

A Novel Framework for Threat Detection in IoT

Abdullah Mohammed Alamri

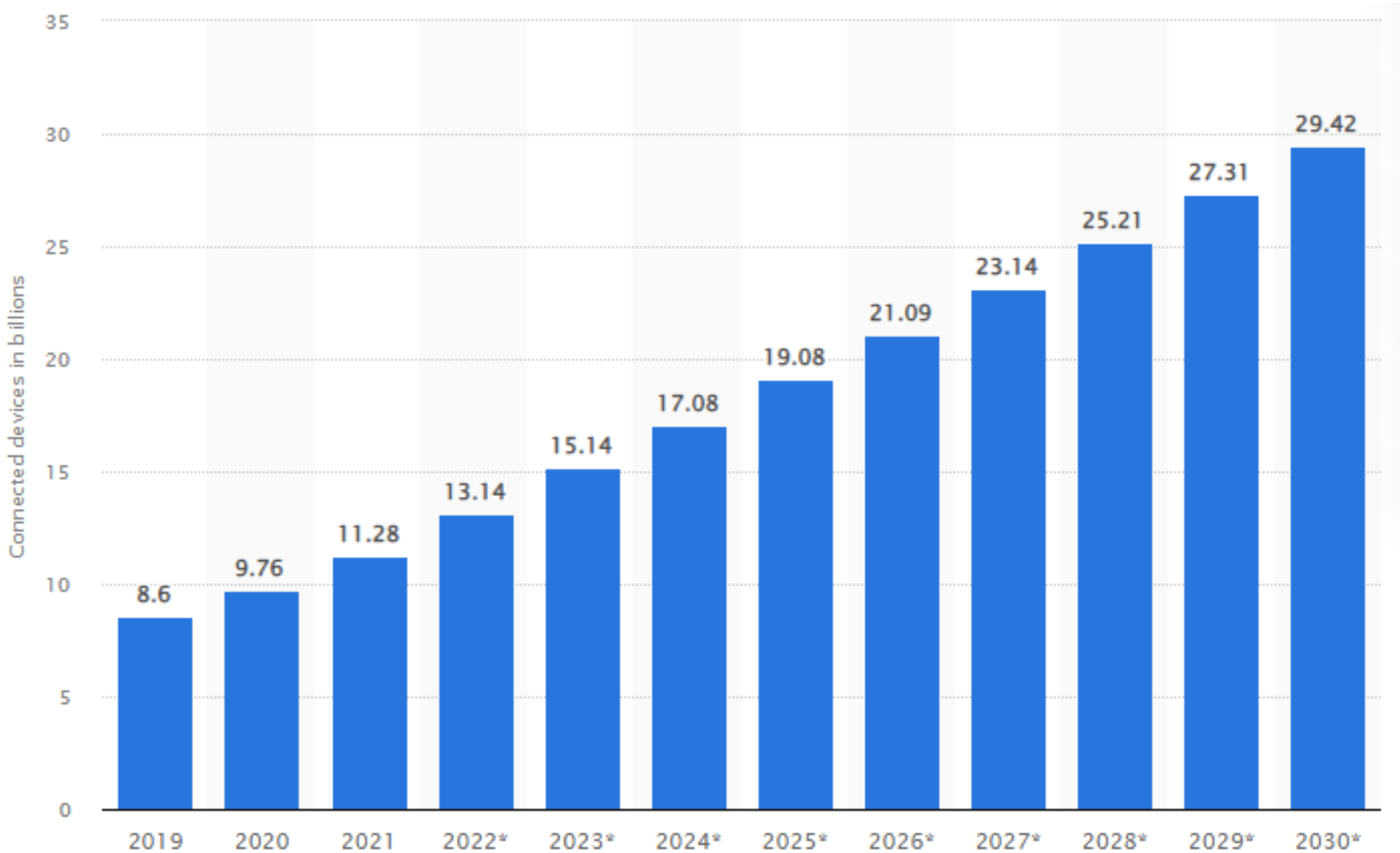
Executive Director of Kafa’at Training Center in Riyadh

