

**International Journal of
Engineering Research and Science & Technology**



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

EFFICIENT BOTNET DETECTION IN IOT NETWORKS USING A HYBRID MACHINE LEARNING APPROACH

¹*Pinjari Ibrahim, MCA Student, Department of MCA*

²*CH Sri Lakshmi Prasanna, M.Tech, (Ph.D), Assistant Professor, Department of MCA*

¹²*Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool*

ABSTRACT

As internet technology advances and becomes more widely used, cyberattacks are becoming more frequent. One of the most damaging assaults was the botnet attack. Because of the many attack channels and the constant development of viruses, botnet detection is getting more difficult. The fast development of Internet of Things (IoT) technology has resulted in botnet assaults targeting several network devices, which have caused significant losses across various industries. Network security is seriously threatened by botnets, and deep learning models have shown the ability to effectively detect botnet activity from network traffic data. This study proposes a botnet detection system based on the stacking of recurrent neural networks (RNN), convolutional neural networks (CNN), artificial neural networks (ANN), and machine learning models (ACLR). Both the individual models and the suggested ACLR model for performance comparison are used in the tests. Nine distinct attack types, including "Normal," "Generic," "Exploits," "Fuzzers," "DoS," "Reconnaissance," "Analysis," "Backdoor," "Shell code," and "Worms," are included in the UNSW-NB15 dataset, which is utilised for botnet assaults. According to experimental data, the suggested ACLR model well captures the complex patterns and features of botnet assaults, achieving a testing accuracy of 0.9698. A K-fold cross-validation accuracy score of 0.9749 for the proposed ACLR model's k values (3, 5, 7, and 10) shows that k = 5 demonstrates the model's generalisability and resilience. Furthermore, the suggested model identifies botnets with a precision-recall area

under the curve (PR-AUC) of 0.9950 and a high receiver operating characteristic area under the curve (ROC-AUC) of 0.9934. The higher performance of the suggested method is further supported by a comparison with current state-of-the-art models. The findings of this study may improve cyber security protocols and provide valuable defence against changing threats.

I. INTRODUCTION

The public quickly and widely adopted Internet technology for everyday, social, cultural, and institutional purposes as a result of its emergence. Like other technology, the internet has its drawbacks, such as targeting individuals to steal their personal information or money. Because it helps avoid and reduce a number of online security problems, bot net attack detection is an essential part of cyber security. By detecting and eliminating botnets, security professionals may shield networks and data from dangerous activities including distributed denial of service (DDOS) attacks, data breaches, and malware dissemination. In addition to reducing potential harm, early detection maintains network efficiency and user trust in digital services. It also encourages cyber resiliency, ensures compliance with legal and regulatory requirements, and fosters innovation in the ongoing fight against ever evolving cyberthreats. Strong machine learning models are used in deep learning-based botnet detection to swiftly identify and classify malicious botnet activity in network data. Rapidly detecting and thwarting botnet-posed cyberthreats helps businesses safeguard their data and systems. In order to gain financial benefit, such assaults have been carried out in a number of methods against both people and

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267>

organisations. Ransomware, one of the most well-known types of attack, attacks a person and locks their data until they pay the ransom the perpetrator demands. The attackers target large organisations by using botnets. Recently, there has been a lot of interest in botnet detection using deep learning algorithms because of its effectiveness in combating the rising threat of botnets.

Deep learning-based network traffic analysers are now a useful tool for identifying and minimising botnet activity. These analysers automatically extract relevant information from raw packet data using deep learning algorithms. In order to find patterns and characteristics indicative of botnet traffic, the first few packets in a flow's headers are particularly collected and analysed. Regardless of the underlying botnet architecture, malicious botnet traffic may be identified using convolutional neural networks (CNNs) and auto encoders [1]. In order to educate the network the basic form of network traffic data, auto encoders are employed to teach the network how to reconstruct its input. This technique helps identify odd patterns that indicate botnet activity. Conversely, CNNs are excellent at analysing structured data, such network traffic, by spotting spatial relationships and hierarchical representations. Because deep learning algorithms are so powerful, researchers and practitioners in the area of botnet detection have made significant strides in detecting and mitigating botnet dangers. Proactive defences against botnet assaults are now possible because to these approaches' promising achievements in accurately classifying and recognising botnet traffic [2].

