

Research Methodology in Computer Science

Course Code: AI800

Final Research Project

Toward Early and Lightweight Machine Learning-Based Detection of IoT Botnet Attacks

Student Name: Ibrahim Mohammed Alsidrani

Student ID: 262000987

Lecturer: Dr. Mohammed Anbar

Hand-out Date: 18 July 2026

Hand-in Date: 22 August 2026

Assignment Weight: 50%

Abstract

The rapid expansion of Internet of Things (IoT) environments has increased the attack surface for cybercriminals. Many IoT devices have limited processing power, weak configurations, and irregular security updates, making them attractive targets for botnet attacks. Previous research has shown that machine learning and deep learning can detect IoT botnet traffic with high accuracy; however, laboratory accuracy alone does not prove practical deployment readiness. Many existing approaches are evaluated on known datasets, provide limited evidence of generalization to unseen botnet families, or require computational resources that may not be suitable for IoT gateways and edge devices. This research proposal investigates how machine learning can support early and lightweight detection of IoT botnet attacks while maintaining reliable performance under changing attack behavior. The proposal reviews ten related studies, including deep learning, class-imbalance-aware learning, ensemble learning, collaborative early detection, transfer learning, and cross-device detection. It then proposes a quantitative experimental methodology using publicly available IoT botnet datasets, preprocessing, feature selection, class imbalance handling, and comparative model evaluation. The expected outcome is a conceptual layered detection approach that combines fast gateway-level screening with deeper analysis for suspicious cases. The research aims to contribute practical evaluation criteria for IoT botnet detection, including accuracy, recall, F1 score, AUC, false positive rate, inference time, feature count, and robustness to unseen attacks.

Table of Contents

Abstract	i
Chapter 1: Introduction	1
1.1 Introduction.....	1
1.2 Background	1
1.3 Problem Statement	2
1.4 Research Aim.....	3
1.5 Research Objectives	3
1.6 Research Scope	3
1.7 Chapter Summary	4
Chapter 2: Literature Review	5
2.1 Introduction.....	5
2.2 Deep Learning-Based IoT Botnet Detection	5
2.3 Handling Class Imbalance in IoT Botnet Datasets	6
2.4 Ensemble Learning for IoT Botnet Detection	6
2.5 Early Detection of IoT Botnets	7
2.6 Transfer Learning and Federated Learning for IoT Cyberattack Detection	8
2.7 Lightweight and Cross-Device Detection.....	8
2.8 Comparative Discussion.....	9
2.9 Research Gap	9
2.9.1 Research Gap Mapping.....	10
2.10 Summary of Reviewed Studies	10
2.11 Chapter Summary	11
Chapter 3: Research Methodology	12
3.1 Introduction.....	12
3.2 Research Design	12
3.2.1 Alignment between Research Objectives and Methodology	12
3.3 Dataset Selection.....	13
3.4 Data Collection Method.....	13

3.4.1 Materials and Method.....	13
3.5 Data Preprocessing	14
3.6 Feature Selection and Feature Engineering	14
3.7 Proposed Layered Detection Approach	15
3.8 Model Development	15
3.9 Evaluation Metrics.....	15
3.10 Data Analysis Plan.....	16
3.11 Ethical Considerations	16
3.12 Expected Outcomes	16
3.13 Gantt Chart.....	17
3.14 Research Milestones	17
3.15 Chapter Summary	17
References	18

Chapter 1: Introduction

1.1 Introduction

The Internet of Things (IoT) has become an important part of modern digital environments. IoT devices are now used in smart homes, universities, healthcare systems, industrial environments, public services, and many other sectors. These devices support automation, remote monitoring, data collection, and real-time interaction between users, systems, and physical objects. However, the rapid expansion of IoT has also increased the number of security risks because many IoT devices are resource-constrained and are often deployed with limited security protection.

Unlike traditional computers and servers, many IoT devices have limited processing power, memory, storage, and energy capacity. They may also use weak default configurations, outdated firmware, and irregular security updates. These limitations make IoT devices attractive targets for attackers. Once compromised, IoT devices can become part of a botnet, where a botmaster remotely controls many infected devices and uses them to perform malicious activities such as distributed denial-of-service attacks, network scanning, credential theft, spam, phishing, and data exfiltration.

Machine learning and deep learning techniques have been widely studied as possible solutions for detecting IoT botnet attacks. These techniques can learn patterns from network traffic or device behavior and classify activities as benign or malicious. Previous studies have shown that deep neural networks, convolutional neural networks, recurrent neural networks, ensemble learning, and collaborative detection models can achieve high detection performance in experimental settings (Jithu et al., 2021; Idrissi et al., 2021; Popoola et al., 2021; Abu Al-Haija & Al-Dalaïen, 2022; Nguyen et al., 2022). However, high accuracy in controlled datasets does not always mean that a model is ready for real-world IoT deployment.

This research focuses on the problem of developing an IoT botnet detection approach that is not only accurate, but also early, lightweight, and reliable under changing attack conditions. The study builds on previous literature by examining the gap between experimental detection performance and practical deployment readiness.

1.2 Background

IoT botnets represent a serious cybersecurity threat because they exploit the weaknesses of connected devices. A botnet is a group of compromised devices that are controlled by an attacker through command-and-control mechanisms. In IoT environments, botnets are especially dangerous because many devices operate continuously, are connected to the internet, and may not be regularly monitored by users or administrators.

Traditional signature-based security tools are useful for detecting known attacks, but they are less effective against new, modified, or unknown botnet behavior. For this reason, many researchers have explored machine learning-based detection approaches. Machine learning can identify hidden patterns in network traffic and classify suspicious behavior even when the exact attack signature is not available.