Supervised By

Dr: Ghada Dahy Fathy Kamel

Prof. Hesham N. Elmahdy

Cairo : 2025

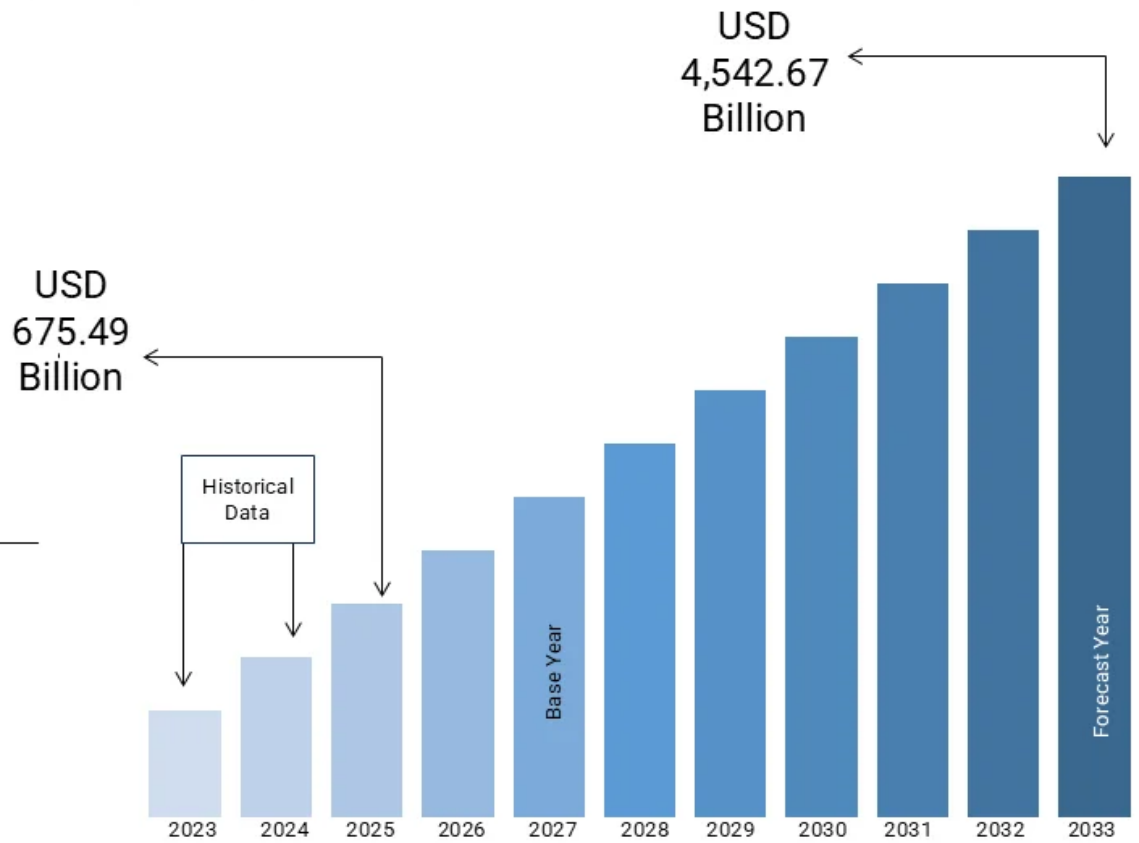

Connected devices in billions
35
30
25
20
15
10
5
0
8.6
9.76
11.28
13.14
15.14
17.08
19.08
21.09
23.14
25.21
27.31
29.42
2019
2020
2021
2022*
2023*
2024*
2025*
2026*
2027*
2028*
2029*
2030*

Global Internet of Things (IoT) Market

Market Size Overview

26.9%

Global market CAGR,
2025 - 2033



www.marketdataforecast.com

Source: Market Data Forecast Analysis

Agenda

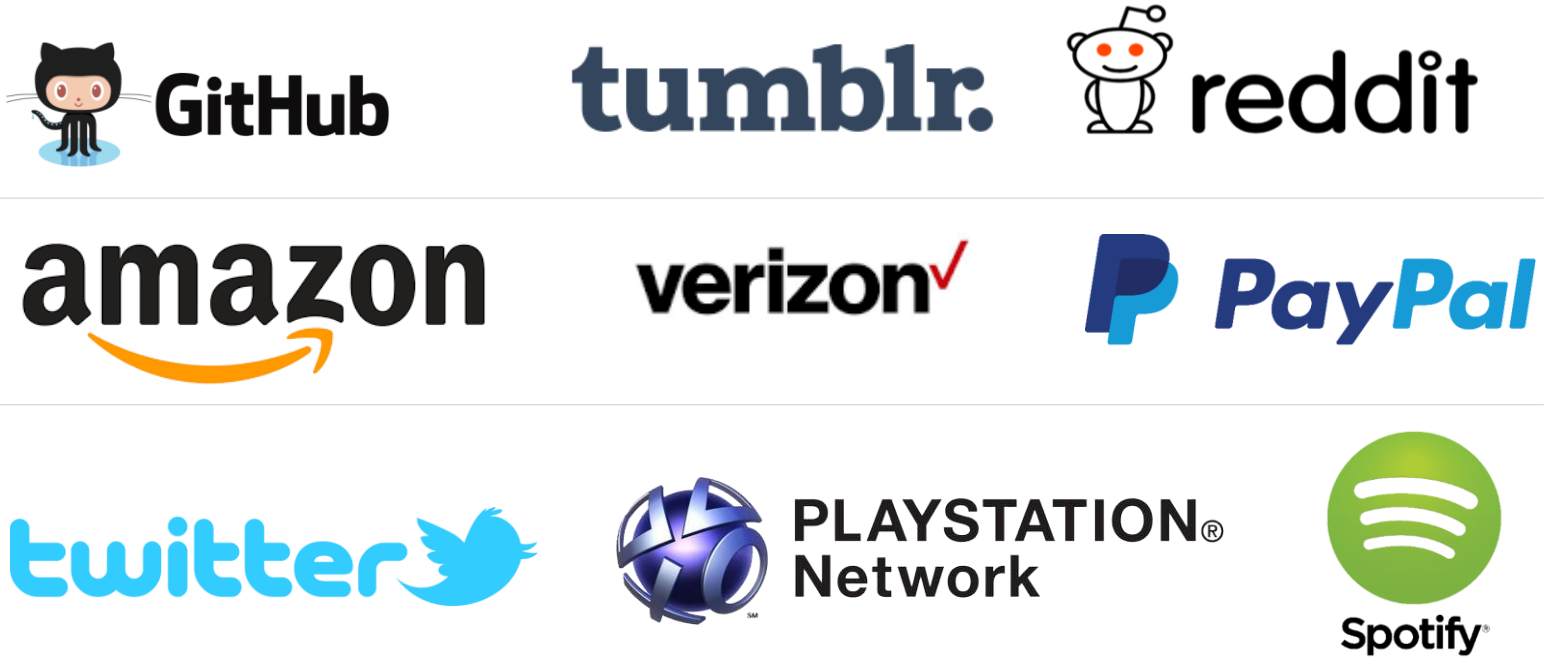
- Introduction
- Problem statement
- Related work
- Research gap
- Research questions
- Research objectives
- Proposed Framework
- Experimental results and Discussing
- Conclusion and future work





