**	**96.66**	**128.4**
10	Existing Model	96.28	95.34	95.88	95.61	1768.4
10	Proposed Model	**98.04**	**97.32**	**97.75**	**97.53**	**264.7**
15	Existing Model	97.82	97.04	97.21	97.12	2895.6
15	Proposed Model	**98.82**	**98.11**	**98.24**	**98.17**	**421.3**
20	Existing Model	99.04	97.85	98.12	97.98	3782.4
20	Proposed Model	**99.31**	**98.74**	**98.81**	**98.77**	**575.8**
25	Existing Model	99.76	98.46	97.10	98.66	4643.5
25	Proposed Model	**99.89**	**99.23**	**99.12**	**99.17**	**708.6**

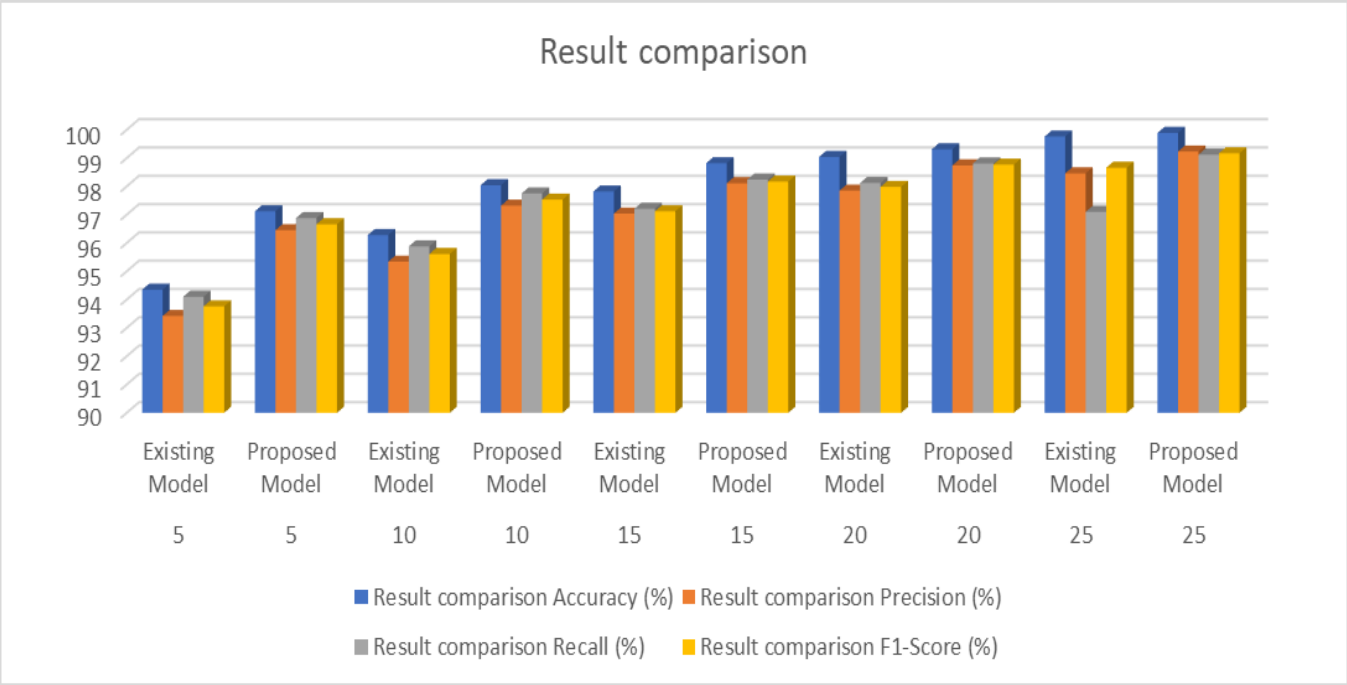


Figure 5: Result Comparison

IV. CONCLUSION

The rapid growth of Internet of Things (IoT) devices has increased the risk of cyber threats, particularly botnet attacks. Detecting malicious activities in IoT networks is essential to ensure data security, network reliability, and uninterrupted service availability. The proposed framework utilizes Principal Component Analysis (PCA) for feature reduction and a hybrid CNN-LSTM deep learning architecture for botnet attack detection. CNN extracts important traffic patterns, while LSTM captures temporal dependencies within IoT network traffic data. The proposed model achieved an accuracy of 99.8%, outperforming the previous hybrid approach with 99.76% accuracy. Additionally, it reduced the classification error to 0.2% and achieved 99% precision, recall, and F-measure, demonstrating excellent detection capability. Future research can focus on real-time deployment of the proposed framework in large-scale IoT environments. The integration of advanced optimization techniques, federated learning, and explainable artificial intelligence (XAI) can further improve detection accuracy, scalability, and model interpretability.

