

Course Title: Research Methodology in Computer Science

Course Code: AI800

Weighting: 20%

Lecturer: Dr. Mohammed Anbar

Hand-out Date: 19 Jun 2026

Hand-in Date: 04 July 2026

Hand-in Method: Online Submission via E-Learning.

Assignment Report: Literature Review – Machine Learning and Deep Learning Approaches for Botnet and Cyberattack Detection in IoT Networks

Student ID: 252000080

Student Name: Mohammed Hussein Bagharib

Table of Contents:

- Summary of the research papers
- Critical Analysis
- Connections
- Conflicts between the findings
- Synthesis Matrix for Literature Review
- References

Summary of the research papers:

Paper 1: SMOTE-DRNN: A Deep Learning Algorithm for Botnet Detection in the Internet-of-Things Networks

SMOTE-DRNN is a hybrid framework that combines SMOTE with DRNN to tackle class imbalance issues in IoT botnet detection. SMOTE solves the class imbalance problem by generating additional artificial attack samples, and the DRNN is used to learn the representation of what an attack looks like. They use the Bot-IoT dataset (UNSW Canberra) with 73,370,320 records of 11 different traffic classes, then apply information gain to reduce the feature space to the 10 most relevant features. SMOTE is used on the training data to augment the synthetic minority class samples before training the DRNN, and the evaluation is done with F1-score, AUC-ROC, and MCC metrics to account for imbalance. The main contributions are the system design and an in-depth ablation study that tested the necessity of both individual SMOTE and DRNN components by comparing the results. Results show that the model performed 99.62% F1-score, 99.87% AUC-ROC, and 99.62% MCC, which outperformed the RNN, BLSTM, CNN, and MLP baselines.

The primary drawbacks are:

1. The model required approximately 1,147 seconds to train, making real-time retraining impractical
2. DRNN has high computing cost. It is very difficult to apply it on the resource constrained IoT edge nodes
3. The work is not generalizable, as it is evaluated only on one dataset. *(Segun I. Popoola, Bamidele Adebisi, Ruth Ande, Mohammad Hammoudeh, Kelvin Anoh and Aderemi A. Atayero, 2021)*

Paper 2: ELBA-IoT: An Ensemble Learning Model for Botnet Attack Detection in IoT Networks

In this paper, the researchers present ELBA-IoT, an anomaly-based intrusion detection system (IDS) for IoT networks, that is based on an ensemble of three decision tree classifiers

1. AdaBoosted
2. RUSBoosted
3. Bagged.

The sizes of the classifiers are small enough to deploy on IoT devices, and the final decision is made based on their combined result. The system is tested on the N-BaloT-2021 dataset of 619,096 samples (611,359 normal, 7,737 botnet) that represent Mirai and Bashlite attacks on five IoT devices. The data preparation steps are Pearson Correlation Coefficient Score feature selection, Z-score normalization, and 5-fold cross-validation.

The model classifies traffic at three granularities:

1. Binary (normal vs. anomaly)
2. Ternary (normal, Mirai, Bashlite)
3. 10-class (normal, MIRAI_DANMINI_DOORBEL, MIRAI_ECOBEE_THERMOSTAT, MIRAI_PHILIPS_BABY_MONITOR , MIRAI_PROVISION_SECURITY_CAMERA , MIRAI_SAMSUNG_WEBCAM , GAFGYT_DANMINI_DOORBELL , GAFGYT_ECOBEE_THERMOSTAT , GAFGYT_PHILIPS_BABY_MONITOR , GAFGYT_PROVISION_SECURITY_CAMERA).

Main findings: ELBA-IoT reached 100% accuracy on both binary and ternary classification, 99.6% on the 10-class task with 97.7% F1-score, and 40 microseconds of inference overhead per sample (lowest of all compared methods)

The primary drawbacks are:

1. Our dataset only includes five IoT device types and two botnet families, and so it is not generalizable
2. Real deployment on IoT hardware is not tested
3. The large class imbalance between normal and botnet traffic is not explicitly addressed. *(Qasem Abu Al-Haija and Mu'away Al-Dala'ien, 2022)*

Paper 3: A Collaborative Approach to Early Detection of IoT Botnet

Contrary to most of the above papers, this paper's work concerns early indicators of a botnet attack. The paper proposes a cooperative machine learning model for early-stage detection of IoT botnet behavior, which could be effective for detection before the malware's complete execution. The authors execute IoT malware samples within a V-Sandbox environment, which is a virtual environment used to safely extract early behavior-based features including system calls, network connections, and file access activities. A late fusion ensemble of KNN, Random Forest, and KNN classifier used the probabilistic output of each model and combine it to reach the final decision regarding the classification. The dataset included in this paper contains a total of 8,911 samples that included benign IoT behavior as well as malicious behavior. The important results included in this paper include the model reaching an accuracy of 99.37% with an AUC of 0.9896. The overall classification time required was 6 seconds, and the model was able to detect botnet behavior in an early stage before it was completely executed.