Introduction

IoT DDOS Attack, Oct 2016

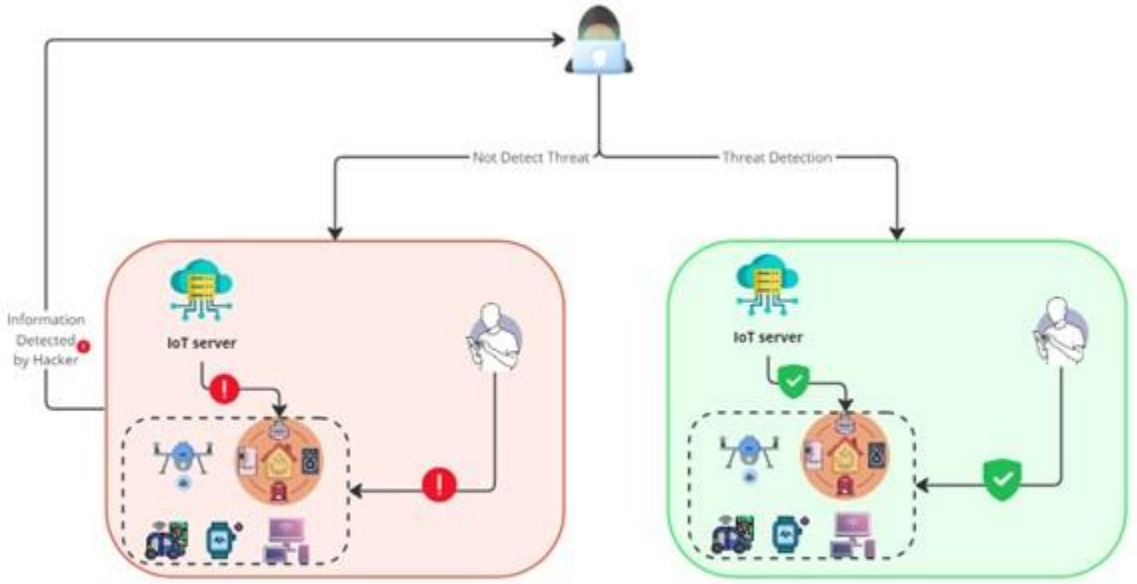


- 2016: Mirai botnet DDoS attack exposed IoT vulnerabilities.
- Study uses TON-IoT dataset with SelectKBest + tuned Random Forest.
- Goal: improve IoT attack detection & mitigation.

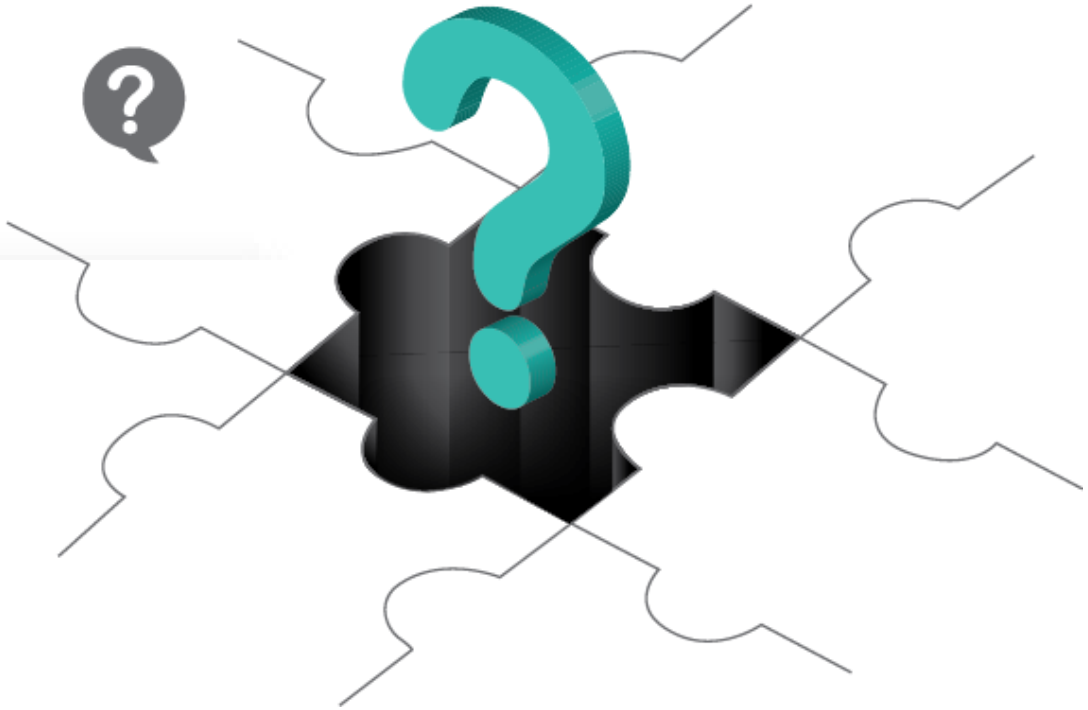


Introduction

- IoT growth (sensors, 5G, cloud) expands connectivity and risks.
- Diverse protocols and limited resources hinder unified security.
- IDS + ML enable early, accurate attack detection with less human effort.