The literature reviewed in the previous assignments shows several important directions in IoT botnet detection. Some studies focus on deep learning-based intrusion detection using IoT network traffic. Others address class imbalance, which is a common issue in cybersecurity datasets because some attack types may appear far less frequently than others. Some studies use ensemble learning to improve prediction stability and reduce inference time. Another important direction is early detection, where the goal is to identify botnet activity before the attack fully develops or causes serious damage.

Despite these developments, several gaps remain. Many models are tested using known datasets and random train-test splits, which may not prove that they can detect unseen botnet families. Some models achieve very high accuracy but require more computational resources than IoT gateways or edge devices can provide. Other approaches support early detection but require richer behavioral data collection and additional processing time. Therefore, there is a need for a practical detection direction that balances accuracy, early response, lightweight deployment, and robustness.

1.3 Problem Statement

The problem domain of this research is IoT cybersecurity, specifically the detection of botnet attacks targeting Internet of Things environments. IoT devices are increasingly used in smart homes, universities, healthcare environments, industrial systems, and public services. These devices provide important benefits by enabling automation, monitoring, and remote access. However, many IoT devices have limited computing power, weak default configurations, and irregular security updates. These weaknesses make them attractive targets for botnet attacks, where compromised devices can be remotely controlled and used to perform malicious activities such as distributed denial-of-service attacks, network scanning, credential theft, and data exfiltration.

Existing solutions show that machine learning and deep learning can detect IoT botnet traffic with high performance in controlled experiments. For example, DNN-based intrusion detection, CNN-based botnet detection, SMOTE-DRNN, ensemble learning, and collaborative early-detection approaches have all been proposed for IoT botnet detection (Jithu et al., 2021; Idrissi et al., 2021; Popoola et al., 2021; Rezaei, 2021; Nguyen et al., 2022; Abu Al-Haija & Al-Dalaien, 2022). These studies demonstrate that data-driven detection can be effective when models are trained and tested on selected datasets.

Despite these promising results, high detection accuracy alone does not fully solve the practical security problem. Many existing approaches are evaluated using known datasets and random train-test splits, which makes it difficult to determine whether the same models can detect unseen botnet families or perform reliably under changing network conditions. In addition, some deep learning models may require more computational resources than IoT gateways or edge devices can provide. Other approaches may detect attacks only after enough malicious traffic has already appeared, reducing the opportunity for early response (Popoola et al., 2021; Nguyen et al., 2022).

The significance of this problem is that IoT environments often operate in resource-constrained and security-sensitive settings. If detection is late, heavy, or unreliable against new botnet behavior, compromised IoT devices may already be used to disrupt services, steal data, or participate in wider cyberattacks. Therefore, there is a clear need for research that moves beyond laboratory accuracy and

investigates how IoT botnet detection can be made early, lightweight, and more reliable in realistic deployment conditions.

Therefore, the research problem addressed in this study is the lack of a practical IoT botnet detection approach that can combine three requirements at the same time: early detection, lightweight operation, and reliable performance against unseen or changing botnet behavior. This research is needed because IoT security solutions should not only perform well on existing datasets but should also support timely response and practical deployment on resource-constrained devices.

1.4 Research Aim

The primary aim of this research is to investigate how machine learning can be used to support early and lightweight detection of IoT botnet attacks while maintaining reliable performance against unseen or changing attack behavior.

1.5 Research Objectives

1. To identify the main limitations of current machine learning-based IoT botnet detection methods in relation to early detection, computational cost, and generalization.
2. To examine the suitability of lightweight models and deep learning models for deployment in IoT gateways or resource-constrained environments.
3. To analyze the effect of class imbalance and unseen botnet behavior on the reliability of reported detection results.
4. To propose a conceptual layered detection approach that combines fast gateway-level screening with deeper analysis for suspicious or uncertain cases.
5. To define practical evaluation criteria for IoT botnet detection models, including accuracy, inference time, resource consumption, early detection capability, and robustness to unseen attacks.

1.6 Research Scope

This research focuses on machine learning-based detection of IoT botnet attacks. The scope includes detection approaches that rely on network traffic features, behavioral features, or a combination of both. The study pays particular attention to four main aspects: early detection, lightweight operation, class imbalance, and generalization to unseen botnet behavior.

The research does not aim to develop a complete commercial intrusion detection system. Instead, it focuses on proposing and evaluating a research direction that can guide the development of a practical IoT botnet detection framework. The proposed direction is mainly suitable for IoT gateways, edge devices, or monitoring systems that need to identify suspicious botnet activity before major damage occurs.

The study also focuses on the relationship between the reviewed literature and the proposed research objectives. Therefore, the scope includes comparing previous detection methods, identifying research gaps, and designing a methodology that can evaluate whether a layered machine learning approach can improve the balance between accuracy, early detection, resource efficiency, and robustness.

1.7 Chapter Summary

This chapter introduced the research topic and explained the background of IoT botnet detection. It discussed the practical limitations of existing machine learning-based approaches and identified the main research problem: the lack of a practical detection approach that combines early detection, lightweight operation, and reliable performance under unseen or changing attack conditions. The chapter also presented the research aim, objectives, and scope. The next chapter reviews the related literature and extends the previous literature review by including additional studies relevant to IoT botnet detection, lightweight learning, early detection, and transfer learning.

