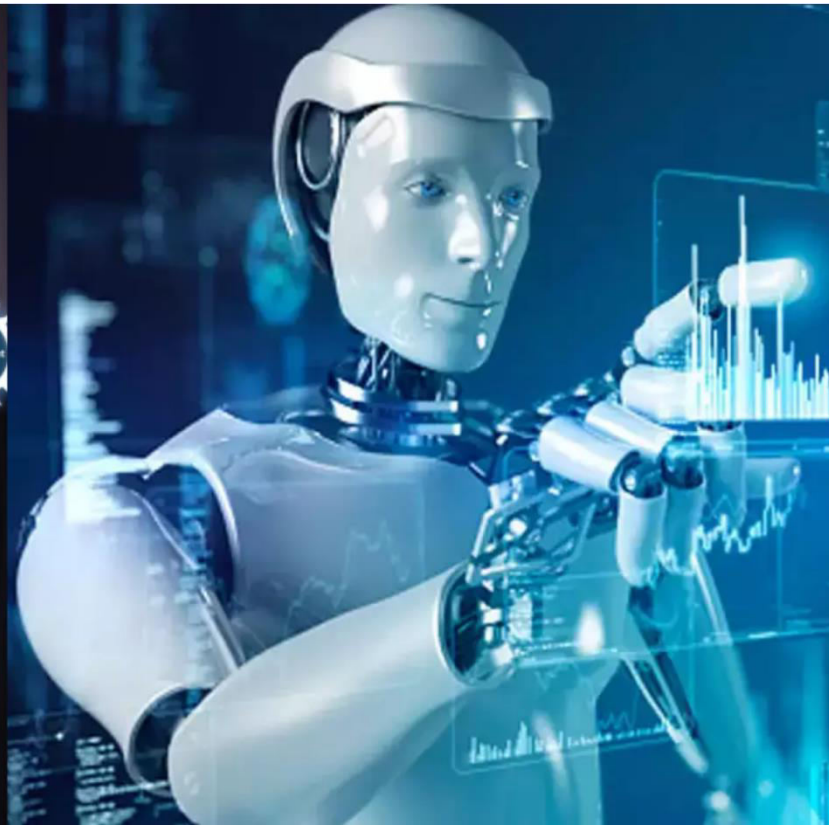
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

Efficient Botnet Attack Detection using a Hybrid Machine Learning Approach

D.Vetriselvi, Garikapati Hemanth Sharma, Geda Andeep Kumar, Gangavarapu Kishore Babu, Gangineni Venkata Venu

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

Department of CSE, Bharath Institute of Higher Education and Research, Chennai, India

ABSTRACT: The advent of Internet of Things (IoT) gadgets brought about a fast increase in cyber-attacks, where botnets are identified as the most hazardous that employ not only complex but also adaptive attack strategies. The detection of these types of attacks is really tough because of IoT devices' diversity to mention nothing about the evolution of malware that enables them to change their patterns. This paper introduces a robust botnet detection model based on the UNSW-NB15 dataset. Feature selection based on Mutual Information is utilized to choose the most relevant features which in turn helps in detection process acceleration. The use of advanced machine learning techniques in hybrid such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), for spatial-temporal dependency extraction in the network traffic data is presented as a new concept. From the results of the studies, the CNN-LSTM-GRU hybrid model achieves the highest accuracy rate thus confirming the capability of exploiting it to the maximum extent for the task of botnet activity detection. This study is a solution with the ability to be used across different devices, and it is accurate in detecting and eliminating botnet-related security challenges thus settling complex attack patterns at the epicenter of the issue.

"KEYWORDS: Machine learning, hyperparameter optimization, classification, cyberattacks, intrusion detection".

