International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online) Volume 15, Issue 09, September 2026)

AI based Techniques for Botnet Attack Detection in IoT Networks

Sakshi Gautam
Department of CSE, Lakshmi Narain College of Technology and Science, Bhopal, M.P., India

Khushbu Rai
Department of CSE, Lakshmi Narain College of Technology and Science, Bhopal, M.P., India

Bhupesh Gour
Department of CSE, Lakshmi Narain College of Technology and Science, Bhopal, M.P., India

Abstract : **The rapid growth of Internet of Things (IoT) networks has increased security risks associated with compromised devices and botnet attacks, including distributed denial-of-service (DDoS), malicious traffic generation, and unauthorized network access. Traditional signature-based detection methods often face limitations in identifying evolving and previously unseen botnet behaviours. Artificial intelligence (AI), including machine learning (ML), deep learning (DL), and hybrid learning techniques, provides promising approaches for detecting and classifying malicious network activities. This review presents a systematic analysis of AI-based techniques for botnet attack detection in IoT networks, focusing on detection architectures, benchmark datasets, traffic features, preprocessing methods, and evaluation metrics. It examines conventional ML algorithms, deep neural networks, hybrid models, anomaly detection, federated learning, and lightweight edge-based approaches. The review also investigates key evaluation challenges, including class imbalance, data leakage, cross-dataset generalization, detection of unseen attacks, explainability, and computational efficiency.**

IndexTerms – IoT Security, Botnet Detection, Machine Learning, Deep Learning, Intrusion Detection, Network Security.

I. INTRODUCTION

The Internet of Things (IoT) has become an important part of modern digital infrastructure, connecting devices used in healthcare, transportation, industrial automation, smart homes, and energy management. Although these connected devices improve communication and operational efficiency, their widespread deployment also increases the potential attack surface of computer networks. Many IoT devices operate with limited computing resources, inadequate security configurations, weak authentication mechanisms, and irregular software updates[1]. These weaknesses can allow attackers to compromise devices and use them for malicious activities. Among the major threats, IoT botnets are particularly concerning because they enable attackers to coordinate multiple infected devices to perform distributed denial-of-service (DDoS) attacks, scan networks, and generate malicious traffic. The growing diversity of IoT devices therefore creates a need for reliable and adaptive security mechanisms capable of identifying suspicious network behaviour [2].

Botnet attacks have demonstrated the potential to disrupt internet services and affect connected systems on a large scale. The Mirai botnet, for example, exploited vulnerable IoT devices and used them to generate large volumes of malicious traffic. Detecting such attacks is challenging because botnet behaviour can vary according to malware families, communication patterns, device types, and attack objectives[3]. Traditional signature-based intrusion detection systems identify threats by matching observed traffic against known attack signatures. Although these systems remain useful for detecting recognized threats, their effectiveness may be limited when attackers modify their behaviour or introduce previously unseen attack patterns. Consequently, researchers have investigated behaviour-based detection approaches that analyse network traffic characteristics and identify suspicious activities without relying exclusively on predefined signatures [4].

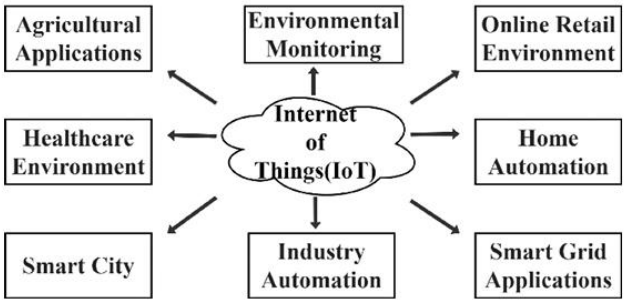


Figure 1: IoT Applications

Artificial intelligence (AI), particularly machine learning (ML), provides opportunities to automate the identification of malicious traffic patterns. Algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbours, and gradient boosting can learn relationships between network traffic features and attack labels. Their effectiveness depends on the quality of the training data, feature representation, class distribution, and validation strategy[5]. Machine learning approaches have demonstrated their potential for identifying suspicious traffic patterns and supporting automated intrusion detection. However, their

performance may vary depending on network conditions, attack characteristics, and the quality of the training datasets. A comprehensive assessment must therefore consider both predictive performance and the limitations of the underlying evaluation procedures [6].