Chapter 2: Literature Review

2.1 Introduction

This chapter reviews previous studies related to machine learning-based IoT botnet detection. The literature review builds on the six studies reviewed in Assignment 1 and extends them by adding four additional studies related to transfer learning, federated learning, ensemble classification, malware detection, and lightweight cross-device detection. The purpose of this chapter is not only to summarize previous work, but also to identify the research gap that supports the current proposal.

The reviewed literature shows that machine learning and deep learning have achieved strong results in IoT botnet detection. However, the studies also show that high classification accuracy is not enough to prove deployment readiness. Practical IoT botnet detection systems must also address early detection, computational cost, class imbalance, cross-device generalization, privacy, and robustness against unseen attack behavior. Therefore, this literature review is organized around the main technical directions that relate to the current research objectives.

2.2 Deep Learning-Based IoT Botnet Detection

Jithu et al. (2021) proposed an intrusion detection system for IoT botnet attacks using a deep neural network. The study used the Bot-IoT dataset, which contains both normal and attack traffic generated in a realistic network environment. The main contribution of this work is that it showed how a DNN model can learn traffic behavior and detect IoT botnet attacks, especially DDoS-related activity. The model achieved strong performance, with an accuracy of about 94% and an AUC of 0.98. This indicates that deep learning can be effective in distinguishing malicious IoT traffic from normal traffic.

The strength of Jithu et al.'s work is its direct relevance to IoT botnet detection and its use of an IoT-specific dataset. However, its main limitation is that the model was mainly evaluated on known Bot-IoT patterns. The study did not fully demonstrate whether the model could generalize to unseen botnet families, different IoT environments, or live deployment conditions. In addition, although the study considered the model relatively light, it did not provide a complete resource evaluation involving memory, energy use, gateway throughput, or deployment on constrained hardware.

Idrissi et al. (2021) proposed BotIDS, a convolutional neural network-based intrusion detection system for IoT botnet attacks. The study used the Bot-IoT dataset and converted traffic features into a form suitable for CNN-based classification. BotIDS achieved very high validation accuracy and fast prediction time, making it a strong benchmark for high-speed deep learning-based multiclass detection. The study also compared CNN with RNN, LSTM, and GRU models, which strengthened its methodological contribution.

However, the BotIDS approach also has limitations. Transforming tabular traffic records into image-like structures may improve CNN compatibility, but it does not necessarily reflect a natural spatial relationship among traffic features. In addition, the model was tested using dataset splits from the same source, which may not be enough to prove robustness under cross-dataset or unseen-attack conditions.

Therefore, while BotIDS supports the use of deep learning for fast classification, further validation is still needed for practical deployment.

2.3 Handling Class Imbalance in IoT Botnet Datasets

Popoola et al. (2021) addressed one of the most important data-related problems in IoT botnet detection: class imbalance. In cybersecurity datasets, some attack classes may have many samples, while rare attack types may have very few. This can make overall accuracy misleading because a model may classify common classes correctly while failing to detect minority attacks.

To solve this issue, Popoola et al. proposed SMOTE-DRNN, which combines Synthetic Minority Oversampling Technique with a deep recurrent neural network. SMOTE was used to generate additional samples for minority classes, while the DRNN learned hierarchical representations from the balanced traffic data. The SMOTE-DRNN model achieved very high performance, including 99.50% precision, 99.75% recall, 99.62% F1 score, 99.87% AUC, 99.74% geometric mean, and 99.62% Matthews correlation coefficient.

The main strength of this study is that it does not rely only on accuracy. Instead, it uses several metrics that are more suitable for imbalanced classification. This makes the evaluation more reliable than studies that only report high overall accuracy. However, the approach may increase training complexity because oversampling expands the dataset and recurrent neural networks can be computationally expensive. In addition, synthetic samples may improve performance on the training distribution but do not necessarily guarantee detection of genuinely unseen botnet behavior.

2.4 Ensemble Learning for IoT Botnet Detection

Abu Al-Haija and Al-Dalaien (2022) proposed ELBA-IoT, an ensemble learning model for botnet attack detection in IoT networks. The model used behavioral traffic features and evaluated three decision-tree ensemble techniques: AdaBoosted trees, RUSBoosted trees, and bagged trees. The study used the N-BaIoT-2021 dataset, which includes normal traffic and botnet traffic from infected IoT devices. The model achieved 99.6% detection accuracy and a low inference overhead of about 40 microseconds.

The main contribution of ELBA-IoT is that it directly addresses lightweight detection. Unlike some deep learning approaches, decision-tree ensembles can provide fast inference and may be more suitable for gateway-level detection. This makes ELBA-IoT especially relevant to the current research, which focuses on lightweight and early IoT botnet detection. However, the model was evaluated offline and was not deployed on a physical IoT gateway. Its performance against unseen botnet families and changing network conditions also requires further investigation.

Rezaei (2021) also studied ensemble learning for IoT botnet detection. The study evaluated supervised, unsupervised, and regression-based algorithms before selecting the best-performing models for a sequential ensemble. The final model combined an artificial neural network and a decision tree while reducing the number of features to 20. The study reported 100% accuracy and no false-positive or false-negative errors in the experimental results.

The strength of Rezaei's study is that it compares many machine learning methods and highlights feature reduction as an important factor in botnet detection. Reducing the number of features can support faster detection and lower processing cost. However, the reported perfect accuracy requires cautious interpretation because the model was not validated on an independent external dataset or live IoT traffic. Therefore, its generalization to different networks, devices, and botnet families remains uncertain.