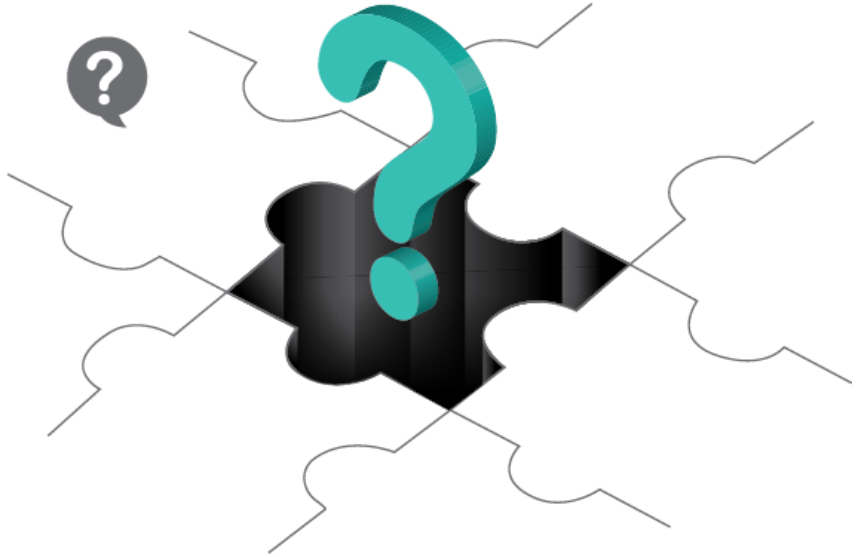


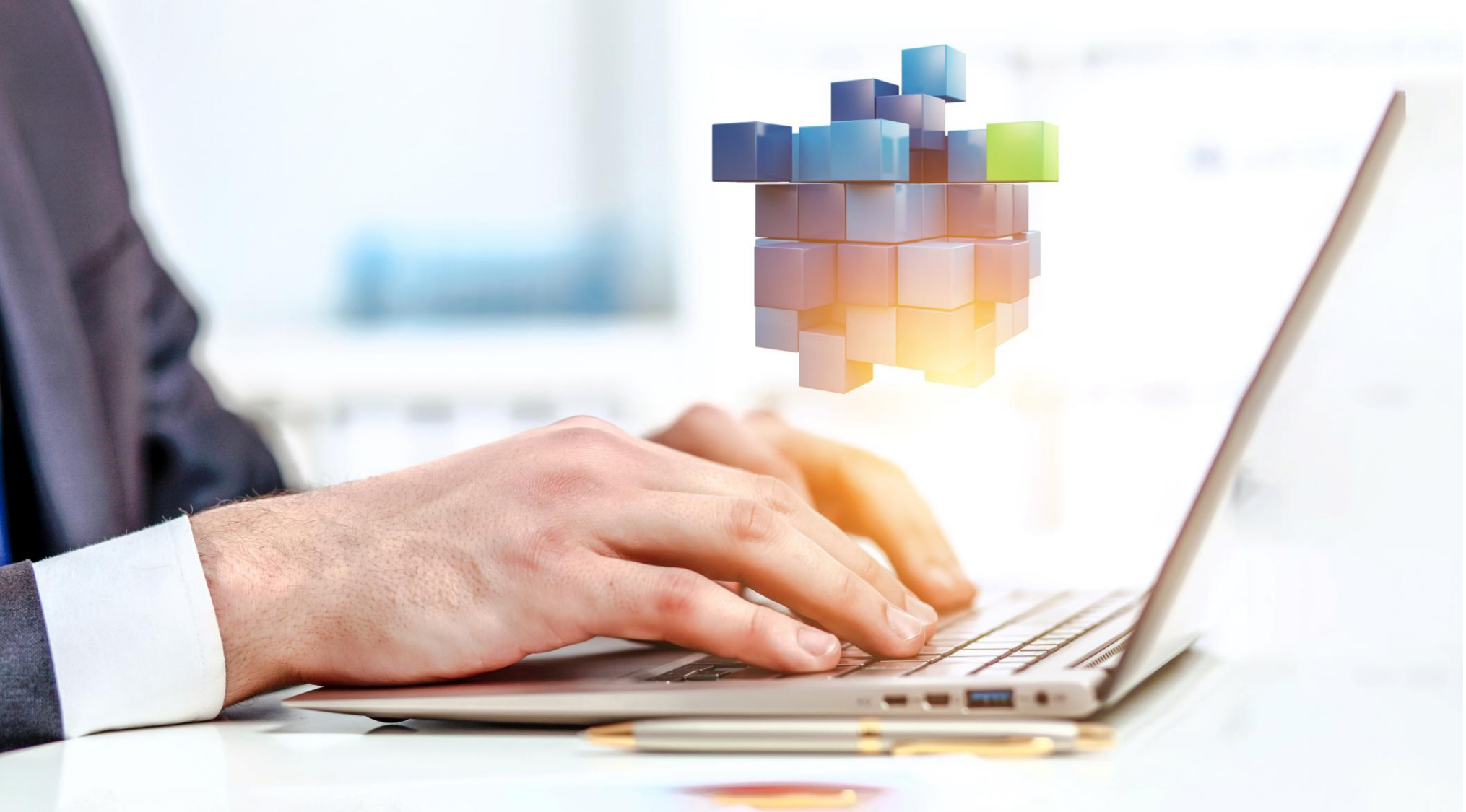
Problem Statement



Problem Statement

- IoT growth increases risks (DDoS, breaches).
- Diverse devices/protocols complicate security.
- Traditional IDS fail against new attacks.
- **Need adaptive, intelligent detection** for IoT.