Deep learning (DL) techniques have also received considerable attention because they can learn complex representations from network traffic data. Models such as Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and autoencoders can support different detection tasks, including malicious traffic classification, sequential behaviour analysis, and anomaly detection. Deep autoencoder-based approaches have been investigated for identifying abnormal traffic associated with compromised IoT devices infected by Mirai and Bashlite[7]. These techniques can support automated feature representation and behavioural analysis. However, their effectiveness depends on the diversity of training data, model complexity, and the extent to which experimental environments represent practical deployment conditions. High classification performance on a particular dataset does not necessarily guarantee reliable detection of unfamiliar attacks or traffic from different IoT environments [8].

The availability of representative datasets is another important factor in developing and evaluating AI-based botnet detection systems. Researchers use benchmark datasets to train models, compare algorithms, and investigate the characteristics of malicious network traffic. Commonly used datasets include N-BaIoT, BoT-IoT, IoT-23, and CICIoT2023, which contain different types of benign and malicious traffic collected or generated under varying experimental conditions[9]. These datasets support the evaluation of detection algorithms and provide opportunities to analyse attack behaviour across different IoT environments. However, differences in traffic generation, device configurations, attack labels, and feature distributions can affect reported results. Therefore, meaningful comparisons should consider dataset characteristics, preprocessing methods, validation strategies, and evaluation metrics rather than relying exclusively on classification accuracy [10].

The present paper provides a systematic literature review of AI-based techniques for botnet attack detection in IoT networks. The review examines traditional machine learning, deep learning, hybrid and ensemble learning, anomaly detection, federated learning, and emerging intelligent detection approaches. It focuses on commonly used datasets, network traffic features, preprocessing strategies, evaluation metrics, and experimental validation methods [11]. Particular attention is given to false-positive rates, class imbalance, data leakage, cross-dataset generalization, unseen attack detection, explainability, and computational requirements. Furthermore, the study identifies existing research gaps and discusses future directions for developing robust, scalable, privacy-preserving, and resource-efficient detection systems [12].

II. BACKGROUND

A. Kumar et al. [1] presented an enhanced hybrid deep learning approach for detecting botnet attacks in IoT environments. The study focused on combining deep learning techniques to improve the identification of malicious network traffic and distinguish botnet activities from legitimate communication. The hybrid approach aimed to exploit complementary learning capabilities for improved detection performance.

Dharmaraj et al. [2] introduced a novel hybrid deep learning approach for botnet attack detection in secure IoT environments. The research addressed the need for intelligent detection mechanisms capable of identifying malicious activities within interconnected IoT devices. The hybrid learning strategy was intended to combine the strengths of different deep learning components for analysing network behaviour and classifying potential attacks. The paper is relevant to understanding how hybrid architectures are being investigated for IoT security.

R. K. S. et al. [3] presented an enhanced convolutional neural network approach for adaptive IoT botnet attack detection using reinforcement learning. The study combined CNN-based pattern extraction with reinforcement learning to support adaptive detection under changing network conditions. CNNs can learn representative patterns from suitably structured network traffic, while reinforcement learning can support decision-making based on feedback from the environment. This combination is relevant to botnet detection scenarios in which attack behaviour may evolve over time.

Guo et al. [4] investigated a differential game method for countering DDoS attacks associated with IoT botnets from holistic and dynamic perspectives. Unlike conventional classification-based approaches, the study focused on modelling the strategic interaction between attackers and defenders. Differential game theory can represent how attack and defence decisions evolve over time, helping analyse resource allocation and mitigation strategies. The approach is relevant to IoT security because effective protection may require coordinated responses as attack conditions change.