Srinivasan and Deepalakshmi (2023) proposed an Ensemble Classifier Algorithm with Stacking Process for botnet detection. The study focused on selecting optimal features and feeding them into machine learning classifiers to estimate botnet detection performance. The proposed method achieved 94.08% accuracy, 86.5% sensitivity, 85.68% specificity, and 78.24% F-measure. This study is useful because it shows another ensemble-based direction and highlights the role of feature extraction and stacking in improving botnet detection.

However, compared with studies such as ELBA-IoT, the reported F-measure is lower, which suggests that ensemble design alone does not guarantee strong performance. The study also appears broader in cyber-world botnet detection rather than being limited only to IoT gateway deployment. For the current research, this paper is useful mainly because it supports the importance of ensemble classification and feature selection, while also showing that performance depends heavily on dataset quality, model design, and evaluation criteria.

2.5 Early Detection of IoT Botnets

Nguyen et al. (2022) proposed a collaborative approach for early detection of IoT botnets. Unlike studies that focus mainly on classifying traffic after malicious behavior has already appeared, this study aimed to detect botnet activity earlier in the malware lifecycle. The framework collected multiple types of evidence, including network behavior, device performance information, and system calls. It then used different classifiers for each data source and combined their outputs using a collaborative voting approach.

The proposed model achieved 99.37% accuracy on a dataset containing 5,023 IoT botnet malware samples and 3,888 benign samples. This study is especially relevant to the current proposal because it directly addresses the timing of detection. Early detection is important because identifying botnet behavior before the attack is fully executed can reduce damage and support faster response.

The main limitation of Nguyen et al.'s approach is deployment complexity. Collecting heterogeneous behavioral data and running sandbox-based analysis can introduce processing overhead. The framework is therefore more suitable for gateways, edge servers, or controlled analysis environments than for direct deployment on low-power IoT endpoints. This limitation supports the need for a layered detection approach where a lightweight model performs fast screening and a deeper behavioral model analyzes suspicious cases.

2.6 Transfer Learning and Federated Learning for IoT Cyberattack Detection

Khoa et al. (2022) proposed a deep transfer learning-based collaborative learning model for cyberattack detection systems in IoT networks. The study focused on the limitations of federated learning when labeled data are unavailable or when different networks have dissimilar feature sets. The proposed framework used transfer learning to allow a target network with unlabeled data to learn from a source network with abundant labeled data. This is significant because practical IoT environments often lack labeled attack data, and different IoT networks may not share identical features.

The main contribution of this study is that it addresses data availability, privacy, and feature mismatch. Instead of assuming that all participating networks use the same feature space, the proposed approach exchanges learned knowledge among different deep learning models. This improves flexibility and scalability for IoT cyberattack detection. The study reported that the proposed framework improved performance by more than 40% compared with state-of-the-art deep learning approaches.

This study is important for the current proposal because it relates to generalization and deployment across different IoT environments. However, it focuses broadly on cyberattack detection systems in IoT networks rather than only on IoT botnet detection. Therefore, its main value is theoretical and methodological: it suggests that transfer learning and collaborative learning can help overcome the lack of labeled data and improve adaptability across heterogeneous networks.

Kadhim et al. (2022) proposed IOT-MDEDTL, an IoT malware detection approach based on ensemble deep transfer learning. The model used two feature extractors, ResNet101 and VGG16, after converting binary data into grayscale images. The extracted feature maps were fused and classified to distinguish malicious and non-malicious samples. The proposed approach achieved 96.31% accuracy, 95.31% precision, and 94.80% recall.

This study supports the idea that transfer learning can be useful for IoT malware detection when the model needs to learn from limited or transformed data. It is also relevant because malware detection and botnet detection are closely connected in IoT security. However, this method relies on binary-to-image transformation and deep feature extraction, which may increase computational cost. For this reason, it may be less suitable for lightweight gateway deployment unless optimized or used only in a second-stage analysis layer.

2.7 Lightweight and Cross-Device Detection

Catillo, Pecchia, and Villano (2023) proposed a lightweight and cross-device IoT botnet detection approach based on a single deep autoencoder model. Many existing intrusion detection approaches require separate models for different devices or attack types, which is not suitable for large and dynamic IoT networks. Their study proposed an all-in-one deep autoencoder trained on normal traffic from different IoT devices. The method aimed to support cross-device detection and improve transferability.

This study is highly relevant to the current proposal because it directly addresses scalability, lightweight design, and cross-device adaptability. The reported results showed very high precision, recall, and F1

score across devices, with low false-positive rates. The main strength is that the approach avoids building separate models for every device, which can reduce management complexity in large IoT environments.

However, even a lightweight autoencoder-based method still needs careful evaluation under real gateway constraints, changing network traffic, and unseen botnet families. The study strengthens the argument that future IoT botnet detection systems should not only be accurate but should also generalize across devices and reduce operational complexity.

2.8 Comparative Discussion

The reviewed literature shows several important patterns. First, deep learning models such as DNN, CNN, and DRNN can achieve high detection performance, especially when trained and tested on structured datasets. However, these models often require careful preprocessing, sufficient labeled data, and further testing under realistic deployment conditions.

Second, class imbalance is a major challenge. Popoola et al. showed that high accuracy can hide poor minority-class detection. This means that evaluation should include recall, F1 score, AUC, geometric mean, MCC, and false-positive rate rather than accuracy alone.

Third, ensemble learning is a strong direction for lightweight and stable detection. ELBA-IoT and Rezaei's sequential ensemble show that ensemble methods can achieve high accuracy and reduce complexity. However, their results still need stronger cross-dataset and live deployment validation.

Fourth, early detection remains underdeveloped compared with traffic classification. Nguyen et al. showed that early detection is possible using partial behavior and multiple data sources, but this introduces data collection and processing overhead. Therefore, early detection should be combined with lightweight screening rather than replacing it.