Related Works

Related Works

Ref	Year	Authors	Methodology	Accuracy	Key Contributions
[3]	2025	Narkedimilli et al.	Adaptive Curriculum Learning + XAI (LIME, SHAP)	97%	Integrated XAI; curriculum learning with transparent explanations for IoT IDS
[7]	2025	Jamshidi et al.	Evaluation of ML-driven IDS (accuracy & energy)	96%	Benchmarked IDS performance vs energy consumption for IoT devices
[46]	2025	Adewole et al.	Ensemble Learning + Rule Induction for Explainability	97%	Rule-explainable ensemble IDS; lightweight for deployment
[14]	2024	Nguyen and Beuran	FedMSE (Semi-supervised FL; Centroid OCC + Shrink Autoencoder)	97%	Addresses heterogeneity and resource limits via semi-supervised FL
[5]	2024	Li et al.	Feature Selection vs Feature Extraction (PCA)	94%	Compared dimensionality reduction methods, showing efficiency & accuracy
[13]	2024	Narayan et al.	Random Forests (CICIoT2023)	96%	Improved detection of underrepresented attack classes, addressed class imbalance
[16]	2023	Karunamurthy et al.	Federated Learning + Chimp Optimization Algorithm (Feature Selection)	95%	Enhanced IoT IDS while preserving privacy
[6]	2023	Mustafa et al.	Ensemble Learning with Feature Selection	96%	Showed effectiveness of feature selection + ensemble classifiers in IoT IDS
[14]	2022	Devine et al.	Federated Learning with SVM	96%	Balanced detection accuracy and efficiency, suitable for edge devices



Research Gap

Research Gap

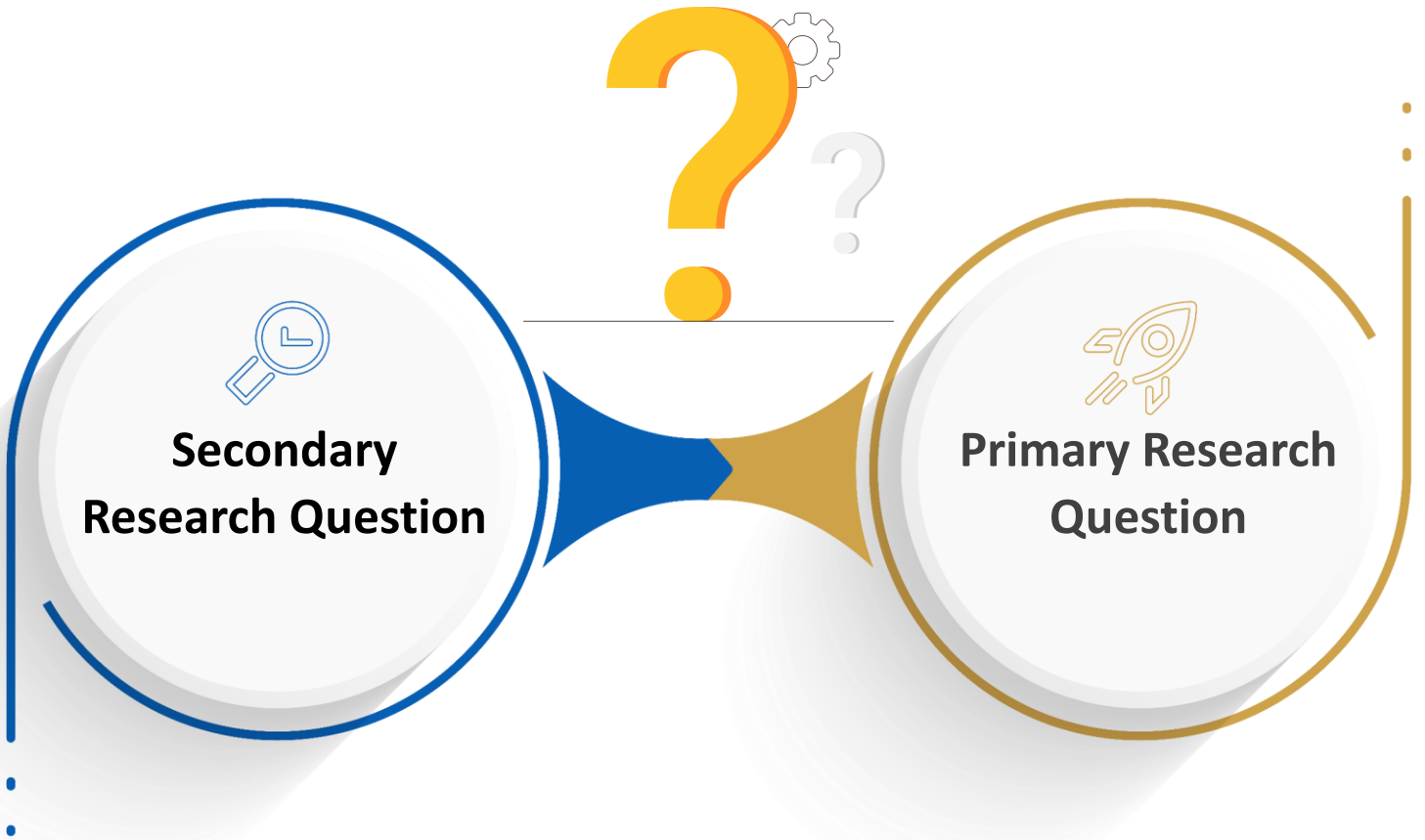
- **Dataset:** Reliance on static benchmarks, limited real-time validation.
- **Analysis:** Lack of statistical exploration of feature relationships.
- **Features:** Limited study on impact of feature reduction/selection for faster training.
- **Tuning:** Underutilization of hyperparameter optimization for optimal performance.
- **Model:** Few studies on Random Forest potential for complex, high-dimensional IoT data.



Research Questions



Research Questions



Research Questions



How can optimized machine learning algorithms, combined with effective feature selection, be leveraged **to enhance accuracy, efficiency, and adaptability** of **IoT intrusion detection systems** using **datasets like TON-IoT**?