Deep learning techniques for botnet detection have shown promise in reducing the danger posed by botnets. It has been proposed that network traffic analysers based on deep learning can effectively identify and stop botnet activity. Bidirectional long-short-term memory

recurrent neural networks (BLSTM-RNN) are a well-known illustration of the use of deep learning to botnet detection processes. Because BLSTM-RNN models excel at gathering both the past and future context of sequential data, they are ideally suited for examining network traffic and identifying patterns linked to botnet activities [3]. The use of deep learning algorithms for botnet detection has many advantages. First of all, these algorithms can automatically learn and adapt to the evolving features of botnets, enabling them to identify new and unidentified botnet behaviours. Second, by extracting features from raw data packets, these algorithms are able to identify hidden patterns and abnormalities that would be invisible using traditional detection methods. Detecting botnet activity in real-time or almost real-time is made feasible by deep learning models' ability to process massive volumes of network traffic data rapidly. The ability to identify botnet activity is essential for network and device security and integrity. Botnets pose a significant risk as they may be used for data theft, distributed denial of service (DDOS) attacks, and spam dissemination. Researchers and practitioners anticipate that using deep learning algorithms to botnet detection would improve detection methods' precision and effectiveness and make proactive defences against botnet assaults possible [4].

The growth of Internet of Things (IOT) devices has created new challenges for the field of botnet detection. Given that there are billions of linked devices globally, the potential possibility of a few devices becoming infected by botnet viruses may have catastrophic consequences. Traditional botnet detection techniques are challenged by the magnitude and variety of IOT networks, highlighting the need for innovative solutions. In this discipline, deep learning methods for identifying botnets have grown in popularity. Effectively identifying and minimising the existence of botnets is the difficulty of botnet detection in IOT networks.

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267>

Deep learning methods, which automatically extract pertinent information from raw packets, address this issue [5]. However, since botnets continue to evolve and adopt sophisticated evasion techniques, detection algorithms must be able to update and adapt in order to accurately identify new botnet behaviours.

II. LITERATURE SURVEY

"Towards the creation of a realistic botnet dataset for network forensic analytics in the Internet of Things: Bot-IoT dataset,"

B. Turnbull, E. Sitnikova, N. Koroniotis, and N. Moustafa,

As IoT systems proliferate, malevolent third parties have begun to attack them. Realistic protection and investigative countermeasures, such network intrusion detection and network forensic systems, must be successfully created in order to meet this challenge. For this reason, training and confirming the legitimacy of the algorithms depend heavily on a representative and well-structured dataset. Despite the existence of several network datasets, the Botnet situations that were used are often not well described. This study suggests a novel dataset, dubbed Bot-IoT, that includes both real and fake IoT network traffic in addition to different kinds of assaults. We also provide a realistic testbed environment to address the shortcomings of current datasets, including accurate labelling, collecting full network information, and the variety of recent and complicated attacks. Lastly, in comparison to the benchmark datasets, we assess the Bot-IoT dataset's dependability for forensic purposes using various statistical and machine learning techniques. The foundation for enabling botnet detection across IoT-specific networks is provided by this study. You may obtain the Bot-IoT dataset at Bot-IoT (2018) [1].

"Evaluating hostile assaults on deep learning for intrusion detection in Internet of Things networks,"

A. Matrawy, O. Shafiq, and O. Ibitoye,

Although adversarial assaults have been extensively researched in computer vision, little is known about how they affect network security applications. Security occurrences and incidents on IoT networks have escalated as IoT, 5G, and AI continue to combine to realise the promise of the fourth industrial revolution (Industry 4.0). Many of these security concerns to IoT networks are being identified and mitigated via the use of deep learning methods. The classification of intrusion threats in Internet of Things networks has made extensive use of feed-forward neural networks (FNN). In this study, we examine the Self-normalizing Neural Network (SNN), a variation of the FNN, and evaluate how well it performs in identifying intrusion threats in an Internet of Things network. The BoT-IoT dataset from the UNSW Canberra Cyber center's Cyber Range Lab is used for our investigation. Based on a variety of performance criteria, including accuracy, precision, and recall, as well as multi-classification metrics like Cohen Cappa's score, our experimental findings show that the FNN performs better than the SNN for intrusion detection in IoT networks. A bright future in the pursuit of safer and more secure deep learning in IoT networks is shown by the SNN's superior resistance against the adversarial samples from the IoT dataset when assessed for adversarial robustness.