Fifth, transfer learning, federated learning, and cross-device detection are important for deployment readiness. Khoa et al. and Catillo et al. show that practical IoT systems require models that can adapt across networks and devices. Kadhim et al. also show that transfer learning can improve IoT malware detection, although computational cost should be considered.

Overall, the literature supports the need for a layered detection framework. A lightweight first-stage model can operate at the gateway level to screen traffic quickly. Suspicious or uncertain cases can then be sent to a deeper second-stage model that uses richer behavioral or transfer-learning-based analysis. This direction combines the strengths of previous studies while addressing their limitations.

2.9 Research Gap

The main research gap identified from the literature is that existing studies usually address only part of the practical IoT botnet detection problem. Some studies achieve high accuracy but do not prove generalization to unseen attacks. Some handle class imbalance but increase computational cost. Some provide fast inference but are not tested on physical gateways. Some support early detection but require richer and slower data collection. Other studies explore transfer learning and cross-device detection but may not focus specifically on botnet attacks or resource-constrained deployment.

Therefore, there is still a need for an IoT botnet detection approach that combines early detection, lightweight operation suitable for IoT gateways, robust performance against unseen or changing botnet behavior, class-imbalance-aware evaluation, cross-device adaptability, and practical deployment metrics such as inference time and resource consumption.

This gap directly supports the current research aim and objectives. The proposed research will therefore focus on a conceptual layered detection approach that combines fast gateway-level screening with deeper analysis for suspicious cases. This approach is expected to improve the balance between accuracy, early response, lightweight deployment, and generalization.

2.9.1 Research Gap Mapping

The following table maps the main gaps identified from the reviewed literature to the proposed research direction. This mapping shows how the current research responds to limitations found in previous studies.

Research Gap	Evidence from Literature	How This Research Addresses the Gap
Limited validation against unseen attacks	Several studies report high accuracy using known datasets or random train-test splits, which may not demonstrate performance against new botnet families.	The proposed research includes generalization as a key evaluation requirement and considers cross-dataset or unseen-attack validation where data availability allows.
High computational cost of deep models	Deep learning approaches such as DRNN and CNN can achieve strong performance but may require more computational resources than IoT gateways can provide.	The proposed approach includes lightweight models and feature selection to reduce computational cost.
Class imbalance in cybersecurity datasets	SMOTE-DRNN shows that imbalanced data can make accuracy misleading and may cause minority attacks to be missed.	The research will evaluate class imbalance using SMOTE, class weighting, and class-sensitive metrics such as recall, F1 score, AUC, and MCC.
Limited early detection capability	Some approaches classify traffic after malicious behavior has already appeared, while early detection requires earlier behavioral evidence.	The proposed layered approach supports fast first-stage screening and deeper analysis for suspicious cases to improve early response.
Weak deployment readiness	Several studies report high experimental accuracy but do not test physical gateway deployment, memory use, energy use, or throughput.	The research will include practical metrics such as inference time, feature count, model complexity, and resource suitability.
Limited cross-device adaptability	IoT environments contain heterogeneous devices, and separate models for each device may not scale well.	The research considers generalization and cross-device adaptability as part of the evaluation direction.

2.10 Summary of Reviewed Studies

No.	Study	Main Focus	Method	Key Strength	Main Limitation
1	Jithu et al. (2021)	IoT botnet IDS	DNN	Learns complex IoT traffic patterns	Limited evidence for unseen botnet generalization
2	Popoola et al. (2021)	Class imbalance	SMOTE-DRNN	Strong minority-class detection	Higher training complexity and resource cost
3	Idrissi et al. (2021)	Deep botnet detection	CNN-based BotIDS	High accuracy and fast prediction	Needs stronger cross-dataset validation
4	Nguyen et al. (2022)	Early detection	Collaborative behavioral analysis	Supports earlier response	Requires richer data collection and processing
5	Abu Al-Haija & Al-Dalaien (2022)	Lightweight detection	Ensemble learning	High accuracy and low inference overhead	Offline evaluation; limited live gateway testing
6	Rezaei (2021)	Feature-efficient detection	Sequential ensemble	Uses selected features and reports high accuracy	Dataset-specific evaluation; no external validation
7	Khoa et al. (2022)	Transfer/federated learning	Deep transfer collaborative learning	Addresses labeled-data shortage and feature mismatch	Broader cyberattack focus, not only botnets
8	Srinivasan & Deepalakshmi (2023)	Ensemble botnet classification	Stacking ensemble	Highlights feature selection and ensemble learning	Lower F-measure and broader botnet context
9	Kadhim et al. (2022)	IoT malware detection	Ensemble deep transfer learning	Uses ResNet101 and VGG16 feature fusion	Potentially high computational cost
10	Catillo et al. (2023)	Lightweight cross-device detection	Deep autoencoder	Supports single cross-device IDS model	Needs further real deployment validation

2.11 Chapter Summary

This chapter reviewed ten studies related to IoT botnet detection, malware detection, ensemble learning, early detection, transfer learning, and lightweight cross-device detection. The literature shows that machine learning and deep learning can achieve strong results, but practical deployment requires more than high accuracy. The main gap is the absence of a unified approach that combines early detection, lightweight operation, class-imbalance-aware evaluation, and generalization to unseen or changing botnet behavior. The next chapter presents the research methodology that will be used to address this gap and achieve the research objectives.

