RESEARCH ARTICLE OPEN ACCESS

Botnet Attack Prevention in Internet of Things (IOT) devices Using AI: A Systematic Review

Sanjeev kumar [1] , Prof. Shivank Soni [2]

[1]Research Scholar, CSE Oriental Institute of science &technology, Bhopal
[2]Assistant professor, CSE Oriental Institute of science &technology, Bhopal

ABSTRACT

Botnet attacks may result in unauthorised access, Distributed Denial-of-Service (DDoS) attacks, and data breaches on Internet of Things devices. Because botnets are dynamic and IoT devices have limited resources, traditional security solutions often fail. This review offers an in-depth analysis of deep learning and machine learning techniques for detecting and preventing botnet attacks in Internet of Things environments. It looks at deep learning, supervised, and unsupervised models, highlighting how they may improve network security, automate anomaly detection, and predict new threats. The study highlights AI's potential to enhance IoT security while examining many important issues, such as unbalanced datasets, adversarial threats, computational constraints, and data privacy issues. Advanced methods like federated learning, explainable AI, hybrid deep learning models, and transfer learning are being researched to overcome these problems. The efficacy of the model is further evaluated by reviewing assessment metrics, feature extraction methods, and frequently used datasets. In order to improve IoT security, future research approaches will include adaptive security mechanisms, ethical AI frameworks, and real-time threat intelligence. This paper highlights the revolutionary potential of AI-driven security solutions and promotes more innovation for real-world application by synthesising previous research.

Keywords: *Botnet Attacks, Internet of Things (IoT), Machine Learning (ML), Deep Learning (DL), Cybersecurity, Anomaly Detection, Threat Intelligence, Network Security, AI-driven Security.*

I. INTRODUCTION

Society is undergoing a significant transformation due to the Internet & the corresponding digital revolution. These days, the majority of people's daily activities, including banking, communication, trade, education, entertainment, and information sharing, are done online. A parallel movement of criminal and malicious entities to cyberspace is occurring as a result of the relocation of these economic and social activities on the Internet as well as the rise in their use. Malware trends released by several organisations, such as Symantec[l], show an ever-increasing quantity of malware and assaults recorded year, demonstrating how these criminal elements take advantage of weak systems and people for financial benefit. Since cyberattacks just need a computer & an Internet connection, they are less expensive to execute than physical attacks since, unlike in the real world, the target's geographic location is not a barrier to attack.

Botnets are networks of hacked computers, each of which is referred to as a "bot" and is managed by a botmaster via a Control and Command (C&C) channel. Bots use a variety of infection techniques to infect host computers linked to the network, such as searching for susceptible computers, compromising websites and infecting users of those hacked sites (a process known as drive-by downloads), or, more recently, social media accounts. The botmaster uses the C&C channel to send instructions and binary updates to these infected zombie devices. Individual bots carry out assaults including spam generation, keylogging, phishing, Distributed Denial of Service attacks, creating phoney clicks to advertising websites, and more based on the orders they get. Additionally, without the genuine users' knowledge, bot computers are utilised to host illicit information and operate as proxy servers. Under the concept of Crimeware as a Service (CaaS), botnets are hired out for illicit purposes [2]. Up to 20% of ad clicks and over 85% of spam emails are thought to be caused by botnets [3]. As shown by the Russian government-sponsored botnet-driven denial-of-service attacks on the government websites of [4] & [5], botnets also represent a serious danger to national security. The most recent instance of botnet-based cybercrime was the May 2014 discovery by Intel Crawler of the massive Point-of-Sale (PoS) network known as Nemanja [5]. PoS systems were home to the Nemanja botnet, which gathered credit card information and

other login credentials in order to access other systems. Research on botnet detection has to be done immediately and actively because of the wide variety of assaults that botnets may be used for and the amount of harm they can do.

A botnet, as defined by [7] and [8], is an overlay network made up of several hosts infected with bots that are controlled by an attacker to carry out harmful actions. According to [9], bot masters have the authority to instruct the server to execute a variety of cyberattacks. These threats include DDoS attacks, spam, phishing, click fraud, and data theft. One of the biggest security threats associated with the internet is these cyberattacks. Given the security risks presented by the spread of botnets, both academic and commercial research continues to place a high priority on the detection and identification of these harmful networks, particularly those that are still in the early phases of development. First, there are a number of complex and unique features that distinguish botnet control and command techniques. In particular, 5G, the Internet of Things, intelligent terminals, cloud platforms, & social media sites like Facebook and Twitter have all been gradually compromised by botnets. [10] [11] [12] Numerous technologies, such as peer-to-peer networks, phishing, rapid flux, anonymised networks, bitcoin networks, lightning networks, zero-day vulnerabilities, and more, are used and disseminated by botnets.