Primary Research Question

Research Questions





1. What are the key vulnerabilities and threats faced by IoT systems?
2. What limitations do traditional security solutions have in IoT?
3. How effective are current ML models at detecting IoT threats?
4. What challenges do ML algorithms face in IoT security, such as efficiency and overfitting?
5. Which feature selection methods best identify relevant threat detection features in the “TON-IoT” dataset?

Research Questions





6. How do feature selection techniques affect ML model performance in IoT threat detection?
7. What are the computational challenges of using ML for IoT security, and how can they be addressed?
8. How can ML models be validated for real-world IoT scenarios?
9. What best practices can generalize the proposed model across different IoT domains?



Research Objectives



Primary Research Objectives

To develop a **machine learning-based** intrusion detection framework for IoT environments, integrating **effective feature selection** and data **balancing techniques**, and validated on the TON-IoT dataset

Research Objectives



ML Model Efficacy

Assess the performance of various ML models in detecting IoT threats.

Optimal Feature Selection

Apply advanced feature selection methods to the ‘TON-IoT' dataset to reduce computational load and enhance detection accuracy.

Evaluation on ‘TON-IoT'

Use TON-IoT dataset to benchmark with accuracy, precision, recall, and F1-score.

ML and Feature Selection Integration

Integrate feature selection + ML models for efficient IoT threat detection.

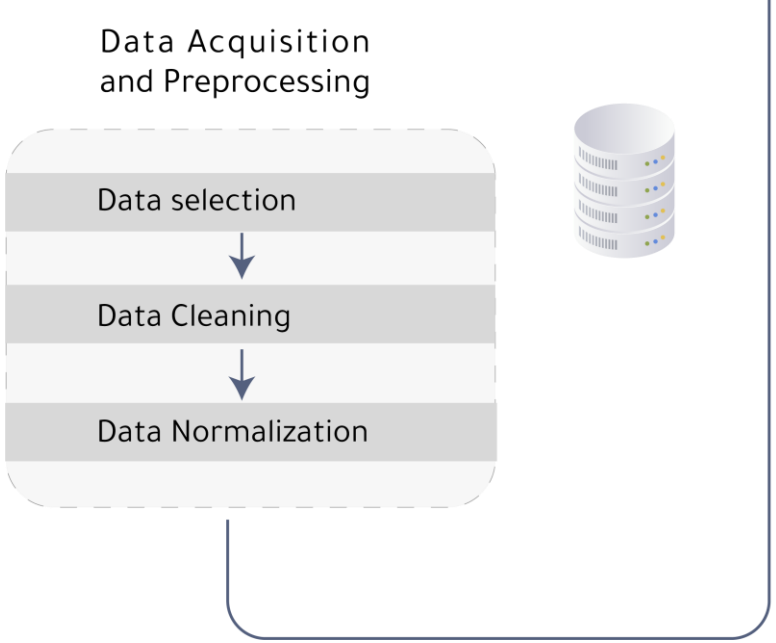
Real-time Threat Detection

Test the framework's ability to detect threats in real-time Internet of things environments.