"A deep learning method for detecting botnets with raw network traffic data,"

M. Rezvani, H. Mashayekhi, and M. Shahhosseini,

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267>

One of the biggest risks to cybersecurity in recent years is thought to be botnets. Despite extensive research, botnets are always changing, growing more complex and resistant to detection techniques. Current methods of detecting botnets sometimes involve human feature engineering or packet content analysis, which compromises user privacy. This method is seldom documented for the large ISCX botnet dataset, despite the fact that several research employ raw packet bytes for botnet detection. The deep learning-based network traffic analyser for botnet identification that we provide in this study automatically derives useful characteristics from unprocessed packet data. Only the headers of a flow's first few packets contain the raw data. The suggested method delivers early detection of fraudulent traffic, protects user privacy, and lowers the costs associated with human feature engineering. In order to create four distinct flow signatures, we further enhance the raw data using field correlations and packet temporal information. The ISCX botnet dataset, which includes novel botnet kinds in its test data, is used for the assessments. By comparing the performance of the suggested method with a number of feature-based techniques, we demonstrate the efficacy of botnet identification based on raw data. According to the assessment findings, the suggested method performs better than a number of cutting-edge research using the same dataset and has a high classification accuracy of 97.13% for network traffic.

"BoTShark: A deep learning method for detecting botnet traffic,"

A. Dehghantanha, R. Khayami, S. Homayoun, M. Ahmadzadeh, S. Hashemi,

Bot malware is always evolving and looking to take advantage of new attack avenues and get beyond current defences, even though botnets have been well researched. Advanced tactics used in recent botnets are unlikely to be successfully countered by intrusion detection systems currently in use. Botnet Traffic Shark (BoTShark), a deep learning-based botnet traffic analyser, is proposed in this chapter. BoTShark avoids intrinsic constraints like the inability to handle encrypted payloads since it just exploits network transactions and is not reliant on the deep packet inspection approach. Additionally, this enables us to find correlations between the initial characteristics and extract additional features in a cascade fashion in each layer of an autoencoder or a convolutional neural network (CNN). Additionally, we use a Softmax classifier as the predictor to effectively identify fraudulent traffic.

III. SYSTEM ANALYSIS & DESIGN EXISTING SYSTEM

Cyber security is an essential responsibility to identify and mitigate one of the largest threats to internet-connected devices. Botnets are groups of compromised computers controlled by a master host and used for nefarious activities such as data theft, DDoS attacks, and spam dissemination. Unknown botnets, encrypted communications, and sophisticated evasion tactics used by attackers are difficult to detect using conventional botnet detection methods like anomaly-based identification and signature-based approaches. However, a more promising approach to resolving these issues is provided by deep learning techniques. The first people to use machine learning for botnet traffic identification were the authors [10]. To that end, they used a CNN model. According to experimental data, the training set's accuracy is 98.62%, its loss is 4.74%, and each epoch of training takes an

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267>

average of 32 seconds. For the test set, the accuracy is 99.57%, the loss is 1.74%, and the test time is 10 seconds per epoch.