II. BACKGROUND AND MOTIVATION

By providing smooth communication between smart devices, from industrial control networks to home automation systems, the Internet of Things has completely transformed a number of sectors. However, this quick growth has also made IoT devices vulnerable to serious cybersecurity risks, of which botnet attacks are now among the most deadly. Cybercriminals may use a botnet, which is a collection of infected devices, to conduct extensive attacks including Distributed Denial-of-Service (DDoS), data breaches, and unauthorised access. High-profile attacks, such as the Mirai botnet, have shown how destructive these threats may be, resulting in compromised personal data, malfunctioning systems, and financial losses. IoT devices have limited computing capabilities, and botnet attacks are constantly changing, thus traditional security methods like firewalls, signature-based intrusion detection systems, and encryption often fail. Detecting zero-day attacks is difficult for rule-based security methods, and keeping threat signatures current is difficult. This makes the need for automated, intelligent, and flexible security systems that can quickly identify and neutralise botnet threats critical.

Deep Learning and Machine Learning have become effective methods for behaviour analysis, anomaly detection, and predictive security in Internet of Things networks in response to these issues. Instead of depending on preset criteria, machine learning models may identify anomalies in network data, learn from past attack patterns, and identify risks that haven't been identified before, unlike conventional approaches. Adaptive defence mechanisms, network traffic analysis, and real-time botnet detection are made possible by supervised, unsupervised, and deep learning approaches.

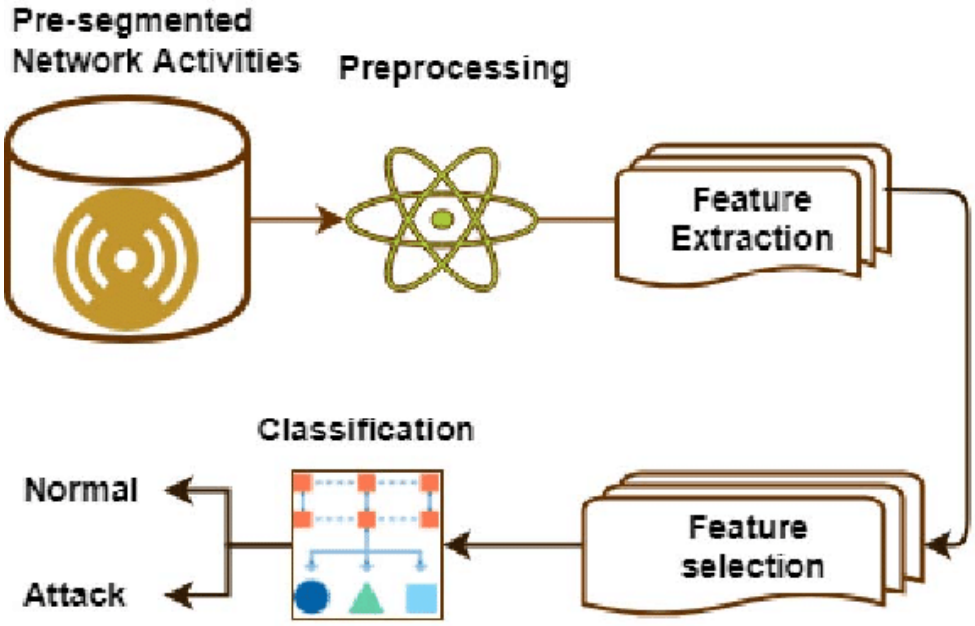

Figure Error! No text of specified style in document.**.1 Flow Diagram for Botnet Detection**

Unbalanced datasets, adversarial AI attacks, high computing costs, and the need for explainability in security choices are some of the obstacles that still exist despite the encouraging developments. Potential remedies are provided by cutting-edge technologies such as explainable AI, hybrid deep learning models, federated learning, and edge-based security frameworks. The purpose of this review is to examine the difficulties associated with ML and DL approaches, assess their efficacy in preventing botnet attacks, and suggest future lines of inquiry. The revolutionary importance of AI-driven cybersecurity in safeguarding IoT networks is highlighted by this study, which bridges the gap between research and practical application.

III. AI-BASED MACHINE LEARNING APPROACHES FOR BOTNET DETECTION

Machine learning methods driven by artificial intelligence are becoming crucial for identifying botnet networks of infected devices being utilised maliciously. AI-driven strategies provide flexibility, automation, and improved accuracy in detecting changing threats in contrast to conventional rule-based techniques. Analysing network traffic patterns, identifying irregularities, and categorising malicious actions are the foundations of AI-based botnet detection. These methods fall under the following categories:

Supervised Learning

Labelling training data is a technique used in supervised learning. The computer "learns" from the labelled patterns to build the classifier, which then uses those patterns to predict labels for future data [13,14]. The training data in unsupervised learning are not labelled. In this method, the classifier is created by the computer "learning" by examining data attributes. The most widely used machine learning algorithms are Decision Tree, BayesNet, J48, Naive Bayes, and Support Vector Machine [13,14]. SL and UL are combined in semi-supervised learning. When using this method, the input training dataset generally has a big number of unlabelled data and a small amount of labelled data with labels. Every strategy has its own application area, advantages, and disadvantages [13,15].