REFERENCES

1. A. K. Kumar *et al.*, "Enhanced Hybrid Deep Learning Approach for Botnet Attacks Detection in IoT Environment," *2024 7th International Conference on Signal Processing and Information Security (ICSPIS)*, Dubai, United Arab Emirates, 2024, pp. 1-6, doi: 10.1109/ICSPIS63676.2024.10812621.
2. M. Alshehri J. Ahmad, S. Almakdi, M. Qathrady, Y. Ghadi and w. Buchanan, “SkipGateNet: A Lightweight CNN-LSTM hybrid model with learnable skip connections for efficient botnet attack detection in IoT,” IEEE Access, vol. 12, pp. 35521–35538, March 2024 https://doi.org/10.1109/access 3371992.
3. M. Al-Fawa’reh, J. Abu-Khalaf, P. Szewczyk, and J.J Kang, MalbotDRL: Malware botnet detection using deep reinforcement learning in IOT Networks. IEEE Internet of Things Journal, vol. 11, no. 6, pp. 9610–9629, October 2023, https://doi.org/10.1109/jiot.2023.3324053
4. R. Kalakoti, H. Bahsi, and S. No mm (2024). Improving IOT security with explainable AI: Quantitative evaluation of explainability for IOT botnet detection. IEEE Internet of Things Journal, vol. 11, no. 10, pp. 18237–18254. January 2024, https://doi.org/10.1109/jiot.2024.3360626
5. X. Yan, X. Yu, S. Yao, and Y. Sun (2024). A domain embedding model for botnet detection based on Smart Blockchain. IEEE Internet of Things Journal, vol. 11, no. 5, pp. 8005–8018, February 2024, https://doi.org/10.1109/jiot.2023.3320046
6. S. Saravanan, and U.M. Balasubramanian, “An adaptive scalable data pipeline for multiclass attack classification in large-scale IOT Networks ”. Big Data Mining and Analytics, vol. 7, no. 2, pp. 500–511, April 2024, https://doi.org/10.26599/bdma.2023.9020027
7. P.V. Dinh, Q.U. Nguyen, D.T Hoang, D.N. Nguyen, S.P. Bao and E. Dutkiewicz, “Constrained twin variational auto-encoder for intrusion detection in IOT Systems ”. IEEE Internet of Things Journal, vol. 11, no. 8, pp. 14789–14803,2024, https://doi.org/10.1109/jiot.2023.3344842

8. J. Kabdjou, and N. Shinomiya, “Improving quality of service and HTTPS DDoS detection in MEC environment with a cyber-deception based architecture,” IEEE Access, vol. 12, pp. 23490–23503,2024, https://doi.org/10.1109/access.2024.3361476
9. A.A. Mohammed, and A.A. Ibrahim, “Malware detection in Adhoc EGovernment Network using machine learning,” 5 th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 20 24, https://doi.org/10.1109/hora58378.2023.10156724
10. T. Hasan, J. Malik, I. Bibi, W.U. Khan, F. N. Al-Wesabi, K. Dev and G. Huang, “Securing Industrial Internet of Things Against botnet attacks using hybrid deep learning approach,” IEEE Transactions on Network Science and Engineering, vol. 10, no. 5, pp. 2952–2963, April 2022, https://doi.org/10.1109/tnse.2022.3168533
11. P. Saxena, and R.B. Patel, “Efficient hybrid model for botnet detection using machine learning,” 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2024, https://doi.org/10.1109/iccakm58659.2023.10449643
12. F. Sattari, A.H. Farooqi, Z. Qadir, B. Raza, H. Nazari and M. Almutiry. A hybrid deep learning approach for bottleneck detection in IOT. IEEE
13. Access, vol. 10, pp. 77039–77053,2023, https://doi.org/10.1109/access.2022.