DDoS assaults have been one of the biggest and most damaging online crimes. One of the most well-known examples of an IoT-based DDoS attack was the Mirai botnet. In a denial-of-service (DDoS) assault, a hacker temporarily subjugates many compromised computers in order to target a single target. The hacker then makes multiple requests for a particular service to a server, overloading the server and persuading it to ignore legitimate requests from end users. Mirai is a piece of malware that infects IoT devices in order to generate and spread a network of robots (botnets) composed of the compromised IoT devices (bots). Using a command and control (C&C) server, the attacker (the "botmaster") then directs the bots to participate in DDoS assaults on Internet targets. A bidirectional LSTM (BLSTM-RNN) method for botnet attack detection was introduced in the study [11]. The model was compared to a unidirectional LSTM-RNN in order to ascertain if the BLSTM-RNN's integration of contextual input from the past and future may result in increased accuracy. The results for Mirai, UDP, and DNS were very promising, with validation accuracy of 99%, 98%, and 98% with validation loss metrics of 0.000809, 0.125630, and 0.116453, respectively.

The three distinct sets of DS1 data were subjected to the study [12], which included classifiers such k-nearest neighbours (KNN), decision tree (DT), AdaBoost (AB), random forest (RF), linear SVM (LSVM), and radial basis function SVM (RSVM). The Naive Bayes (NB) and logistic regression (LR) classifiers produced noticeably inferior results. Using the CICIDS 2017 dataset, the authors in [13] integrated the CNN-LSTM model to detect DDoS assaults. The accuracy, precision, and dependability are 97.16%, 97.41%, and 99.1%, respectively. In [14], an attack using a domain creation technique is examined, and 94.9% accuracy is noted. Cost minimisation is a key component of cyber

security as putting countermeasures in place for cyberattacks is expensive. The suggested model in [15] beat state-of-the-art IoT botnet detection methods in terms of accuracy while using just 25% of the implementation expenditure. This results in a nearly 75% reduction in the implementation expenditure. CNN and LSTM are used in the study [16] to identify botnet attacks. The average training accuracy for DNNBoT1 and DNNBoT2 was 90.71% and 91.44%, respectively, compared to the average validation accuracy of 90.54% and 91.24%. Deep learning techniques for botnet identification have become more popular in recent years. Botnets, which are networks of hacked hosts utilised by a master host to carry out malicious actions, pose a severe danger to cyber security. Popular deep learning techniques, such as their ensembles and hybrid approaches, may be used to successfully handle the range of cyber security challenges [17].

The ANN is a comprehensive, practical method for learning vector-, discrete-, and real-valued functions. The CTU-13 dataset was used in the study [18] to evaluate an ANN model that was presented. With an accuracy of almost 99%, the model outperforms support vector machines (SVM) and NB. In a similar vein, [19] uses an ANN to automatically detect DDoS assaults. The results showed that ANN in particular was more effective against DDoS assaults and had a very excellent accuracy of 99%.

To identify botnet attacks, several CNN models are also used. In [20], a CNN-LSTM model is used to identify attacks in IoT environments, classify network behaviour, and stop it by cutting off wifi connections. While LSTM is used for detection, CNN layers are used to extract features from the input data. With an F1 score of 100% and a specificity of 93%, the authors claim positive outcomes. The findings show the creative effects of using the CNN-LSTM model to the analysis of flood attacks, fuzzing attacks, and ordinary packets. The following were the weighted average results of the author's

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267>

recommended method for locating the botnet on the Provision PT-737E camera: 88% for F1 score, 87% for recall, and 88% for camera accuracy. The system's classification results for botnet attacks and normal packets on the Provision PT-838 camera were 89% for recall, 85% for F1 score, and 94% for accuracy. [21]

Disadvantages

- 1) Artificial neural networks (ANN), convolutional neural networks (CNN), and machine learning models are not stacked in the system.
- 2) The technique for identifying distributed denial of service (DDoS) assaults is not implemented by the system.

PROPOSED SYSTEM

In order to strengthen security protocols for Internet of Things systems, this study suggests a stacking model ACLR for botnet attack detection. The suggested approach makes use of the advantages of ML models, recurrent neural networks (RNN), convolutional neural networks (CNN), and artificial neural networks (ANN).