- **Logistic Regression:** A simple but powerful approach for botnet identification, logistic regression is very useful for spotting unwanted network activity. It uses extracted data such as packet size, flow length, and communication patterns to estimate the likelihood that a network flow is botnet-infected. With well-defined feature selection, LR can differentiate between benign & malicious behaviour since botnets often display unique traffic fingerprints. Its linear design, however, restricts its ability to identify intricate botnet systems, particularly in adaptive botnets that exhibit dynamic behaviour. Because of its ease of use and interpretability, LR is often used as a baseline classifier in intrusion detection systems.

- **Random Forest:** Because Random Forest can manage high-dimensional network traffic and recognise intricate assault patterns, it is often used for botnet detection. It is robust to noisy or incomplete data by constructing numerous decision trees using randomly chosen network parameters including connection length, protocol kinds, and packet intervals. Both centralised and peer-to-peer (P2P) botnets may be successfully detected using RF, which offers excellent accuracy while reducing overfitting. However, parsing large-scale network records may be computationally demanding. For security analysts working with intrusion detection systems (IDS), its interpretability and feature relevance rating make it useful.
- **KNN:** K-Nearest Neighbours is a simple yet effective method for detecting botnets that groups network data according to how closely it resembles known botnet behaviours. KNN compares labelled network traffic samples to distance metrics like cosine or Euclidean similarity to identify malicious flows. When trained on well-structured datasets, it performs effectively, especially when identifying botnets that exhibit recurring communication patterns. However, real-time applications may be limited by its sensitivity to feature selection and high computational cost in large-scale network systems. Approximate Nearest Neighbours and KD-Trees are two optimised variants that aid in enhancing performance in high-dimensional network traffic analysis.
- **Decision Tree:** By recursively separating characteristics like packet time, source-destination relationships, and protocol use, decision trees are able to categorise network traffic. They work well for rule-based botnet identification because botnets often display recognisable traffic patterns. Because DTs provide data that are easy to read, cybersecurity researchers may utilise them to better analyse attack patterns. Deep trees, however, have a propensity to overfit, which might result in false positives or decreased effectiveness on novel botnet variations. Nevertheless, decision trees are the basis for more resilient ensemble techniques, such as Random Forest, which increase detection precision while preserving botnet classification efficiency.
- **SVM:** By identifying an ideal decision boundary, Support Vector Machine, a reliable classification model for botnet detection, successfully separates botnet traffic from normal network flows. It is helpful for identifying covert botnets that imitate typical user behaviour since it can employ nonlinear kernels to capture complicated traffic behaviours. SVM has trouble scaling in big network datasets but does well in high-dimensional feature spaces. For the best detection performance, hyperparameter adjustment and kernel selection are essential. SVM has been effectively used in intrusion detection systems (IDS) to identify zero-day botnet attacks, despite its processing demands.
- **MLP:** Because it can recognise complex assault patterns, the Multilayer Perceptron, a kind of artificial neural network, is very good at detecting botnets. MLP distinguishes between legitimate and botnet traffic by examining characteristics such as packet frequency, traffic entropy, and domain name system (DNS) requests. Backpropagation-trained MLP is very effective in identifying adaptive botnets, which alter their behaviour to avoid detection. But it needs a lot of processing power and big datasets. Dropout and regularisation help to reduce the problem of overfitting. MLP increases resistance to changing cyberthreats by providing a solid basis for intrusion detection systems based on deep learning.

Unsupervised Learning

Because unsupervised learning techniques don't need labelled data, they can effectively identify botnet behaviours that haven't been seen before. These methods examine network traffic patterns and spot irregularities that can point to botnet activity.

- **K-Means :** Botnet identification in unlabelled datasets may benefit from the use of the K-Means clustering technique, which classifies network traffic according to feature similarity. While botnet activity creates distinct clusters or outliers, regular traffic is clustered together to identify aberrant behaviours. Botnet activity may be distinguished by characteristics such as protocol use, flow size, and packet timing. K-Means, on the other hand, makes the assumption that cluster numbers are fixed, which may not match changing botnet trends. It requires pre-processing and tweaking and works well with structured datasets but has trouble with dynamic botnet behaviours. K-Means is often employed for anomaly identification in intrusion detection systems (IDS) in spite of these difficulties.
- **DBSCAN:** Because it detects dense traffic clusters and marks sparse or odd network flows as anomalies, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a useful tool for botnet identification. It is appropriate for identifying botnets with different communication patterns since, in contrast to K-Means, it does not need a set number of clusters. Peer-to-peer (P2P) botnets with decentralised traffic

architectures benefit greatly from DBSCAN. It is susceptible to parameter manipulation, however, and might mistakenly identify normal low-density traffic as botnet activity. Notwithstanding its drawbacks, DBSCAN is often used in intrusion detection systems to spot covert and hidden botnet activity.

- **Autoencoders:** Neural networks such as autoencoders pick up on typical network behaviours and use reconstruction mistakes to spot abnormalities. The autoencoder reconstructs network traffic characteristics after compressing them into a latent space during training. A significant reconstruction error indicates possible botnet activity by implying departures from typical patterns. Zero-day botnet assaults may be successfully detected by autoencoders without the need for labelled data. However, they need high-quality normal traffic samples for training in order to function well. Poor generalisation may result from overfitting, and large false-positive rates are frequent. In spite of this, autoencoders are essential to anomaly detection systems that rely on deep learning.