I. INTRODUCTION

The Internet is everywhere now, and it shapes our daily lives, social interactions, cultural activities, and how organizations operate. But, like all technologies, it can also be misused. People and companies can become targets for money theft or having their personal information stolen. A significant issue in cybersecurity today is botnet attacks. These are a big concern because they change easily and use complex methods to attack. It's crucial to find and stop these attacks to protect against security problems like DDoS attacks, stealing data, and spreading viruses. Recognizing botnets early helps reduce damage, maintains network efficiency, and ensures user trust in online services. It also supports strength against cyber threats, meeting legal requirements, and developing new strategies in the fight against cyber dangers. Deep learning technology is becoming very popular because it can quickly spot and identify harmful botnet actions in network data. It works by finding important features in raw network data. Tools like Convolutional Neural Networks (CNNs) and autoencoders are useful because they notice spatial connections and can discover patterns, even in new types of botnets. Another tool, the Bidirectional Long Short-Term Memory Recurrent Neural Network (BLSTM-RNN), is good at recognizing trends in network data by looking at both past and future information, making it ideal for finding botnets.

With the rise of IoT devices, finding botnets is even harder due to the vast and varied nature of these networks. Older detection methods struggle with this, so more advanced solutions are needed. Deep learning is a solid approach, as it adapts to changing botnet tactics and can handle large amounts of data quickly. These methods can discover new botnet behaviors and learn to counter more sophisticated tricks. Deep learning's ability to manage complex and ever-changing IoT traffic makes it the best choice for large-scale, real-time botnet detection.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

II. RELATED WORK

The Internet has changed how people and businesses communicate. However, this change has also brought the danger of cyber-attacks, especially botnet attacks. Botnets are groups of infected devices controlled remotely and used for harmful activities like DDoS attacks, data theft, and spreading malware. As the Internet of Things (IoT) grows, these attacks become harder to detect, making it crucial to find better detection methods. Recently, deep learning techniques have proven effective in identifying and handling botnet attacks with greater accuracy and speed than older methods.

A big challenge in detecting botnets is their changing nature. Attackers often update their methods, so older detection systems can miss new threats. This has led to the increased use of machine learning and deep learning, which can quickly learn to recognize new botnet behavior and detect unusual patterns. Various deep learning models, like CNNs, RNNs, and autoencoders, are employed to analyze large amounts of network data to find signs of botnet activity.

In one study, Hasan et al. proposed using a hybrid deep learning method for securing industrial IoT systems from botnet attacks. They used CNNs together with LSTM networks to improve how well they could detect botnet traffic in IoT environments. CNNs help analyze spatial features, while LSTMs focus on identifying time-related patterns in network traffic. This hybrid approach performed better than traditional machine learning methods and showed deep learning's potential in protecting IoT networks from complex botnet attacks. Similarly, Popoola et al. used a model combining CNNs and LSTMs for identifying botnet activity in IoT networks. By capturing both spatial and time-related patterns, it achieved more accurate detection, even in large networks with diverse traffic.

Other deep learning techniques have also been tested for detecting botnets. Son et al. used CNNs and autoencoders to identify botnet traffic in typical network settings. CNNs effectively captured traffic patterns associated with botnet behavior, while autoencoders helped find anomalies and outliers in the data. Since autoencoders learn to reproduce input data, they are particularly useful for identifying previously unseen botnet traffic. This research highlighted deep learning techniques' versatility in detecting various botnet attacks, including DDoS and data exfiltration.

Another significant contribution was made by Haq and Khan, who introduced a deep neural network (DNN) model for botnet detection and classification. Their DNNBoT model used a multi-layer DNN to understand complex patterns in network traffic data, achieving high detection accuracy. Using a large dataset, it identified many botnet behaviors, making it suitable for real-time detection in dynamic environments. This approach emphasized the need for deep learning models with many layers to capture complex relationships in network data, essential for detecting advanced botnet attacks.