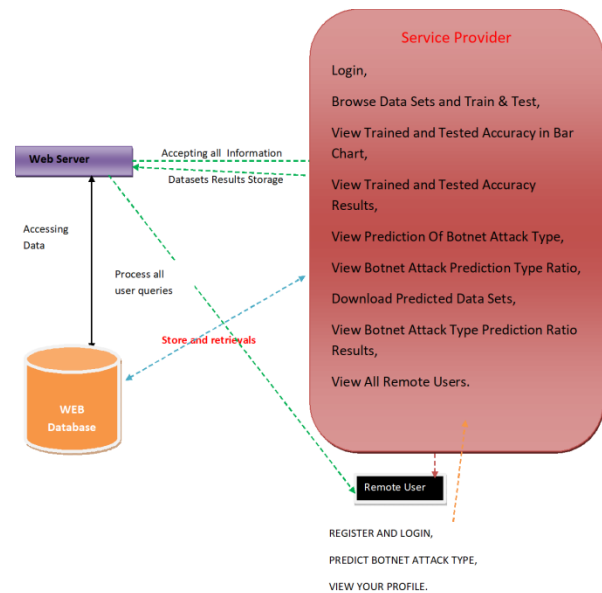
- Various attack kinds, such as worms, backdoors, shell code, fuzzers, DoS, reconnaissance, and analysis, as well as common ones like normal and generic, are classified in the experiments. The dataset is preprocessed for experiments, which includes label encoding for categorical values required for machine learning model training and null value removal.
- The effectiveness of the suggested method is carefully evaluated using a broad range of commonly accepted performance assessment criteria, including accuracy, precision, recall, and the F1 score. The performance is fully validated using k-fold cross-validation, with k values of 3,5, 7, and 10, to further strengthen the findings' durability. Additionally, the receiver operating characteristic area under the curve (ROC-AUC) measure is used to assess the model's discriminative capacity. Additionally, a

performance comparison with the most advanced models is conducted.

Advantages

- Preprocessing must be completed to make the data ready for the model training and testing. The dataset are loaded, cleaned, modified, and transformed into a form that is appropriate for machine learning models.
- An advanced form of an ANN is CNN, which was created to be particularly effective at processing and analyzing visual data, such as pictures and movies. CNNs are very good at extracting significant patterns and characteristics from datasets.

SYSTEM ARCHITECTURE



IV. IMPLEMENTATIONS

Modules

Service Provider

The Service Provider must use a working user name and password to log in to this module. Following a successful login, he may do several tasks including browsing data sets and training and testing. See the results of the trained and tested accuracy, the bar chart showing the accuracy, the prediction of the kind of botnet attack, the prediction type ratio, and the predicted

https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267

IP	Port	Protocol	...
151.101.21.140	80	HTTP	...
151.101.21.140	80	HTTP	...
151.101.21.140	80	HTTP	...

PREDICTION OF BOTNET ATTACK TYPE

Enter All Datasets Details Here !!!

Enter IP	151.101.21.140-104.0.0	Enter SourceIp	SarkHainMessage
Enter Port	1,316-17	Enter SourceIp	203.106.149.10
Enter SourcePort	8019	Enter SourceIp	AS4788
Enter TargetIp	204.95.99.31	Enter TargetPort	
Enter Payload		Enter SourceIpCountryCode	
Enter SourceIpRegion		Enter SourceIpCity	
Enter SourceIpLatitude		Enter SourceIpLongitude	
Enter SourceIpAlphaCode		Enter SourceIpAlphaCode	
Enter HttpRequest			

Predict

Prediction Of Botnet Attack Type →

VSRRMS

V. CONCLUSION

Recently, there has been an increase in the frequency and severity of cyberattacks, and botnet assaults have surfaced with the ability to do significant harm. Automated botnet detection has shown the promise of deep learning-based models; ensemble models outperform individual models in terms of prediction. For botnet identification, this study suggests ANN+CNN+LSTM+RNN (ACLR), a hybrid stacking model. The UNSW-NB15 dataset is used to build ACLR in the Google COLAB environment as part of the experimental setup. ANN, CNN, LSTM, and RNN models were also used in this study to compare their performance to the suggested ACLR model. With an accuracy score of 0.9698, the experimental data indicate that ACLR performs better than k-fold cross-validation, which has an accuracy score of 0.9749 when k is 3, 5, 7, and 10, respectively.