Deep Learning

- **Recurrent Neural Networks (RNNs):** Because recurrent neural networks examine sequential network traffic data to spot suspicious patterns, they are useful for botnet identification. RNNs analyse packet flows, timestamps, and behavioural sequences to identify abnormalities since botnet activity often displays temporal relationships. Nevertheless, vanishing gradient problems plague conventional RNNs, which restricts their capacity to identify long-term relationships. RNNs demand a lot of computer power and vast datasets, even if they increase detection accuracy over conventional techniques. They are useful in intrusion detection systems, especially for identifying changing botnet communication patterns, because of their capacity to simulate dynamic botnet behaviours.
- **Long Short-Term Memory (LSTMs) :** By identifying long-term dependencies in network traffic sequences, Long Short-Term Memory (LSTM) networks—a subset of RNNs—are very good at identifying botnets. By addressing the vanishing gradient issue, LSTMs are able to identify persistent botnet behaviours over time. LSTMs can differentiate between legitimate and botnet traffic by examining packet interarrival delays, connection durations, and traffic quantities. They are quite good at identifying covert botnets that use sporadic or sluggish communication patterns. But LSTMs need a lot of labelled training data and are computationally demanding. In spite of this, they are often used in botnet categorisation in deep learning-based intrusion detection systems.
- **Convolutional Neural Networks (CNNs) :** By considering network data as a structured feature map, Convolutional Neural Networks (CNNs), which are mainly employed in picture recognition, have been modified for botnet detection. CNNs discover patterns suggestive of botnet activity by extracting spatial associations from protocol distributions, flow sequences, and packet headers. They can handle encrypted communication without requiring thorough packet inspection and are especially useful for high-dimensional traffic data. CNNs classify botnets with great accuracy, but they need a lot of computing power and training data. For more thorough botnet detection systems, they are often paired with RNNs or LSTMs to improve time-series analysis.
- **Federated Learning (FL) :** A decentralised method for detecting botnets, federated learning (FL) enables many devices or organisations to work together to build a model without exchanging raw network traffic data. This technique allows for large-scale botnet identification in various situations while improving user security and privacy. FL is especially helpful for IoT-based botnets, because dispersed devices help with training without disclosing private data. Communication cost, model synchronisation, and vulnerability to hostile assaults are obstacles, nevertheless. FL is becoming more popular in cybersecurity applications despite these drawbacks since it offers scalable and private botnet detection in edge and cloud-based networks.

IV. DATASETS

In SDN-based networks, a number of publicly accessible statistics are utilised for botnet and DDoS detection. Researchers may create and evaluate machine learning as well as deep learning models for botnet detection using these datasets, which combine simulated and real-world attack traffic. Some of the most widely used datasets are shown here, arranged according to their areas of interest.

1. Bot-IoT Dataset

- Description: A dataset designed to study botnet behavior in IoT environments, including DDoS floods targeting SDN infrastructures.
- Key Features: Includes large-scale IoT traffic data with centralized and P2P botnets.
- Strengths: Highly relevant for IoT security research, focusing on emerging threats in smart networks.
- Limitations: Due to its high data volume, real-time processing can be challenging.
- Availability: Publicly available at University of New South Wales (UNSW) and IEEE DataPort.

2. CIC-DDoS2019 Dataset

- Description: A large-scale dataset designed for DDoS attack detection, covering modern attack vectors in SDN environments.
- Key Features: Contains multiple DDoS attack types such as UDP flood, SYN flood, HTTP flood, and more, with labeled network traffic.
- Strengths: Extensively used in intrusion detection systems (IDS) and anomaly-based DDoS detection for SDN networks.
- Limitations: Due to its large dataset size, it requires efficient data preprocessing and feature selection for ML/DL models.
- Availability: Publicly available on the UNB CIC website.

3. TON_IoT Dataset

- Description: This dataset integrates IoT, network traffic, and system logs, making it useful for detecting both DDoS and botnet attacks in SDN-based networks.
- Key Features: Includes network flow records, system logs, and telemetry data, allowing multi-dimensional threat detection.
- Strengths: Useful for next-generation IDS solutions, particularly in IoT-SDN security research.
- Limitations: Requires data fusion techniques due to its diverse data sources.
- Availability: Can be accessed on IEEE DataPort and Kaggle.

4. UNSW-NB15 Dataset

- Description: A hybrid dataset containing normal and attack traffic, including DDoS and botnet samples.
- Key Features: Provides 49 network features, such as protocol types, traffic analysis, and attack labels, making it useful for IDS research.
- Strengths: Covers a wide range of cyberattacks beyond botnets, making it beneficial for general intrusion detection in SDN.
- Limitations: The dataset contains synthetic attack data, which may not fully represent real-world botnet behaviors.
- Availability: Publicly available via Australian Centre for Cyber Security (ACCS).