Detecting botnet activity in IoT networks is crucial due to the increasing number of connected devices. IoT presents unique challenges for conventional security systems, involving numerous different devices with varying capabilities. Alzahrani and Bamhdi tackled these challenges by developing a hybrid deep learning model for IoT networks. Their approach combined a CNN with a fully connected network to detect botnet attacks, demonstrating strong performance in identifying botnet traffic in IoT settings. This illustrates how deep learning can manage the complex and dynamic nature of IoT traffic.

Furthermore, research by Ahmed et al. explored using deep learning for botnet detection, proposing a classification model based on deep neural networks. This model aimed to classify network traffic into botnet and non-botnet categories, improving detection accuracy and reducing false positives. This demonstrates deep learning's potential in enhancing cybersecurity across various network environments.

III. MATERIALS AND METHODS

This document outlines a system to detect botnets in IoT setups using sophisticated techniques. The system utilizes several machine learning tools to achieve its goals. It begins with selecting crucial attributes from network traffic data by using Mutual Information, which helps the system focus on the most important details to improve efficiency. The system employs different machine learning tools like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Artificial Neural Networks (ANN). These are combined in various ways, such as CNN + LSTM and RNN + ANN, to better understand the patterns in network traffic. Additional

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

models like CNN + LSTM + GRU and CNN + BiLSTM + BiGRU are also evaluated to address more complex data patterns. Integrating these approaches offers enhanced botnet attack detection and adaptability to new IoT threats.

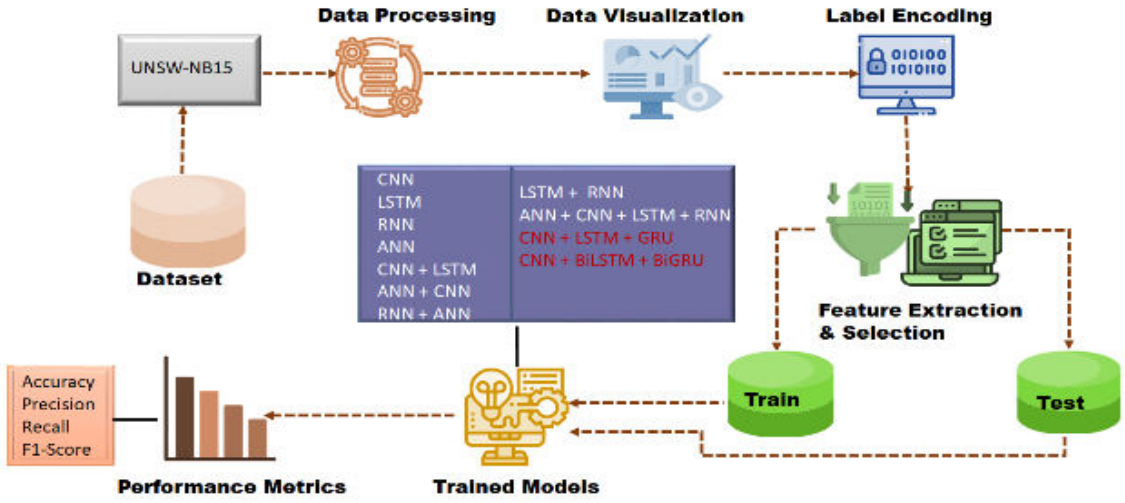


Figure 1: Proposed Architecture

Figure 1 displays a method using deep learning for intrusion detection, relying on a dataset called UNSW-NB15. The procedure starts with preparing and visualizing the data, converting it into a format suitable for machine learning through label encoding. Important features are chosen to focus on key data points. Deep learning models, including CNN, LSTM, and RNN, plus their combinations, are trained and assessed. Their effectiveness is judged on accuracy, precision, recall, and F1-score. Once trained, these models can spot new intrusions in network traffic.