Furthermore, it has been shown that improved performance is obtained by increasing the number of layers in the deployed models; however, this comes at the expense of more computational complexity and longer training times. LSTM, RNN, and CNN perform better than the ANN model, which performs poorly among the models used. The comparison results show that the suggested method performs better and is more accurate for botnet identification than ANN, CNN, LSTM, and RNN models. The proposed ACLR model has the highest ROC AUC (0.9934) and PR AUC (0.9950) values when compared to earlier models. ACLR can outperform state-of-the-art models, according to performance study using current models. It is crucial to remember that deep learning algorithms for botnet detection still have drawbacks, including the potential for hostile assaults and a lack of labelled training information. More study and advancement in this field are needed to improve the accuracy, scalability, and resilience of deep learning-based botnet detection systems. Based on four deep learning models, the stacking model in the proposed study produces more effective outcomes than a single model while using more time in predictions. Data synchronisation and exchange are also required. This illustrates how important it is to properly balance model efficacy and complexity across a variety of applications. Future research will be entirely automated, thus additional training using reinforcement learning—which can be more successful—should be conducted.

REFERENCES

[1] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset," *Future Gener. Comput. Syst.*, vol. 100, pp. 779–796, Nov. 2019.

[2] O. Ibitoye, O. Shafiq, and A. Matrawy, "Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks,"

<https://doi.org/10.62643/ijerst.2025.v21.i2.pp1259-1267>

in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2019, pp. 1–6.

[3] M. Shahhosseini, H. Mashayekhi, and M. Rezvani, “A deep learning approach for botnet detection using raw network traffic data,” *J. Netw. Syst. Manage.*, vol. 30, no. 3, p. 44, Jul. 2022.

[4] S. Homyoun, M. Ahmadzadeh, S. Hashemi, A. Dehghantanha, and R. Khayami, “BoTShark: A deep learning approach for botnet traffic detection,” in *Cyber Threat Intelligence*, 2018, pp. 137–153.

[5] M. Ge, X. Fu, N. Syed, Z. Baig, G. Teo, and A. Robles-Kelly, “Deep learning-based intrusion detection for IoT networks,” in Proc. IEEE 24th Pacific Rim Int. Symp. Dependable Comput. (PRDC), Dec. 2019, p. 256.

[6] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, “Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,” *J. Inf. Secur. Appl.*, vol. 50, Feb. 2020, Art. no. 102419.

[7] 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 Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 2952–2963, Sep./Oct. 2023.

[8] D. T. Son, N. T. K. Tram, and P. M. Hieu, “Deep learning techniques to detect botnet,” *J. Sci. Technol. Inf. Secur.*, vol. 1, no. 15, pp. 85–91, Jun. 2022.

[9] M. Gandhi and S. Srivatsa, “Detecting and preventing attacks using network intrusion detection systems,” *Int. J. Comput. Sci. Secur.*, vol. 2, no. 1, pp. 49–60, 2008.

[10] J. Liu, S. Liu, and S. Zhang, “Detection of IoT botnet based on deep learning,” in Proc. Chin. Control Conf. (CCC), 2019, pp. 8381–8385.

[11] C. D. McDermott, F. Majdani, and A. V. Petrovski, “Botnet detection in the Internet of Things using deep learning approaches,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.

[12] S. Sriram, R. Vinayakumar, M. Alazab, and K. Soman, “Network flow based IoT botnet attack detection using deep learning,” in Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), Jul. 2020, pp. 189–194.

[13] B. Nugraha, A. Nambiar, and T. Bauschert, “Performance evaluation of botnet detection using deep learning techniques,” in Proc. 11th Int. Conf. Netw. Future (NoF), Oct. 2020, pp. 141–149.

[14] P. Karunakaran, “Deep learning approach to DGA classification foreffective cyber security,” *J. Ubiquitous Comput. Commun. Technol. (UCCT)*, vol. 2, no. 4, pp. 203–213, 2020.

[15] N. Elsayed, Z. ElSayed, and M. Bayoumi, “IoT botnet detection using an economic deep learning model,” 2023, arXiv:2302.02013.