5. CTU-13 Dataset

- Description: One of the earliest datasets focused on real-world botnet detection, collected from Czech Technical University.
- Key Features: Contains 13 botnet scenarios with real-world botnet communication and attack behaviors.
- Strengths: Useful for flow-based anomaly detection in SDN and traditional networks.
- Limitations: Some attack traces are outdated, requiring feature selection and preprocessing for modern IDS applications.

- Availability: Available at Czech Technical University and research platforms like Kaggle.

V. RELATED WORK

A comprehensive analysis of existing research is essential to understand the developments and challenges in a specific field. This paper explores advancements in botnet attack prevention in Internet of Things (IoT) devices using Artificial Intelligence (AI) by systematically reviewing recent studies. As IoT devices become more prevalent, they face increasing security threats, particularly botnet attacks that compromise device integrity and user privacy. AI-driven techniques, including machine learning and deep learning, have shown promising results in detecting and preventing such attacks in real time. This review aims to evaluate current approaches, identify research gaps, and highlight the effectiveness of AI-based security solutions in protecting IoT ecosystems from evolving cyber threats.

David Concejal Muñoz et. al [16] Intrusion detection systems (IDS) have been widely explored for detecting botnet attacks, with cloud-based machine learning and deep learning models playing a key role. However, the rapid growth of the Internet of Things (IoT) demands a more decentralized approach due to high data volume and latency concerns. This study proposes an anomaly-based IDS deployed at the IoT-edge, utilizing software-defined networking (SDN). In this architecture, IoT-edge devices request behavioral insights from the SDN controller, representing device activity as communication graphs rather than traditional network traffic analysis. This method reduces data volume while improving detection accuracy. The proposed approach is validated using botnet attack simulations with the IoT-23 dataset. Experimental results demonstrate high detection accuracy with minimal memory and storage requirements, making it suitable for resource-constrained IoT devices. By integrating edge computing and SDN, this approach enhances efficiency, real-time detection, and scalability in IoT network security.

Khalid Alissa et. al [17] With the increasing number of Internet of Things (IoT) devices connected to networks, security threats and cyberattacks, such as botnets, have become more sophisticated and pose significant risks. These attacks disrupt network operations and services, compromising the reliability and security of IoT ecosystems. To address this challenge, recent studies have explored machine learningand deep learning techniques for detecting and classifying botnet attacks in IoT environments. This study focuses on implementing machine learning methods for binary classification using the publicly available UNSW-NB15 dataset. To overcome the class imbalance issue, the SMOTE OverSampling technique was applied. A complete machine learning pipeline was developed, including exploratory data analysis to gain insights into the dataset, followed by a structured preprocessing approach with six essential steps. Three machine learning models decision tree, XGBoost, and logistic regression—were trained, tested, and evaluated based on multiple performance metrics, including accuracy, F1-score, recall, and precision. Experimental results indicate that the decision tree model achieved the highest performance, with a 94% test accuracy, demonstrating its effectiveness in classifying botnet attacks in IoT networks.

Table 1 Related Work on Malware and Botnet Detection Techniques

Study	Focus Area	Methodology	Key Findings	Limitations
Alsmadi & Alqudah (2021)[18]	Cyber Attacks	Analysis of modern attack trends and malware types	Rise in malware attacks due to increased internet speed and evolving threats	Detection is difficult due to rapid changes in attack architectures
Tran et al. (2018) [19]	Zero-Day Attacks	Research on malware detection techniques	Emphasized challenges in detecting advanced malware types (e.g., ransomware, DDoS)	Malware continuously evolves, making detection harder
Ianelli & Hackworth (2005) [20]	Botnets	Study of botnet threats	Botnets facilitate cybercrime like DDoS, phishing, and fraud	Hard to detect due to their hidden nature
Li et al. (2009) [21]	Botnet Evolution	Review of botnet history (e.g.,	P2P botnets emerged as a more	Detection techniques need to

		Eggdrop, Sinit, Phatbot)	sophisticated attack method	evolve with botnet structures
Taylor (2019) [22]	Behavior-based Detection	Evaluates program code and executed actions	Detects malware based on access to critical files and OS instructions	Ineffective if malware alters its code dynamically
Kugisaki et al. (2007) [23]	Botnet Communication	IRC-based botnet detection	Differentiates client-server from bot-server communication	Fails if bots use non-IRC communication
Zhao et al. (2013) [24]	BotMiner Model	Clusters bot behavior into communication (C-plane) and activity (A-plane)	Detects similar bot behaviors in network traffic	Less effective for highly dynamic botnets
Ji et al. (2015) [25]	BotCatch Model	Combines signature and behavior-based detection	Uses multi-feedback mechanism for malware detection	Requires continuous updating of malware signatures
Strayer et al. (2008) [26]	Botnet Traffic Analysis	Filters and clusters botnet-related traffic based on packet timing and bandwidth	Effectively groups botnet flows based on similarities	May miss botnets with irregular traffic patterns
Barsamian (2009) [27]	Synchronous Bot Behavior	Detects periodic botnet activity	Identifies predictable bot behaviors in network traffic	Not effective for botnets that change behavior dynamically
Torres et al. (2016) [28]	RNN for Botnet Detection	Uses RNN and k-fold validation for anomaly detection	Achieves minimal false alarms	Struggles with traffic imbalance and indistinguishable patterns
Rehak et al. (2009) [29]	Cooperative Adaptive Detection	Multi-stage approach (anomaly, trust model, collector)	Reduces false positive predictions	Performance depends on accurate anomaly definitions
García et al. (2014) [30]	BClus Framework	Uses known bot behaviors to cluster similar threats	Recognizes bot behaviors based on IP clusters	May not detect novel botnet patterns