i) Dataset Collection:
We use the UNSW-NB15 dataset from Kaggle for analyzing network behavior. Originally collected by the University of New South Wales (UNSW), this dataset has been used widely in network security research, including IoT. It serves as a testbed for systems that detect various network threats, even those aimed at IoT devices. Using Mutual Information-based feature selection, nine features were picked: 'sbytes', 'dbytes', 'rate', 'dinpkt', 'tcprtt', 'synack', 'ackdat', 'smean', and 'dmean'. These are crucial for making the botnet detection more effective.

	dur	proto	service	state	spkts	dpkts	sbytes	dbytes	rate	sload
0	0.121478	tcp	-	FIN	6	4	258	172	74.087486	14158.942383
1	0.649902	tcp	-	FIN	14	38	734	42014	78.473373	8395.112305
2	1.623129	tcp	-	FIN	8	16	364	13186	14.170161	1572.271851
3	1.681642	tcp	ftp	FIN	12	12	628	770	13.677108	2740.178955
4	0.449454	tcp	-	FIN	10	6	534	268	33.373825	8561.499023

5 rows × 36 columns

Figure 2: Dataset Collection Table – UNSW-NB15

ii) Pre-Processing:
Pre-processing is essential for making the dataset ready for machine learning, boosting performance and precision. This has several key steps:

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

a) Data Processing: Duplicate entries are removed to ensure each dataset entry is distinct. We clean the dataset by dropping unnecessary columns and handle missing values to maintain consistency and readiness for model training.
b) Data Visualization: This process helps reveal patterns and trends within the dataset. Using graphs like histograms and scatter plots, we analyze the data structure and detect any anomalies. This understanding guides further data processing efforts.
c) Label Encoding: Categorical data is converted into numerical format so that machine learning models can process it. Each data category is given a unique numerical value, enhancing model efficiency and accuracy, mainly for classification purposes.
d) Feature Extraction: Relevant input (X) and output (y) variables are selected for training models. Input variables aid in prediction, while the output variable is the target of prediction. Proper feature extraction ensures that models train on the most informative data.
e)Feature Selection:
In feature selection, we identify the most important parts of our data. We use a method called Mutual Information (MI) to see how each part relates to the final prediction target. This helps us get rid of information that doesn't help much, making the model work better. We focus on the parts with the highest MI scores since they are crucial for accurate predictions.

iii)Training & Testing:
We split our data into two parts: one for training and one for testing. We usually put 80% of the data in the training section and 20% in the testing section. The training data helps the model learn, while the testing data lets us see how well the model can predict new information. This step is important to make sure the model works well in real-world applications.

Algorithms:
CNNs: These models are great for working with images but are also useful for network traffic analysis. They can detect unusual patterns, such as botnet activities, by examining behaviors in smart device networks.
LSTMs: These models handle sequences expertly, remembering important information over time. We use them to detect repeating patterns or odd behavior in smart device traffic that might signal an attack.
RNNs: These models focus on processing data sequences and keeping track of prior information. In our work, they help identify botnet attacks by understanding shifts or repeated patterns in network traffic.
ANNs: These models excel in recognizing complex data patterns. In our project, they help distinguish between normal and botnet-related network activities by processing inputs and making accurate classifications.

Hybrid Models:
CNN + LSTM: This combination benefits from CNN's knack for extracting meaningful features and LSTM's ability to track sequences over time. It enhances the detection of botnet attacks by considering detailed patterns and temporal behaviors.
ANN + CNN: This blend uses CNN for feature extraction and ANNs for classification tasks, enhancing the ability to recognize whether network traffic is normal or linked to botnet activities.
RNN + ANN: In this setup, RNN analyzes the sequence of data, while ANN handles classification, which helps better identify traffic associated with botnets.
LSTM + RNN: This pairing captures both short and long-term patterns in time-series data, making it useful for catching botnet activities in smart device networks.
ANN + CNN + LSTM + RNN: This comprehensive approach takes advantage of CNN's, RNN's, LSTM's, and ANN's respective strengths, providing a robust framework for detecting botnets by evaluating both feature details and timing aspects.
CNN + LSTM + GRU: Together, these components improve how models learn and predict botnet behaviors, resulting in better detection in smart device data.
CNN + BiLSTM + BiGRU: This sophisticated model captures patterns from both directions of data flow for a thorough investigation, enhancing botnet behavior detection accuracy in network traffic.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