Sendhil et al. [5] investigated the application of machine learning techniques for detecting botnet attacks in IoT–5G networks. The study addressed the security challenges arising from the integration of IoT devices with fifth-generation communication infrastructure. Machine learning provides a mechanism for analysing network traffic features and identifying patterns associated with malicious activities. This research is relevant because IoT–5G environments require detection methods capable of handling heterogeneous devices and changing communication conditions.

Bhadoriya et al. [6] presented a hybrid deep learning framework for IoT botnet detection using network flow features and autoencoder-based analysis. The approach combines traffic-flow information with representation learning to support the identification of suspicious network behaviour. Autoencoders can learn compact representations of traffic data and potentially identify abnormal patterns, while a subsequent detection component can classify malicious activities. This framework is relevant to developing automated detection systems that analyse network-level characteristics without relying exclusively on packet payloads.

Wang et al. [7] investigated malware propagation in heterogeneous IoT networks by examining propagation

thresholds and cost-effective patching strategies. Their approach considered the differences among IoT devices and communication technologies when modelling the spread of malware. The study focused on identifying the conditions under which malware propagation can occur and determining strategies for reducing the cost of defence. This research complements AI-based botnet detection because understanding propagation behaviour can help security administrators prioritize vulnerable devices and implement preventive measures.

Chen et al. [8] presented a collaborative framework for defending IoT networks against distributed denial-of-service attacks using a Stackelberg game model. The framework examined strategic interactions between network defenders and attackers, considering the role of device identity information and signalling data in detecting and blocking malicious traffic. The game-theoretic approach supports the analysis of coordinated defence decisions and resource allocation among network participants. This work is relevant to botnet security because DDoS attacks often involve multiple compromised devices and require coordinated mitigation.

Wu et al. [9] developed a differential game approach for analysing attack–defence strategies in IoT networks. The researchers modelled interactions between attackers and defenders as a dynamic competition over network resources, with changes in device security states represented through differential equations. Optimal control theory was used to analyse the strategies of the competing parties, while a numerical method was employed to obtain the game equilibrium. The reported numerical analysis demonstrated that the model could characterize the evolution of network security states under strategic interaction.

Zhang et al. [10] investigated the modelling and defence of advanced metering infrastructure against distributed denial-of-service attacks. Advanced metering infrastructure is an important component of smart-grid communication systems, where communication disruptions can affect monitoring and operational functions. The study examined the relationship between DDoS attacks and the behaviour of the targeted infrastructure, with an emphasis on modelling attack impacts and evaluating defensive strategies. Its findings are relevant to IoT security applications involving connected energy meters and networked control devices.

Zhang et al. [11] presented a cybersecurity threat assessment approach integrating qualitative differential analysis and evolutionary game theory. The research addressed the assessment of cyber threats by combining qualitative information with the strategic behaviour of interacting entities. Evolutionary game theory can be used to analyse how attack and defence strategies change according to their associated costs, benefits, and outcomes. This approach offers a broader perspective on security assessment by considering both threat characteristics and strategic responses.

Mahboubi et al. [12] presented a survey of stochastic models for understanding the spread of IoT botnets, drawing on epidemiological models originally developed for malware propagation. The study discussed susceptible–infected–recovered–susceptible (SIRS) and susceptible–exposed–infected–recovered–susceptible (SEIRS) models to represent different stages of malware transmission. The authors used MATLAB and R to illustrate stochastic botnet propagation and examine how malware outbreaks can influence DDoS-related threats.

III. CHALLENGES

AI-based botnet attack detection in IoT networks faces several technical and operational challenges due to the heterogeneous nature of connected devices, evolving attack behaviours, limited computing resources, and the complexity of network traffic. Although machine learning and deep learning techniques can identify malicious traffic patterns, their effectiveness depends on the availability of representative datasets, appropriate feature selection, and reliable evaluation methods. Furthermore, changing botnet strategies, encrypted communications, class imbalance, and differences between training and real-world traffic can reduce detection reliability. Addressing these challenges is essential for developing accurate, adaptive, scalable, and resource-efficient security systems for modern IoT environments.