The primary drawbacks are:

1. The V-Sandbox environment introduces approximately 3 minutes of overhead per sample, which is not realistic for real-time detection
2. The dataset of 8,911 samples is small compared to large-scale traffic datasets
3. Sophisticated botnets and malware actively try to detect if they are running inside a sandbox, these evasion techniques can easily defeat approach. *(Giang L. Nguyen, Braulio Dumba, Quoc-Dung Ngo , Hai-Viet Le and Tu N. Nguyen, 2022)*

Paper 4: Toward a Deep Learning-Based Intrusion Detection System for IoT Against Botnet Attacks

The researchers present in this paper BotIDS, a comparison of 4 deep learning architectures. Convolutional Neural Network (CNN), SimpleRNN, LSTM and GRU are trained and evaluated for 11-class botnet detection on Bot-IoT dataset with standard preprocessing techniques of normalization and one-hot encoding and all four architectures are trained and tested in the same conditions. Accuracy, loss and prediction time per inference are used to evaluate performance. Results reveal that CNN outperforms all the compared recurrent architectures with 99.94% of accuracy, 0.58% loss and 0.34 ms of prediction time. This indicates that local feature correlation using convolutional filters are better suited than sequential modelling for the task of network traffic classification.

The primary drawbacks are:

1. Class imbalance in the Bot-IoT dataset is not addressed. Given that the majority of the dataset is normal traffic this might lead to unrealistic result if the class imbalance is not handled using techniques like SMOTE.
2. An attacker can know that CNN is used and accordingly design his attack to bypass the CNN. Adversarial robustness is not assessed
3. The proposed CNN architecture is not well-documented, this could make it difficult to reproduce the experiment to verify the results. *(Idriss Idrissi, Mohammed Boukabous, Mostafa Azizi, Omar Moussaoui and Hakim El Fadili5 ,2021)*

Paper 5: Def-IDS: An Ensemble Defense Mechanism Against Adversarial Attacks for Deep Learning-Based Network Intrusion Detection

This paper introduces Def-IDS, a defense framework to protect against adversarial attacks on deep learning based intrusion detection systems (IDSs). Def-IDS combines a Multi-source Generative Adversarial Network (MGAN) of four generators where each generator is trained to generate adversarial examples based on one of the following four methods: FGSM, BIM, DeepFool, or JSMA with Multi-source Adversarial Training (MAT), which retrains the IDS classifier with an augmented dataset of both the original clean and a generated adversarial traffic. Def-IDS is evaluated on a state-of-the-art CSE-CIC-IDS2018 large scale modern dataset. Results show that without a defense the performance of DL classifiers on the adversarial testing samples dropped to 32.8 - 43.6% accuracy while Def-IDS can maintain an accuracy of about 97.9% and perform better than all single-source adversarial training baselines with a total training cost of 111.51 seconds.

The primary drawbacks are:

1. The system is not evaluated on IoT-specific traffic or deployed on IoT hardware.
2. Due to the expensive training cost of GAN, its deployment on resource constrained IoT edge devices are prohibitive
3. The dataset CSE-CIC-IDS2018 does not include IoT-specific botnet traffic such as Mirai or Bashlite. *(Jianyu Wang , Jianli Pan, Ismail AlQerm, and Yuanni Liu, 2021)*

Paper 6: Deep Transfer Learning: A Novel Collaborative Learning Model for Cyberattack Detection Systems in IoT Networks

This paper proposes Federated Transfer Learning (FTL), a cyberattack detection framework combining both Federated Learning (FL) and Transfer Learning (TL). The combined use of FL and TL enables participating IoT networks, all heterogeneous in their feature spaces and non-IID data distributions, to collaboratively train a distributed detection model, while not having to share their raw data. Transfer Learning allows heterogeneous feature spaces from different networks to be connected through a common representation space by using only around 10% of the data from each network that is exchanged, in order to align their shared layers for representation. The new FTL framework has been instantiated and evaluated on two cyberattack detection networks, using the KDD Cup 1999, NSL-KDD and UNSW-NB15 datasets. The major results showed that FTL achieved as high as 99% accuracy in attack detection, while increasing the accuracy against a baseline unsupervised distributed learning (UDL) model by as much as 40%. Only a marginal improvement was seen in FL when the networks' feature spaces were heterogeneous, making TL the critical component.