Proposed Framework

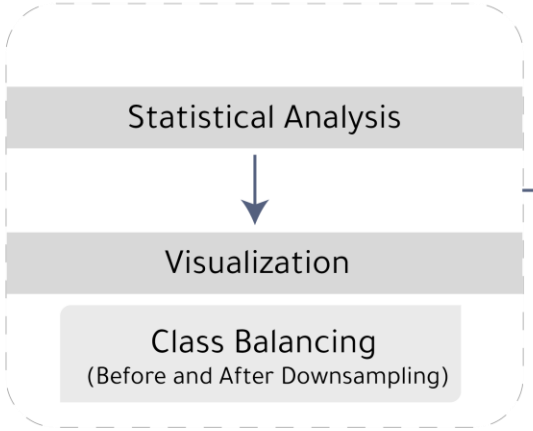


Proposed Framework



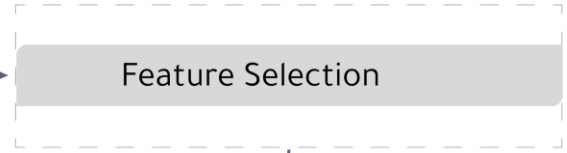


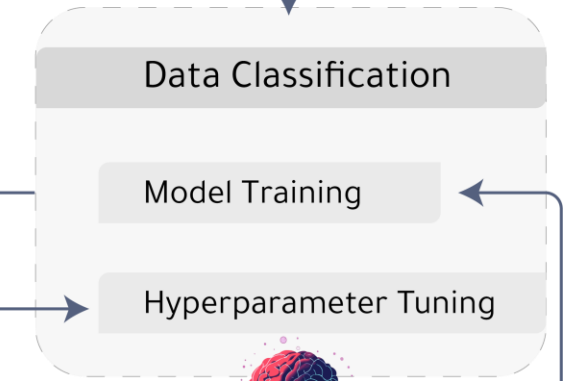
Initial Feature Assessment





Feature Selection



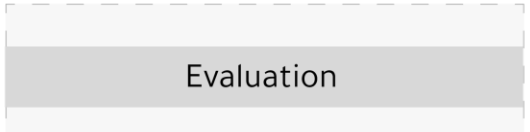


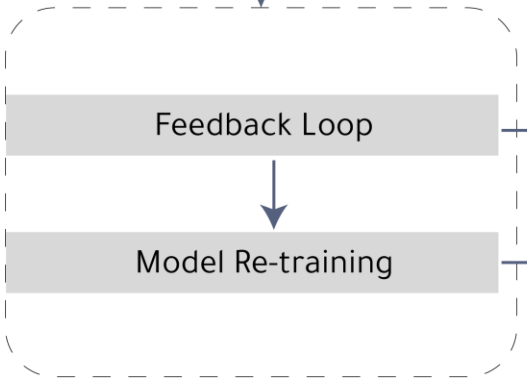


Machine Learning Model Development



Evaluation Performance



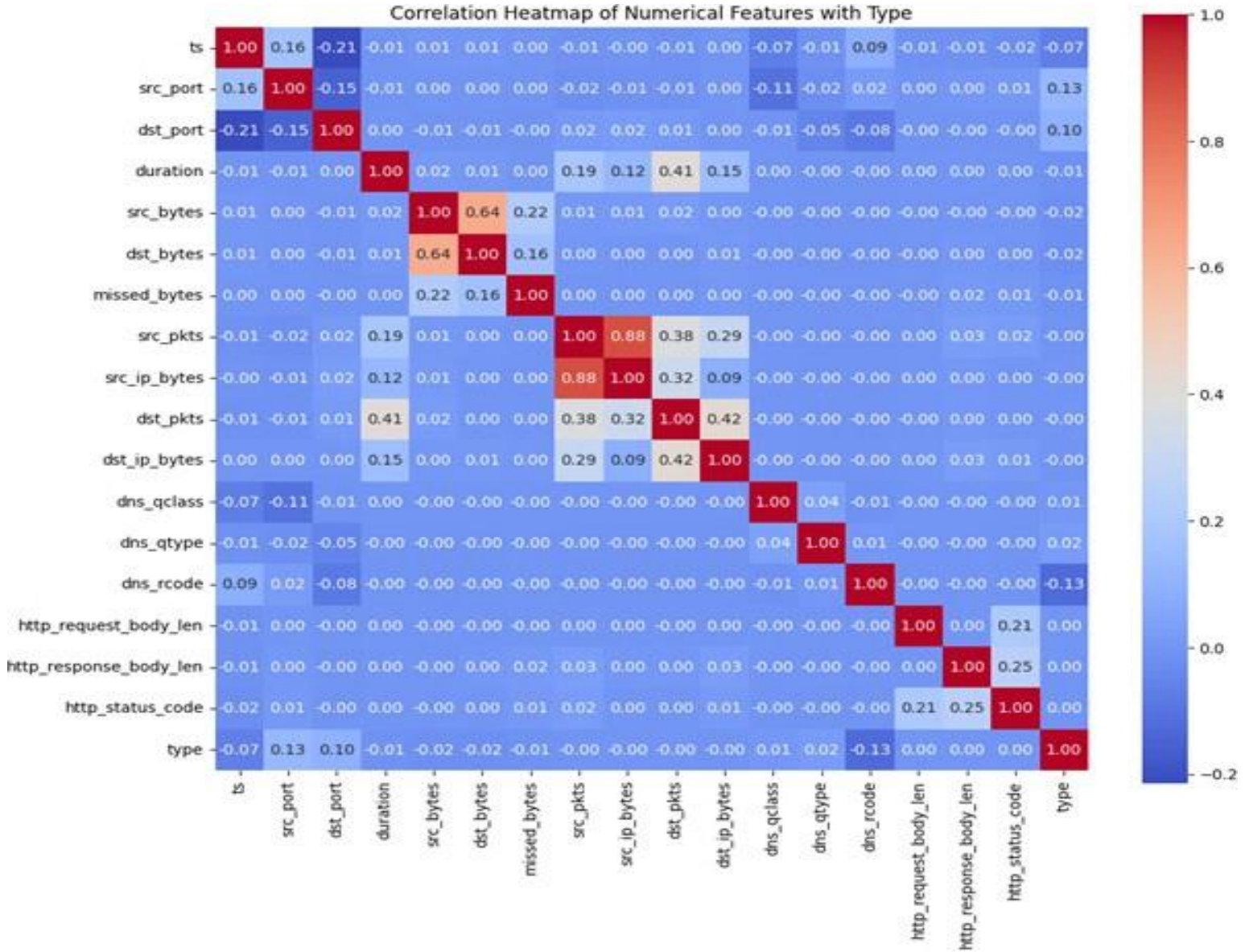




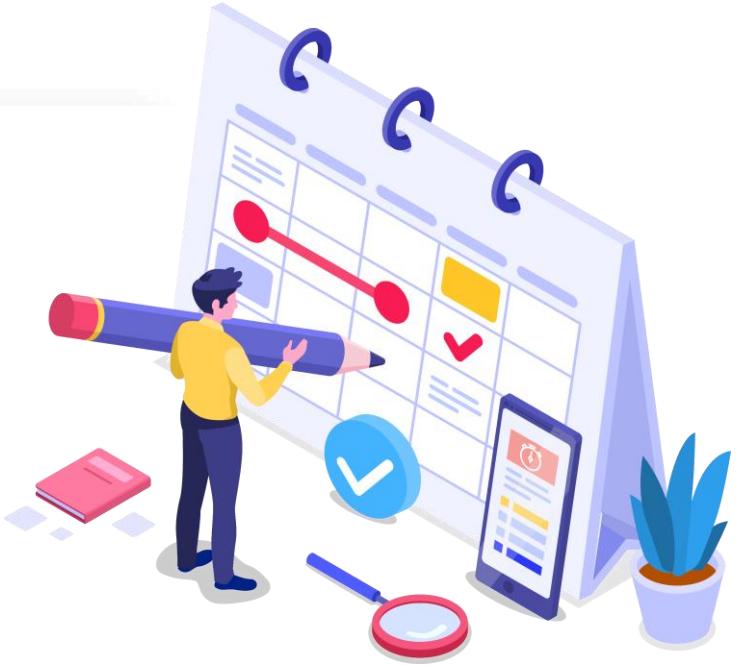
Iterative Refinement

Proposed Framework

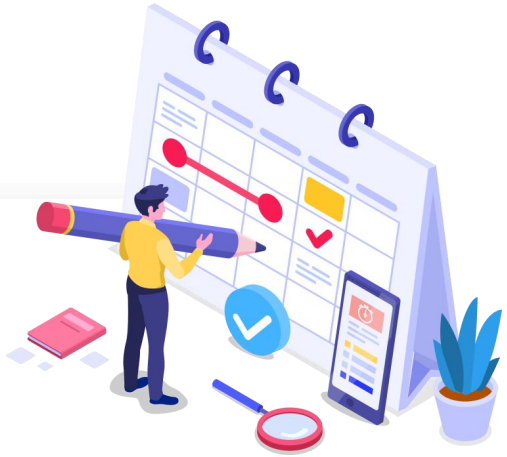