1. Limited Availability of Representative Datasets: Existing datasets may not adequately represent the diversity of IoT devices, network conditions, and emerging botnet attacks. Models trained on limited datasets may perform poorly in different environments.

2. Detection of Unknown Attacks: AI models trained on previously identified botnet families may fail to recognize zero-day attacks or unfamiliar malicious behaviours. Developing models that generalize to unseen attacks remains a significant challenge.

3. Class Imbalance: Network datasets frequently contain unequal numbers of benign and malicious traffic samples. This imbalance can bias classification models toward majority classes and reduce the detection of less frequent attacks.

4. High Computational Requirements: Complex deep learning models may require substantial processing power, memory, and energy. These requirements can limit deployment on resource-constrained IoT devices and gateways.

5. Encrypted Network Traffic: Encryption protects communication privacy but can restrict access to packet payload information. Detection systems must therefore rely on available metadata, flow statistics, and behavioural characteristics.

6. False Positives and False Negatives: Incorrectly identifying legitimate traffic as malicious can generate excessive alerts, while failing to detect actual attacks can compromise network security. Maintaining an appropriate balance between precision and recall is essential.

7. Changing Attack Behaviour and Concept Drift: Botnet communication patterns and attack strategies may change over time. Detection models can lose effectiveness when the characteristics of current network traffic differ from those observed during training.

8. Privacy, Explainability, and Deployment: Centralized data collection may introduce privacy concerns, while complex models can make their decisions difficult to interpret. Integrating privacy-preserving learning, understandable alerts, and real-time detection into existing IoT infrastructure remains challenging.

IV. AI TECHNIQUES

A. MACHINE LEARNING TECHNIQUES

Random Forest (RF): Random Forest is a supervised machine learning algorithm that combines multiple decision trees to classify network traffic as benign or malicious. In IoT botnet detection, the model uses features such as packet counts, flow duration, packet size, connection frequency, protocol information, and incoming or outgoing traffic statistics. Each decision tree learns different classification rules from training samples, and the final prediction is determined by combining the outputs of the individual trees. Random Forest can handle complex relationships among network features and is generally less prone to overfitting than a single decision tree. It can also provide feature-importance estimates to help identify traffic characteristics associated with botnet activities. However, its performance depends on the quality of the training data, feature selection, and class distribution. For practical deployment, the model should be evaluated using precision, recall, F1-score, false-positive rate, and inference time to determine its suitability for resource-constrained IoT gateways.

Support Vector Machine (SVM): Support Vector Machine is a supervised learning technique that identifies an optimal decision boundary for separating different classes of network traffic. In botnet detection, SVM can classify traffic records using extracted network-flow features, including packet size, flow duration, transmission rate, connection behaviour, and protocol-related attributes. The algorithm identifies support vectors, which are the training samples closest to the decision boundary and which influence the classification model. Linear SVM can be used when the classes are approximately separable by a linear boundary, while kernel-based SVM can model more complex nonlinear relationships. This makes SVM a potential solution for distinguishing legitimate IoT communication from malicious botnet traffic. However, its performance may be affected by noisy data, class imbalance, and parameter selection. Training and prediction can also become computationally expensive for large datasets, particularly with nonlinear kernels. Therefore, SVM should be evaluated using representative IoT traffic and compared with other classifiers under consistent experimental conditions.