Chapter 3: Research Methodology

3.1 Introduction

This chapter presents the research methodology proposed for achieving the research aim and objectives. The purpose of this research is to investigate how machine learning can support early and lightweight detection of IoT botnet attacks while maintaining reliable performance against unseen or changing attack behavior. To achieve this purpose, the study will follow an experimental and comparative research methodology based on existing IoT botnet datasets, machine learning model development, and performance evaluation.

The methodology is designed to address the research gaps identified in the literature review. These gaps include limited generalization to unseen botnet behavior, insufficient evaluation of lightweight deployment readiness, class imbalance in IoT botnet datasets, and limited integration between early detection and low-cost gateway-level detection. Therefore, the proposed methodology focuses not only on detection accuracy, but also on inference time, class-sensitive performance, resource efficiency, and robustness.

3.2 Research Design

This research will use a quantitative experimental research design. The study will compare selected machine learning and deep learning approaches for IoT botnet detection using benchmark datasets. The experimental design will focus on evaluating whether a layered detection approach can improve the balance between accuracy, early detection, lightweight operation, and generalization.

The research design will include dataset selection and preparation, data preprocessing and cleaning, feature selection and feature engineering, class imbalance handling, model development and training, model evaluation using classification and deployment-related metrics, comparative analysis of results, and interpretation of findings in relation to the research objectives. This design is suitable because the research objectives require measurable comparison between detection methods.

3.2.1 Alignment between Research Objectives and Methodology

To ensure that the research methodology directly addresses the research objectives, each objective is linked to a specific methodological action. This alignment demonstrates how the proposed research design will achieve the aim of developing an early, lightweight, and reliable IoT botnet detection approach.

Research Objective	Methodological Approach
To identify the main limitations of current machine learning-based IoT botnet detection methods.	This objective will be achieved through an extended literature review and research gap analysis focusing on early detection, computational cost, class imbalance, and generalization.
To examine the suitability of lightweight and deep learning models for resource-constrained environments.	This objective will be achieved by comparing lightweight machine learning models such as Decision Tree, Random Forest, AdaBoost, and Gradient Boosting with deeper models such as DNN or DRNN.
To analyze the effect of class imbalance and unseen botnet behavior.	This objective will be achieved by evaluating models using class-sensitive metrics such as recall, F1 score, AUC, MCC, false-positive rate, and false-negative rate, and by applying imbalance handling techniques such as SMOTE or class weighting.
To propose a conceptual layered detection approach.	This objective will be achieved by designing a two-stage detection framework in which a lightweight first layer screens traffic and a second layer performs deeper analysis for suspicious cases.

Research Objective	Methodological Approach
To define practical evaluation criteria for IoT botnet detection models.	This objective will be achieved by using predictive metrics and deployment-related metrics, including accuracy, precision, recall, F1 score, AUC, inference time, feature count, resource suitability, and generalization performance.

3.3 Dataset Selection

The study will use publicly available IoT botnet and cybersecurity datasets that have been commonly used in previous research. The main dataset expected to be used is the Bot-IoT dataset because it was used in several reviewed studies and contains both normal and attack traffic. It includes attack categories such as DDoS, DoS, scanning, keylogging, and data exfiltration, which are relevant to IoT botnet detection. The Bot-IoT dataset is appropriate for this study because it has been used in previous IoT botnet detection studies involving DNN, SMOTE-DRNN, and CNN-based BotIDS models (Jithu et al., 2021; Popoola et al., 2021; Idrissi et al., 2021).

If available, an additional dataset such as N-BaIoT may be used for cross-dataset validation. N-BaIoT is relevant because it includes botnet traffic generated by infected IoT devices, including Mirai and Bashlite attacks. Using more than one dataset can help evaluate whether the proposed detection approach performs well beyond one controlled dataset. The N-BaIoT dataset is also relevant because it was used in ELBA-IoT to evaluate lightweight ensemble-based botnet detection against Mirai and Bashlite traffic (Abu Al-Haija & Al-Dalaien, 2022).

The dataset selection criteria include IoT-related normal and malicious traffic, botnet or malware-related attack behavior, features suitable for machine learning classification, support for binary or multiclass classification, and suitability for evaluating class imbalance and generalization.

3.4 Data Collection Method

This study will rely on secondary data collection because it uses existing publicly available cybersecurity datasets rather than collecting new traffic from live IoT devices. Secondary data is appropriate for this research because the main objective is to design and evaluate a machine learning-based detection approach using established data sources.

The selected datasets will be downloaded from their official or recognized research repositories. Dataset documentation will be reviewed to understand the meaning of features, attack labels, data collection environment, and known limitations. This step is important to ensure that the data is interpreted correctly and used consistently during preprocessing and evaluation.

3.4.1 Materials and Method

The materials used in this research will consist of publicly available IoT botnet and cybersecurity datasets, selected machine learning tools, and evaluation metrics suitable for both classification performance and deployment readiness.

The main dataset planned for this study is the Bot-IoT dataset because it has been used in several previous IoT botnet detection studies and includes normal and attack traffic such as DDoS, DoS, scanning, keylogging, and data exfiltration. If available, the N-BaIoT dataset will also be considered

for cross-dataset validation because it contains botnet traffic generated from infected IoT devices, including Mirai and Bashlite attacks.

The proposed tools for model development and analysis include Python-based machine learning libraries such as Pandas and NumPy for data preparation, Scikit-learn for classical machine learning models, and TensorFlow or Keras for deep learning models if required. These tools are appropriate because they support preprocessing, model training, feature selection, class imbalance handling, and evaluation using multiple metrics.