VI. EVALUATION METRICS

To effectively detect botnet attacks in Internet of Things environments using Artificial Intelligence, various performance metrics are employed to evaluate detection models. These metrics provide insights into a model's correctness, reliability, and practicality, allowing researchers and practitioners to compare approaches and select the most suitable one for their needs. Common metrics include F1-score, recall, precision, and area under the receiver operating characteristic curve (AUC-ROC), each representing a different aspect of the model's performance. Precision measures the proportion of correctly identified botnet attacks among all predicted attacks and is calculated as the number of true positives (TP) divided by the sum of true positives and false positives (FP). Since investigating false positives can be costly, a model with a low false alarm rate is preferred, though being overly cautious may result in undetected threats. Recall, also known as sensitivity or the true positive rate, evaluates how many actual botnet attacks the model correctly identifies and is determined by dividing true positives by the sum of true positives and false negatives (FN). A high recall indicates strong detection capability, which is crucial in security-sensitive applications where undetected attacks can have severe consequences. However, high recall may lead to excessive false positives,

causing unnecessary alerts and disruptions. The F1-score, which is the harmonic mean of precision and recall, helps balance these trade-offs, making it particularly useful for handling imbalanced datasets common in botnet detection.

It is calculated as F1 = 2 × (precision × recall) / (precision + recall), with a high score indicating a good balance between identifying threats and minimizing false positives. Another key metric, AUC-ROC, evaluates a model's ability to differentiate between botnet and normal traffic at different thresholds. The ROC curve compares recall and false positive rate (FPR), where an AUC value of 0.5 indicates no discriminative ability, and 1.0 represents perfect classification. This metric is useful when the ideal classification threshold is uncertain or varies across different applications. Additional metrics such as specificity, accuracy, and the Matthews correlation coefficient (MCC) further help assess model performance. Accuracy measures the proportion of correct predictions but can be misleading in imbalanced datasets where botnet traffic is rare. Specificity quantifies the percentage of correctly identified normal traffic, complementing recall. MCC, a correlation coefficient between actual and predicted classifications, is particularly useful for unbalanced datasets, providing a comprehensive performance measure. Selecting the appropriate evaluation metrics depends on the specific goals and constraints of botnet detection systems. In critical security applications, recall may be prioritized to ensure all botnet threats are detected, even at the cost of false positives. Conversely, in resource-constrained environments, precision may be more important to reduce unnecessary investigations. Since false positives and false negatives impact security differently, no single metric fully captures the real-world consequences of classification errors. By carefully analyzing these trade-offs, researchers and practitioners can select the best combination of metrics to develop and improve AI-driven botnet detection systems, ensuring optimal security in IoT networks.

Table 2 Metrics Used in Classification Problems

Metric	Formula	Interpretation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall performance of model
Precision	$\frac{TP}{TP+FP}$	How accurate the positive predictions are
Recall	$\frac{TP}{TP+FN}$	Coverage of actual positive sample
F1 score	$\frac{2TP}{2TP+FP+FN}$	Hybrid metric useful for unbalanced classes

VII. CHALLENGES AND LIMITATIONS

The efficacy of AI-driven botnet detection in IoT systems is impacted by a number of important issues. Data imbalance is one of the main issues since botnet assaults are far less common than regular network traffic. A significant percentage of false negatives, in which real botnet infections go unnoticed, might arise from biassed algorithms that are unable to identify minority-class threats. The growth of attack methods is another significant obstacle, as fraudsters are always creating more complex botnets that may evade detection systems. AI models that have been trained on historical data may not be able to identify novel attack patterns, necessitating frequent updates and retraining—a process that may be costly and time-consuming. Accurate detection also depends heavily on feature selection and data quality. Large volumes of diverse data are produced by IoT devices, making it challenging to identify the most relevant characteristics for categorisation. Model performance is further deteriorated by poor data quality, such as noise and missing values. Due to the high processing power and memory requirements of sophisticated AI models, especially those based on deep learning, computational cost & scalability are also major challenges. Complex models for real-time detection are difficult to deploy on many IoT devices due to their restricted hardware resources.