Extreme Gradient Boosting (XGBoost): Extreme Gradient Boosting (XGBoost) is an ensemble learning algorithm that builds decision trees sequentially, with each new tree attempting to correct errors made by the existing ensemble. For IoT botnet detection, XGBoost can analyse structured network-flow data and learn relationships among features associated with benign and malicious activities. Relevant features may include packet transmission rates, connection duration, protocol distribution, source and destination statistics, and traffic-volume variations. The algorithm incorporates regularization and optimization techniques to control model complexity and improve generalization. It can also support feature-importance analysis, helping researchers identify traffic characteristics that contribute to attack classification. XGBoost is particularly relevant when the dataset contains numerous structured features and multiple attack categories. Nevertheless, its effectiveness depends on appropriate hyperparameter tuning, class-imbalance handling, and leakage-free validation. Its memory consumption and prediction latency should also be measured before deployment on IoT gateways or other resource-constrained platforms.

B. DEEP LEARNING TECHNIQUES

Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a deep learning model that learns representative patterns from structured input data through convolutional filters and successive feature-processing layers. In IoT botnet detection, network-flow features can be arranged into suitable feature vectors or matrices and supplied to the CNN for classification. The convolutional layers learn local feature relationships, while subsequent layers combine these representations to distinguish normal traffic from malicious activities. The final classification layer can identify benign traffic and predefined botnet or attack categories. CNN-based methods can reduce the need for manually specifying every classification rule because relevant feature representations are learned during training. However, network-flow data must be represented appropriately, since arranging features into a matrix does not automatically create meaningful spatial relationships. The model also requires sufficient training data and careful regularization to avoid overfitting. Its performance should be evaluated on separate test data, including unseen traffic scenarios where possible, to determine whether learned patterns generalize beyond the training environment.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network designed to learn temporal dependencies in sequential data. It uses memory cells and gating mechanisms to control the retention and updating of information across a sequence. In IoT botnet detection, network traffic observations can be arranged according to time, device, flow, or communication session. The LSTM model can then learn patterns involving repeated connection attempts, variations in traffic volume, periodic communication, and changes in device behaviour. These temporal characteristics may help distinguish certain botnet activities from normal IoT communication. For example, repeated connection attempts over a short period may indicate scanning behaviour, although such activity must be distinguished from legitimate network operations. LSTM models require carefully constructed sequences and sufficient representative training data. They can also introduce higher computational and memory requirements than conventional machine learning classifiers. Evaluation should therefore consider detection performance, inference

latency, sequence length, and the ability to detect attacks that differ from the training examples.

Autoencoder

An autoencoder is a neural network architecture that learns to represent input data in a compressed form and reconstruct the original input from that representation. It generally consists of an encoder, a latent representation, and a decoder. In IoT botnet detection, an autoencoder can be trained primarily on benign network traffic to learn patterns associated with normal device communication. When new traffic is processed, the difference between the original input and its reconstructed output can be measured using a reconstruction-error function. Traffic with an error exceeding a selected threshold may be classified as anomalous and flagged for further investigation. This approach is particularly relevant when labelled examples of every possible botnet attack are unavailable. However, anomalous traffic is not necessarily malicious, and some attacks may be reconstructed accurately if their characteristics resemble normal traffic. The threshold-selection process therefore has a direct effect on false-positive and false-negative rates. The model should be evaluated on benign traffic, known attacks, and previously unseen attack scenarios to establish its practical usefulness.

Table 2.1: Summary of the Techniques

Technique	Learning type	Main application	Key limitation
Random Forest	Supervised ML	Classification of network-flow features	Ensemble size and model generalization
SVM	Supervised ML	Separation of benign and malicious traffic	Scalability on large datasets
XGBoost	Supervised ML	Structured traffic classification	Hyperparameter sensitivity
CNN	Deep learning	Feature-pattern extraction	Input representation and training requirements
LSTM	Deep learning	Sequential traffic analysis	Computational cost and sequence design
Autoencoder	Unsupervised or semi-supervised DL	Anomaly detection	Threshold selection and false alarms