The method will include four main experimental stages. First, baseline models will be trained using the original preprocessed dataset. Second, class imbalance handling techniques such as SMOTE, class weighting, or undersampling will be applied and compared. Third, feature selection will be used to evaluate whether a reduced feature set can maintain detection performance while lowering computational cost. Fourth, a conceptual layered detection approach will be evaluated by combining a lightweight first-stage model with a deeper second-stage model for suspicious or uncertain cases.

This materials and method structure supports the research aim by ensuring that the proposed detection approach is evaluated not only by accuracy, but also by practical factors such as inference time, feature count, class-sensitive performance, and generalization.

3.5 Data Preprocessing

Data preprocessing will be performed before model training and evaluation. The preprocessing stage is necessary because raw cybersecurity datasets often contain redundant columns, categorical values, missing values, highly skewed features, and imbalanced labels.

The preprocessing process will include data cleaning, label preparation, categorical encoding, feature scaling, train-test splitting, and class imbalance handling. Binary classification will distinguish benign and malicious traffic, while multiclass classification will distinguish different attack types. Class imbalance may be handled using SMOTE, class weighting, undersampling, or cost-sensitive learning.

3.6 Feature Selection and Feature Engineering

Feature selection is important for lightweight IoT botnet detection because using too many features may increase processing cost and reduce deployment suitability. The study will investigate feature reduction methods to identify the most useful features for detecting botnet behavior.

The feature selection stage may include removing features with low variance, removing highly correlated features, ranking features using statistical or model-based importance scores, and comparing model performance using full features and reduced feature sets. Feature engineering may also be used to create traffic-based indicators such as flow duration, packet rate, byte rate, protocol-related patterns, or behavioral indicators if these are available in the dataset.

3.7 Proposed Layered Detection Approach

Based on the literature review, this study proposes a conceptual layered detection approach. The purpose of this approach is to combine lightweight detection with deeper analysis.

The first layer will be a lightweight gateway-level screening model. This layer should be fast and suitable for resource-constrained environments. Models such as decision tree ensembles, random forest, gradient boosting, or other lightweight classifiers may be considered for this stage. The first layer will classify traffic quickly and identify clearly benign or clearly malicious behavior. This layer is inspired by lightweight ensemble-based detection, especially studies showing that decision-tree ensembles can achieve high accuracy with low inference overhead (Abu Al-Haija & Al-Dalaien, 2022).

The second layer will be used for suspicious or uncertain cases. This layer may use a deeper or more advanced model, such as a deep neural network, recurrent model, or transfer-learning-based approach. The purpose of the second layer is to improve detection reliability for difficult cases, minority attack classes, or uncertain traffic patterns. This layer is motivated by early and collaborative detection approaches that use richer behavioral evidence for suspicious cases (Nguyen et al., 2022).

The layered approach is expected to reduce computational cost because not all traffic needs to be processed by a heavy model. It also supports early response because the first layer can provide fast screening, while the second layer provides deeper analysis only when necessary.

3.8 Model Development

Several models will be considered and compared in this research. The exact models may depend on the selected dataset and available features, but the study will focus on models that represent different detection approaches.

The candidate models may include Decision Tree as a simple baseline, Random Forest as a stable ensemble model, Gradient Boosting or AdaBoost as lightweight ensemble methods, Deep Neural Network as a deep learning baseline, Recurrent Neural Network or DRNN if the dataset supports temporal representation, and a layered hybrid model combining lightweight screening with deeper analysis.

The models will be trained using the same training data and evaluated using the same testing protocol where possible. This will support fair comparison and allow the study to determine whether the proposed layered direction provides practical advantages.

3.9 Evaluation Metrics

The evaluation will use both classification metrics and deployment-related metrics. This is important because accuracy alone may be misleading, especially when datasets are imbalanced or when practical IoT deployment is considered. Class-sensitive metrics are necessary because previous work showed that high overall accuracy can hide weak minority-class detection in imbalanced IoT botnet datasets (Popoola et al., 2021).

The classification metrics will include accuracy, precision, recall, F1 score, false positive rate, false negative rate, AUC, and Matthews correlation coefficient. The deployment-related metrics will include inference time, feature count, model complexity, resource suitability, and generalization performance under unseen or cross-dataset conditions if data availability allows. Therefore, this study evaluates detection quality and practical deployment readiness rather than relying only on headline accuracy.

3.10 Data Analysis Plan

The data analysis will compare the performance of different models using the selected metrics. The analysis will not rely only on the highest accuracy score. Instead, it will examine the relationship between accuracy, class-sensitive performance, inference time, feature count, and generalization.

The analysis will compare baseline models, analyze performance on minority attack classes, compare full-feature and reduced-feature models, evaluate the effect of class imbalance handling methods, compare lightweight models with deep learning models, assess whether the layered approach improves the balance between speed and detection reliability, and identify the model or approach most suitable for practical IoT botnet detection.

3.11 Ethical Considerations

This research uses publicly available cybersecurity datasets and does not collect personal data directly from users. Therefore, the ethical risk is limited. However, cybersecurity datasets may still contain network-related information, so the study will use the data only for academic research purposes.

The study will not attempt to create, deploy, or execute malicious code. It will focus only on detection and analysis. All sources, datasets, and previous studies will be properly cited using APA style. The research will also avoid plagiarism by paraphrasing, synthesizing, and clearly acknowledging the work of other researchers.

3.12 Expected Outcomes

The expected outcomes of this research are directly linked to the research objectives and the gaps identified in the literature review. First, the study is expected to provide a clearer understanding of the limitations of current machine learning-based IoT botnet detection methods, especially in relation to early detection, computational cost, class imbalance, and generalization.