IV. RESULTS & DISCUSSION

Accuracy:
Accuracy tells us how well a test can correctly identify patients and healthy people. To figure out accuracy, we calculate how many true positive (TP) and true negative (TN) cases there are compared to all cases checked. The formula is:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} (1)$$

Precision:
Precision measures how many of the cases classified as positive are actually correct. It's the number of true positive cases divided by all cases marked as positive, both true and false. The formula is:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)$$

Recall:
Recall is a term used in machine learning to see how well a model finds all the true positive cases of a specific class. It is the ratio of true positives to all actual positive cases. This gives us an idea of the model's ability to capture all the right instances.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}} (3)$$

F1-Score:
The F1 score helps us understand how accurate a model is by combining both precision and recall. It shows how often a model makes correct predictions when looking at the entire dataset. The formula for the F1 score is:

$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100 (4)$$

In Table (1), you'll find the performance of each algorithm measured by accuracy, precision, recall, and F1-Score. This helps to compare how each algorithm performs in identifying correct cases.

ML Model	Accuracy	Precision	Recall	F1_score
ANN	0.896	0.885	0.913	0.899
CNN	0.854	0.251	0.140	0.177
RNN	0.856	0.859	0.855	0.855
LSTM	0.859	0.792	0.924	0.843
CNN+LSTM	0.874	0.874	0.874	0.874
CNN+ANN	0.795	0.510	0.993	0.673
RNN+ANN	0.719	0.519	0.465	0.465
RNN+LSTM	0.898	0.919	0.861	0.888
CNN+RNN+LSTM+ANN	0.892	0.895	0.889	0.892
CNN+LSTM+GRU	0.889	0.888	0.889	0.889
CNN+BiLSTM+BiGRU	0.879	0.880	0.879	0.879

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

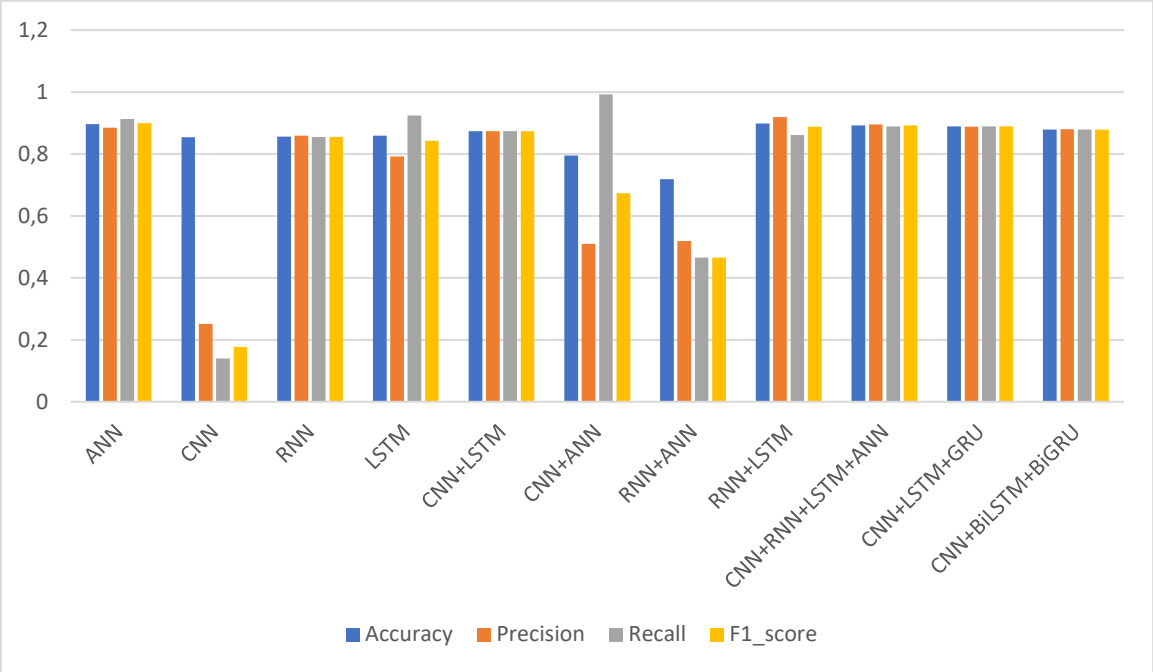
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