V. CONCLUSION

AI-based techniques have significant potential for improving botnet attack detection in Internet of Things (IoT) networks by identifying malicious traffic patterns and supporting automated security analysis. This review examined traditional machine learning techniques, including Random Forest, Support Vector Machine, and XGBoost, along with deep learning approaches such as Convolutional Neural Networks, Long Short-Term Memory networks, and Autoencoders. Each technique offers different capabilities for traffic classification, sequential behaviour analysis, and anomaly detection. However, their effectiveness depends on dataset quality, feature selection, class imbalance, and appropriate evaluation procedures. Major challenges include detecting previously unseen attacks, reducing false alarms, adapting to evolving botnet behaviours, protecting data privacy, and meeting the computational requirements of resource-constrained IoT devices. Future research should focus on developing hybrid and lightweight detection models, improving cross-dataset generalization, incorporating explainable AI, and evaluating performance under realistic network conditions. Overall, reliable botnet detection requires a balanced approach that considers detection performance, robustness, computational efficiency, and practical deployment requirements to strengthen the security of heterogeneous IoT networks.

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, pp. 1-6, doi: 10.1109/ICSPIS63676.2024.10812621.
2. N. Dharmaraj, J. Jenifa, S. Subasree, G. D. Dharshini and T. Ragunath, "A Novel Hybrid Deep Learning Approach for Botnet Attacks Detection in a Secure Internet of Things Environment," *2025 6th International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS)*, Bengaluru, India, 2025, pp. 617-621, doi: 10.1109/ICICNIS66685.2025.11315501.
3. R. K. S, S. P and A. R, "Enhanced Convolutional Neural Networks for Adaptive IoT Botnet Attack Detection Using Reinforcement Learning," *2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA)*, Namakkal, India, 2025, pp. 1-5, doi: 10.1109/AIMLA63829.2025.11041316.
4. C. Guo, S. Wang, K. Yu, Y. Zhu and X. Tao, "A Differential Game Method Against DDoS Attacks in IoT Botnets: Holistic and Dynamic Perspectives," in *IEEE Internet of Things Journal*, vol. 12, no. 12, pp. 19414-19427, 15 June15, 2025, doi: 10.1109/JIOT.2025.3541852.
5. R. Sendhil, J. N P and J. S, "Leveraging the power of Machine Learning techniques to Detect Botnet Attacks in IoT-5G Networks," *2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3)*, Bhubaneswar, India, 2025, pp. 1-6, doi: 10.1109/ISAC364032.2025.11156468.
6. A. Bhadoriya, S. K. Tiwari, A. S. Bisen and S. Swami, "Hybrid Deep Learning Framework for IoT Botnet Detection Using Network Flow Features and Autoencoder-Based Analysis," *2025 International Conference on Digital Innovations for Sustainable Solutions (ICDISS)*, Faridabad, India, 2025, pp. 1-6, doi: 10.1109/ICDISS68238.2025.11320590.
7. X. Wang, X. Zhang, S. Wang, J. Xiao, and X. Tao, “Modeling, critical threshold, and lowest-cost patching strategy of malware propagation in heterogeneous IoT networks,” *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 3531–3545, 2023.
8. X. Chen, L. Xiao, W. Feng, N. Ge, and X. Wang, “DDoS defense for IoT: A Stackelberg game model-enabled

collaborative framework," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9659–9674, Jun. 2022.

9. H. Wu, Q. Gao, X. Tao, N. Zhang, D. Chen, and Z. Han, "Differential game approach for attack–defense strategy analysis in Internet of Things networks," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 10340–10353, Jun. 2022.
10. C. Zhang, F. Luo, M. Sun, and G. Ranzi, "Modeling and defending advanced metering infrastructure subjected to Distributed Denial-of-Service attacks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2106–2117, Jul.–Sep. 2021.
11. H. Zhang, J. Tan, X. Liu, S. Huang, H. Hu, and Y. Zhang, "Cybersecurity threat assessment integrating qualitative differential and evolutionary games," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 3, pp. 3425–3437, Sep. 2022.
12. A. Mahboubi, S. Camtepe, and K. Ansari, "Stochastic modeling of IoT botnet spread: A short survey on mobile malware spread modeling," *IEEE Access*, vol. 8, pp. 228818–228830, 2020.