Second, the study is expected to identify which types of models are more suitable for lightweight IoT gateway deployment. This includes comparing classical machine learning models, ensemble learning models, and deep learning models using both predictive and deployment-related metrics.

Third, the research is expected to show how class imbalance affects the reliability of IoT botnet detection results. By using metrics such as recall, F1 score, AUC, MCC, false-positive rate, and false-negative rate, the study will provide a more balanced evaluation than accuracy alone.

Fourth, the research is expected to propose a conceptual layered detection approach that combines fast gateway-level screening with deeper analysis for suspicious or uncertain cases. This outcome supports

the practical need for a detection approach that is both lightweight and capable of deeper analysis when required.

Finally, the study is expected to provide practical evaluation criteria for future IoT botnet detection research. These criteria include accuracy, precision, recall, F1 score, AUC, MCC, inference time, feature count, model complexity, resource suitability, and generalization performance.

3.13 Gantt Chart

The following Gantt chart presents the planned research activities and milestones across the research project preparation period.

Task / Week	Week 1 22-28 Jul 2026	Week 2 29 Jul-4 Aug 2026	Week 3 5-11 Aug 2026	Week 4 12-22 Aug 2026
Finalize research title and objectives	X			
Review and organize previous literature	X	X		
Add additional studies to literature review	X	X		
Identify research gaps		X		
Design research methodology		X		
Select datasets and define preprocessing plan		X	X	
Define candidate models and evaluation metrics			X	
Develop proposed layered detection framework			X	
Write expected outcomes			X	
Prepare final report structure			X	X
Review citations and references				X
Final formatting and proofreading				X
Final submission				X

3.14 Research Milestones

- Milestone 1: Research topic, aim, and objectives finalized by the end of Week 1.
- Milestone 2: Literature review expanded and research gaps identified by the end of Week 2.
- Milestone 3: Research methodology and evaluation plan completed by the middle of Week 3.
- Milestone 4: Expected outcomes and Gantt chart completed by the end of Week 3.
- Milestone 5: Final report reviewed, formatted, and submitted by the end of Week 4.

3.15 Chapter Summary

This chapter presented the research methodology for the proposed study. The methodology follows a quantitative experimental design using publicly available IoT botnet datasets. The chapter explained dataset selection, data preprocessing, feature selection, model development, evaluation metrics, and the proposed layered detection approach. It also presented the expected outcomes, ethical considerations, Gantt chart, and project milestones. The proposed methodology directly addresses the research objectives by focusing on early detection, lightweight operation, class imbalance, and generalization to unseen botnet behavior.

References

- Abu Al-Haija, Q., & Al-Dalaien, M. (2022). ELBA-IoT: An ensemble learning model for botnet attack detection in IoT networks. *Journal of Sensor and Actuator Networks*, 11(1), 18. <https://doi.org/10.3390/jsan11010018>
- Catillo, M., Pecchia, A., & Villano, U. (2023). A deep learning method for lightweight and cross-device IoT botnet detection. *Applied Sciences*, 13(2), 837. <https://doi.org/10.3390/app13020837>
- Idrissi, I., Boukabous, M., Azizi, M., Moussaoui, O., & El Fadili, H. (2021). Toward a deep learning-based intrusion detection system for IoT against botnet attacks. *IAES International Journal of Artificial Intelligence*, 10(1), 110-120. <https://doi.org/10.11591/ijai.v10.i1.pp110-120>
- Jithu, P., Shareena, J., Ramdas, A., & Haripriya, A. P. (2021). Intrusion detection system for IoT botnet attacks using deep learning. *SN Computer Science*, 2, 205. <https://doi.org/10.1007/s42979-021-00516-9>
- Kadhim, Q. K., Sharhan Al-Sudani, A. Q. A., Almani, I. A., Alghazali, T., Dabis, H. K., Mohammed, A. T., Talib, S. G., Mahmood, R. A., Sahi, Z. T., & Mezaal, Y. S. (2022). IOT-MDEDTL: IoT malware detection based on ensemble deep transfer learning. *Majlesi Journal of Electrical Engineering*, 16(3), 47-54. <https://doi.org/10.52547/mjee.16.3.47>
- Khoa, T. V., Hoang, D. T., Trung, N. L., Nguyen, C. T., Quynh, T. T. T., Nguyen, D. N., Ha, N. V., & Dutkiewicz, E. (2022). Deep transfer learning: A novel collaborative learning model for cyberattack detection systems in IoT networks. *IEEE Internet of Things Journal*. <https://doi.org/10.1109/JIOT.2022.3202029>
- Nguyen, G. L., Dumba, B., Ngo, Q.-D., Le, H.-V., & Nguyen, T. N. (2022). A collaborative approach to early detection of IoT botnet. *Computers & Electrical Engineering*, 97, 107525. <https://doi.org/10.1016/j.compeleceng.2021.107525>
- Popoola, S. I., Adebisi, B., Ande, R., Hammoudeh, M., Anoh, K., & Atayero, A. A. (2021). SMOTE-DRNN: A deep learning algorithm for botnet detection in the Internet-of-Things networks. *Sensors*, 21(9), 2985. <https://doi.org/10.3390/s21092985>
- Rezaei, A. (2021). Using ensemble learning technique for detecting botnet on IoT. *SN Computer Science*, 2, 148. <https://doi.org/10.1007/s42979-021-00585-w>
- Srinivasan, S., & Deepalakshmi, P. (2023). Enhancing the security in cyber-world by detecting the botnets using ensemble classification based machine learning. *Measurement: Sensors*, 25, 100624. <https://doi.org/10.1016/j.measen.2022.100624>