The primary drawbacks are:

1. The framework requires approximately 10% of mutual data between networks, which may be difficult to obtain in practice
2. Negative transfer can occur when participating networks have too dissimilar traffic patterns
3. The framework is only validated for two networks, and the KDD/NSL-KDD datasets are not IoT-specific. *(Tran Viet Khoa, Dinh Thai Hoang, Nguyen Linh Trung, Cong T. Nguyen, Tran Thi Thuy Quynh, Diep N. Nguyen, Nguyen Viet Ha and Eryk Dutkiewicz, 2022)*

Critical Analysis:

The six reviewed studies collectively demonstrate that ensemble and hybrid approaches consistently outperform single-model baselines for IoT botnet and cyberattack detection, and that class imbalance in network traffic datasets must be explicitly addressed — studies that ignore it and report only accuracy figures likely overstate their true detection capability. The most critical gap, however, is the near-universal assumption that test data will be clean and unperturbed. Evidence from within the research (research paper 5) shows that an undefended deep learning classifier operating above 99% accuracy on clean traffic can collapse to below 44% under adversarial manipulation, this means that the high accuracy claims of most reviewed studies may significantly overstate real-world security value. Compounding this, all studies evaluate in laboratory or simulated environments with no validation on actual IoT hardware, datasets vary widely in age, scope, and attack coverage with no common benchmark, and all but one study design for a single organization rather than the distributed, cross-network reality of IoT threat landscapes. Until adversarial robustness, deployment validation, dataset standardization, and collaborative learning become standard requirements rather than optional extensions, the field's results should be treated as promising but not yet deployable.

Connections:

1. Shared objective: All six papers aim to improve automated detection of malicious traffic in IoT networks, establishing a common security research agenda.
2. Ensemble approaches: Papers 2 (AdaBoosted / RUSBoosted /Bagged DTs), 3 (KNN+RF+KNN late fusion), and 5 (MGAN+MAT) all employ explicit ensemble strategies, and Papers 1 and 6 combine two complementary techniques, reflecting a field-wide consensus that single-model approaches are insufficient.
3. Common benchmark Bot-IoT: Papers 1 and 4 both evaluate on the same 11-class Bot-IoT dataset, enabling direct comparison and illustrating how metric choice and imbalance handling jointly determine apparent performance.
4. Collaborative and distributed paradigms: Papers 3 (collaborative ML with V-Sandbox) and 6 (federated transfer learning) both move beyond centralized single-organization models, connecting on the theme that IoT security benefits from coordinated multi-party approaches.
5. Research progression: The papers reflect natural maturation from maximizing detection accuracy in centralized settings (Papers 1, 2, 4), to early behavioral detection (Paper 3), to hardening the IDS against active adversarial manipulation (Paper 5), to enabling collaborative security across organizations (Paper 6).

Conflicts between the findings:

1. CNN vs. recurrent architectures: One study concludes that CNN outperforms LSTM, GRU, and SimpleRNN on the Bot-IoT dataset, while another demonstrates strong performance using a deep recurrent architecture on the same dataset. The conflict is not necessarily a genuine architectural disagreement, it most likely a result of one study ignoring class imbalance entirely, which inflates the apparent advantage of architectures that are biased toward majority traffic classes, rather than reflecting a true difference in detection capability.
2. Computational efficiency vs. detection thoroughness: The studies conflict fundamentally with what the primary design priority should be. One approach achieves 40-microsecond inference using lightweight decision tree ensembles explicitly targeting IoT gateway deployment. Another accepts over 1,100 seconds of training time to achieve rigorous imbalance correction with a deep recurrent model. A third requires approximately three minutes per sample for behavioral sandbox analysis. These are not minor engineering trade-offs, they reflect incompatible assumptions about where the system will run, how fast it must respond, and what resources are available.
3. Adversarial threat model: All but one study evaluates their models on clean, unperturbed test traffic, implicitly assuming attackers will not attempt to manipulate the detection system itself. The one study that tests this assumption directly finds that an undefended classifier drops from above 99% accuracy on clean data to between 32.8% and 43.4% when real adversary evasion techniques are employed (paper 5 table I). This means the high accuracy claims of the other studies describe performance against non-adaptive attackers only, and their real-world security value under active adversarial conditions is unknown.