The actual implementation of AI-based botnet detection systems is further limited by a number of constraints. Interpretability and explainability are two major drawbacks; many AI models, particularly deep neural networks, operate as "black boxes," making it difficult for security analysts to comprehend how they arrive at their conclusions. This lack of transparency makes it harder to comply with regulations and erodes confidence in AI-driven security solutions. The possibility for adversarial assaults, in which attackers alter input data to trick AI algorithms, resulting in misclassification and possible security breaches, is another significant drawback. As a result, AI-based detection systems become less reliable. Analysing network traffic also raises privacy issues since handling private user data may be necessary for data packet inspection and monitoring. When implementing AI-driven security solutions, ethical and legal issues pertaining to data protection and adherence to laws like GDPR must be addressed. Lastly, because AI-based detection systems must cooperate with conventional cybersecurity measures like firewalls and intrusion detection systems, integration with current security frameworks continues to be a constraint. Maintaining high detection accuracy while minimising false alarms and ensuring smooth interoperability is still a difficult task that calls for constant innovation and research.

VIII. RECENT TRENDS AND FUTURE DIRECTIONS

In recent years, AI-driven botnet detection in IoT contexts has advanced quickly, and a number of new developments are influencing cybersecurity going forward. Combining deep learning (DL) with reinforcement learning (RL) for real-time botnet identification is one of the most prominent developments. While deep learning methods like convolutional neural networks (CNNs) & recurrent neural networks (RNNs) automatically identify intricate patterns from network traffic data, increasing detection accuracy, traditional machine learning models need human feature selection. By allowing models to develop adaptive defence techniques against changing botnet threats, reinforcement learning significantly improves security. The use of federated learning for decentralised security is another significant development. This approach addresses privacy issues while preserving detection efficacy by enabling AI models to be trained across several IoT devices without exchanging raw data.

Since moving detection techniques closer to IoT devices lowers latency and improves real-time threat mitigation, edge computing & AI-driven anomaly detection are also becoming more popular. By processing data locally, edge AI models reduce dependency on cloud computing and enhance scalability for extensive IoT networks. Furthermore, a viable avenue for botnet protection is the integration of blockchain technology. By limiting unwanted access and guaranteeing transparent monitoring of network activity, blockchain's decentralised and impenetrable nature improves security. The use of generative adversarial networks (GANs) for adversarial attack resistance is another expanding trend. In GANs, artificial intelligence (AI) models are taught to recognise and counteract complex evasion strategies that attackers use to get around detection systems.

Future studies will concentrate on creating self-learning AI models that can identify unidentified botnet variations without the need for regular retraining. Explainable AI will also be prioritised as it will increase regulatory compliance and foster confidence by making detection models easier to understand and comprehend. Furthermore, by allowing AI models to analyse enormous volumes of data at previously unheard-of rates, the use of quantum computing might completely transform botnet detection. A multi-layered security strategy that combines AI with conventional cybersecurity techniques will be crucial to improving resistance against botnet assaults as cyber threats continue to change. AI-driven botnet detection solutions will become more effective, flexible, and safe by using these developments, guaranteeing strong defence for IoT networks going forward.

IX. CONCLUSION

In this research, we investigate cutting-edge AI-powered methods for detecting botnets and their revolutionary potential to improve cybersecurity. Real-time botnet activity detection, anomaly detection, and network traffic analysis have all been made possible by machine learning & deep learning approaches. These models are essential for protecting contemporary networks because they continually learn and adapt to changing attack patterns, outperforming conventional rule-based systems. However, a number of obstacles prevent their broad use. AI models may be manipulated by adversarial assaults, which can result in incorrect categorisation and detection evasion. Furthermore, the deployment of sophisticated ML models is constrained by the resource limitations of IoT devices. Concerns about trust and transparency in cybersecurity applications are also raised by the inability to comprehend AI-driven conclusions. Promising answers to these problems are provided by recent developments. By improving model transparency, explainable AI (XAI) enables security analysts to comprehend and have faith in AI-based judgements. Decentralised training is made possible by federated learning, which enhances detection accuracy over dispersed IoT networks while protecting data privacy. Furthermore, transfer learning methods and hybrid deep learning models

improve flexibility and resistance to changing botnet tactics. To guarantee reliable and responsible botnet identification, future research should concentrate on combining cost-sensitive learning, real-time adaptive security measures, and ethical AI frameworks. The effect and expense of cyberattacks will be decreased by proactive threat mitigation made possible by developments in incremental learning. A secure digital environment may be ensured by cybersecurity by tackling these issues and using AI advancements to create more robust, scalable, and efficient botnet defence systems.