Graph.1 Comparison Graphs



Accuracy is represented in light blue, precision in orange, recall in grey, and F1-Score in light yellow, ***Graph (1)***.

V. CONCLUSION

In the past few years, cyber-attacks have become more common and serious. Botnet attacks are especially dangerous for network security, particularly when it comes to devices connected to the Internet of Things (IoT). Spotting these attacks is getting tougher because malware keeps changing, and there are many different types of IoT devices. This study demonstrates that using advanced hybrid machine learning models can help address these challenges effectively. The research utilized the UNSW-NB15 dataset and used Mutual Information to choose the key features of network traffic, which improves the attack detection process. Among the different models tested, the combination of Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) proved most effective. It achieved the best accuracy in detecting botnet attacks. This hybrid model is strong because it captures both fixed and changing elements in network traffic, making it powerful for detecting botnet activities. The study's results show that hybrid machine learning approaches can significantly improve the accuracy and efficiency of detecting botnets in complex IoT environments.

Looking ahead, the system can be further improved by exploring advanced ways to extract features, such as statistical or time-domain methods, to enhance model accuracy. Trying out different combination methods, like stacking or boosting, could also sharpen the detection skills. Including attention mechanisms in deep learning models might help spot crucial patterns in IoT traffic. Additionally, using transfer learning with pre-trained models could help detect new types of botnets without needing vast real-world datasets for training. These strategies could lead to a stronger, more adaptable, and more efficient solution for detecting botnets in IoT environments, helping to tackle new threats and challenges effectively.

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.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404050|

[2] O. Ibitoye, O. Shafiq, and A. Matrawy, "Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks," 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. Homayoun, 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] N. Elsayed, Z. ElSayed, and M. Bayoumi, "IoT botnet detection using an economic deep learning model," 2023, arXiv:2302.02013.
[10] M. A. Haq and M. A. Rahim Khan, "DNNBoT: Deep neural network-based botnet detection and classification," Comput., Mater. Continua, vol. 71, no. 1, pp. 1729–1750, 2022.
[11] A. A. Ahmed, W. A. Jabbar, A. S. Sadiq, and H. Patel, "Deep learning-based classification model for botnet attack detection," J. Ambient Intell. Humanized Comput., vol. 13, no. 7, pp. 3457–3466, Jul. 2022.
[12] M. Y. Alzahrani and A. M. Bamhdi, "Hybrid deep-learning model to detect botnet attacks over Internet of Things environments," Soft Comput., vol. 26, no. 16, pp. 7721–7735, Aug. 2022.
[13] S. I. Popoola, B. Adebisi, M. Hammoudeh, G. Gui, and H. Gacanin, "Hybrid deep learning for botnet attack detection in the Internet-of-Things networks," IEEE Internet Things J., vol. 8, no. 6, pp. 4944–4956, Mar. 2021.
[14] S. I. Popoola, B. Adebisi, M. Hammoudeh, H. Gacanin, and G. Gui, "Stacked recurrent neural network for botnet detection in smart homes," Comput. Electr. Eng., vol. 92, Jun. 2021, Art. no. 107039.
[15] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Proc. MilCIS, Nov. 2015, pp. 1–6.
[16] 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.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details