Synthesis Matrix for Literature Review:

Study	Authors	Year	Research Design	Key Findings/Results	Methodology	Limitations
SMOTE-DRNN: A Deep Learning Algorithm for Botnet Detection in the Internet-of-Things Networks	Popoola et al.	2021	Experimental	99.62% F1; 99.87% AUC-ROC; 99.62% MCC; outperforms RNN, BLSTM, CNN, MLP	Dataset: Bot-IoT (73M records, 11 classes); 10-feature selection via information gain; SMOTE oversampling; DRNN classifier; Metrics: F1, AUC-ROC, MCC	High training time (~1,147s); single dataset; high compute cost; no adversarial robustness test
ELBA-IoT: An Ensemble Learning Model for Botnet Attack Detection in IoT Networks	Abu Al-Haija & Al-Dala'ien	2022	Experimental	100% accuracy (binary & ternary); 99.6% accuracy, 97.7% F1 (10-class); 40 μ s inference overhead	Dataset: N-BaloT-2021 (619K samples; Mirai + Bashlite; 5 IoT devices); Pearson CCS feature selection; Z-score normalization; 5-fold CV; AdaBoosted + RUSBoosted + Bagged DT ensemble	Dataset limited to 5 devices and 2 botnet families; no real IoT hardware deployment; class imbalance not explicitly addressed
A Collaborative Approach to Early Detection of IoT Botnet	Nguyen et al.	2022	Experimental	99.37% accuracy; AUC=0.9896; ~6s classification time; pre-execution botnet detection validated	Dataset: 8,911 behavioral samples from V-Sandbox; Late fusion ensemble: KNN + Random Forest + KNN; Metrics: ACC, AUC-ROC, detection time	~3-min V-Sandbox overhead; small dataset (8,911 samples); vulnerable to sandbox evasion; simulated environment may not reflect real IoT behavior
Toward a Deep Learning-Based Intrusion Detection System for IoT Against Botnet Attacks	Idrissi et al.	2021	Experimental	CNN: 99.94% accuracy, 0.58% loss, 0.34ms prediction time; CNN outperforms SimpleRNN, LSTM, and GRU	Dataset: Bot-IoT (11 classes); Comparative study: CNN vs. SimpleRNN vs. LSTM vs. GRU; Standard normalization and one-hot encoding; Metrics: accuracy, loss, prediction time	Class imbalance not addressed; no adversarial robustness evaluation; single dataset; CNN architecture details insufficiently documented
Def-IDS: An Ensemble Defense Mechanism Against Adversarial Attacks for Deep Learning-Based Network	Wang et al.	2021	Experimental	~97.9% accuracy under adversarial attack; undefended DL drops to 40–70%; total training cost ~111.51s	Dataset: CSE-CIC-IDS2018; MGAN with 4 generators (FGSM, BIM, DeepFool, JSMA); Multi-source Adversarial Training (MAT) on augmented dataset; Metrics:	Not evaluated on IoT traffic or hardware; high compute cost; no IoT-specific botnet data; GAN training instability

Intrusion Detection					accuracy, F1 on clean and adversarial traffic	
Deep Transfer Learning: A Novel Collaborative Learning Model for Cyberattack Detection Systems in IoT Networks	Khoa et al.	2022	Experimental	Up to 99% accuracy; 40% improvement over UDL baseline; FL alone insufficient for heterogeneous features	Datasets: KDD Cup 1999, NSL-KDD, UNSW-NB15; Two-network FTL = Federated Learning (FL) + Transfer Learning (TL); ~10% mutual data for shared representation; Metric: accuracy	Requires ~10% mutual data; negative transfer risk for dissimilar networks; validated for only two networks; KDD/NSL-KDD are legacy non-IoT datasets

References:

- Abu Al-Haija, Q., & Al-Dala'ien, 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>
- 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 (IJ-AI)*, 10(1), 110–120. <https://doi.org/10.11591/ijai.v10.i1.pp110-120>
- 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*, 9(11), 8825–8836.
<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 and 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>
- Wang, J., Pan, J., AlQerm, I., & Liu, Y. (2021). Def-IDS: An Ensemble Defense Mechanism Against Adversarial Attacks for Deep Learning-Based Network Intrusion Detection. *Proceedings of the 30th International Conference on Computer Communications and Networks (ICCCN 2021)*. IEEE.
<https://doi.org/10.1109/ICCCN52240.2021.9522215>