REFERENCES

1. 'Internet Security Threat Report', Symantec Corporation, Volume 17, Apr. 2012.
2. A. K. Sood and R. J. Enbody, 'Crimeware-as-a-service— A survey of commoditized crimeware in the underground market'. Int. J. Crit. Infrastruct. Prot., vol. 6, no. 1, pp. 28-38, Mar. 2013.
3. R. Gummadi. H. Balakrishnan, P. Maniatis, and S. Ratnasamy, 'Not-a-Bot: Improving Service Availability in the Face of Botnet Attacks.', in NSDI, 2009, vol. 9, pp. 307-320.
4. M. Lesk, 'The new front line: Estonia under cyberassault', Secur. Priv. IEEE, vol. 5,no. 4, pp. 76-79, 2007.
5. G. Hogben, D. Plohmann, E. Gerhards-Padilla. and F. Leder, 'Botnets: Detection, measurement, disinfection and defence', Eur. Netw. Inf. Secur. Agency ENISA, 2011 .
6. 'Point - of - Sale and Modem Cybercrime Detection of "Nemanja" Botnet'. May 2014.
7. Fang, X Cui & Wang, W 2011, 'Survey of botnets',Journal of Computer Research and Development, vol. 48, no. 8, pp. 1315–1331.
8. Vormayr, G, Zseby, T & Fabini, J 2017, 'Botnet communication patterns', IEEE Communications Survey's; Tutorials, vol. 19, no. 4, pp. 2768–2796
9. Karim, RB, Salleh, M & Shiraz 2014, 'Botnet detection techniques: review, future trends, and issues', Journal of Zhejiang University Science, vol.15, no. 11, pp. 943–983.
10. Casenove, M & Miraglia, A 2014, 'Botnet over tor: the illusion of hiding', in Proceedings of the 6th International Conference on Cyber Conflict, CyCon, Tallinn, Estonia, pp. 273–282.
11. Curran, T & Geist, D 2016, 'Using the bitcoin blockchain as a botnet resilience mechanism', pp. 1-29.
12. Kurt, E, Erdin, M, Cebe, K, Akkaya & Uluagac, AS 2020, 'LNBot: a covert hybrid botnet on bitcoin lightning network for fun and profit', in Computer Security – ESORICS, ESORICS, L. Chen, N. Li, K. Liang, and S. Schneider, Eds, Springer, Berlin, Germany, pp.734-755.
13. Miller, S.; Busby-Earle, C. The role of machine learning in botnet detection. In Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), Barcelona, Spain, 5–7 December 2016; pp. 359–364
14. Hyslip, T.S.; Pittman, J.M. A survey of botnet detection techniques by command and control infrastructure. *J. Digit. Forensics Secur. Law* **2015**, *10*, 2.
15. Wang, X.; Xu, Y.; Chen, C.; Yang, X.; Chen, J.; Ruan, L.; Xu, Y.; Chen, R. Machine Learning Empowered Spectrum Sharing in Intelligent Unmanned Swarm Communication Systems: Challenges, Requirements and Solutions. *IEEE Access* **2020**, *8*, 89839–89849.
16. Muñoz, D. C., & Valiente, A. D. C. (2023). A novel botnet attack detection for IoT networks based on communication graphs. Cybersecurity, 6(1), 33.
17. Alissa, K., Alyas, T., Zafar, K., Abbas, Q., Tabassum, N., & Sakib, S. (2022). Botnet attack detection in iot using machine learning. Computational Intelligence and Neuroscience, 2022(1), 4515642.
18. Alsmadi, T., & Alqudah, N. (2021). A Survey on malware detection techniques. , 371–376. doi: 10.1109/ICIT52682.2021.9491765
19. Tran, D., Mac, H., Tong, V., Tran, H. A., Nguyen, L. G., Fernandez Maimo, L., . . . Caianiello, P. (2018). Overview of botnet detection based on machine learning. ,(3), 29–35
20. Ianelli, N., & Hackworth, A. (2005). Botnets as a vehicle for online crime. ,(1), 28.
21. Li, C., Jiang, W., & Zou, X. (2009). Botnet: Survey and case study. , 1184–1187. doi: 10.1109/ICICIC.2009.127
22. Taylor, T. (2019, nov). techgenix.com.

23. Kugisaki, Y., Kasahara, Y., Hori, Y., & Sakurai, K. (2007). Bot Detection Based on Traffic Analysis. In (pp. 303–306). doi: 10.1109/IPC.2007.91

24. Zhao, D., Traore, I., Sayed, B., Lu, W., Saad, S., Ghorbani, A., & Garant, D. (2013). Botnet detection based on traffic behavior analysis and flow intervals. ,(PARTA), 2–16.

25. Ji, Y., Li, Q., He, Y., & Guo, D. (2015). Botcatch: leveraging signature and behavior for bot detection. ,(6), 952–969.

26. Strayer, W. T., Lapsely, D., Walsh, R., & Livadas, C. (2008). Botnet detection based on network behavior. ,, 1–24. doi: 10.1007/978-0-387-68768-1_1

27. Barsamian, A. V. (2009). (Unpublished doctoral dissertation). Citeseer.

28. Torres, P., Catania, C., Garcia, S., & Garino, C. G. (2016). An analysis of Recurrent Neural Networks for Botnet detection behavior. . doi: 10.1109/ARGENCON.2016.7585247

29. Rehak, M., Celeda, P., Pechoucek, M., & Novotny, J. (2009). Camnep: multistage collective network behavior analysis system with hardware accelerated netflow probes.

30. García, S., Grill, M., Stiborek, J., & Zunino, A. (2014). An empirical comparison of botnet detection methods. ,, 100–123. doi: 10.1016/j.cose.2014.05.011