- **Feature Selection** is performed utilizing the SelectKBest method, with mutual information that considered as the scoring criterion. Mutual information measures the reduction in uncertainty about one variable when another is known as in Eq.



EXPERIMENTAL RESULTS

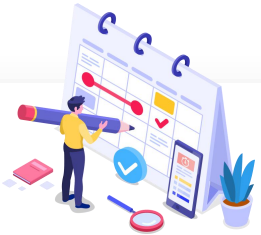


EXPERIMENTAL RESULTS

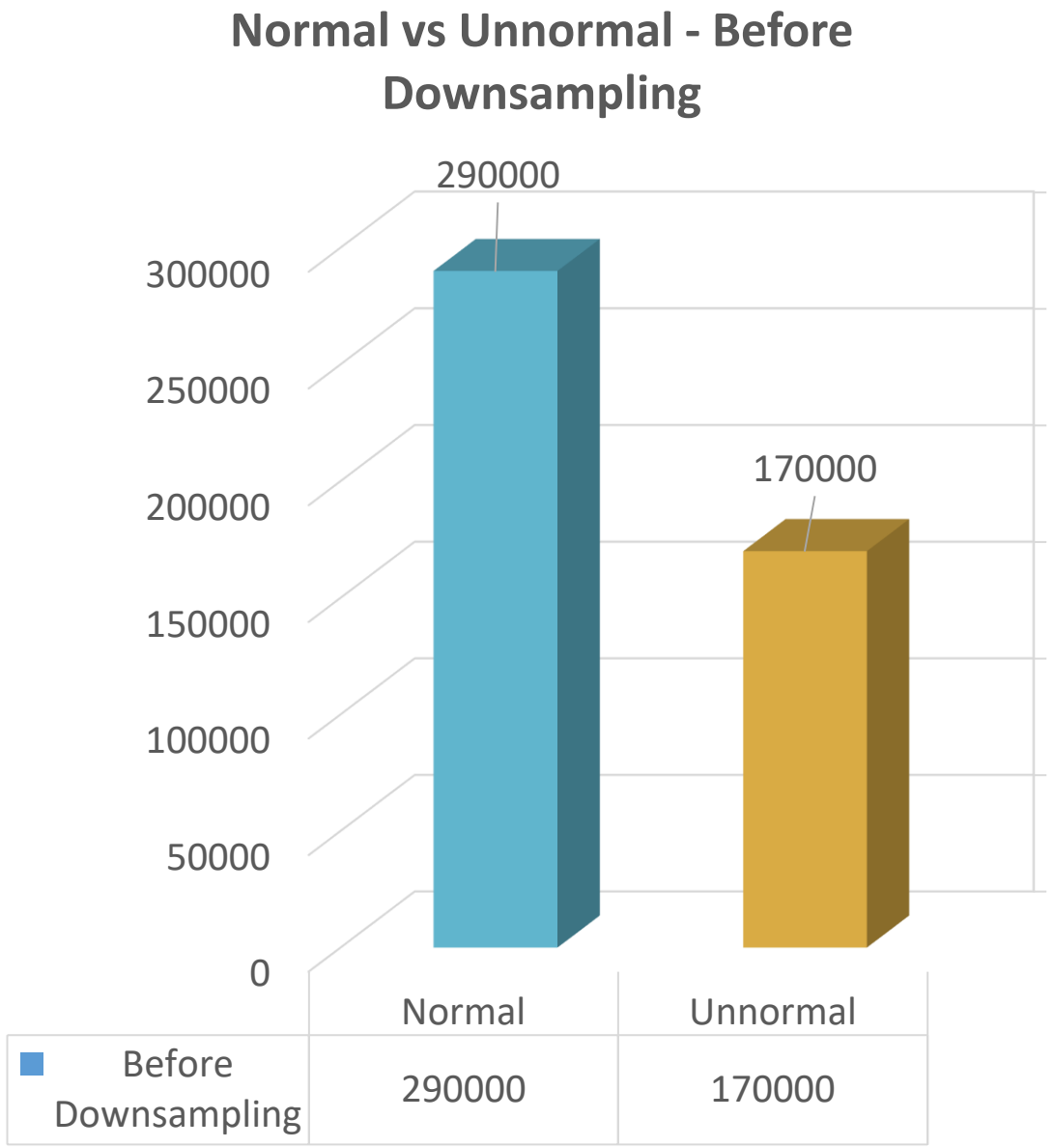


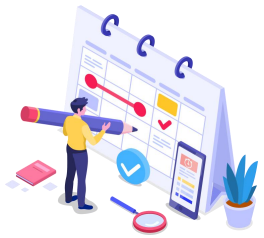
Experimental setup: Three phases—data preprocessing, feature selection, and ML model development/evaluation—run on **an Intel i7-12700K, 32GB RAM, RTX 3080 system.** Implemented in Python 3.9 with **Scikit-learn and related libraries**, evaluated by accuracy.

EXPERIMENTAL RESULTS



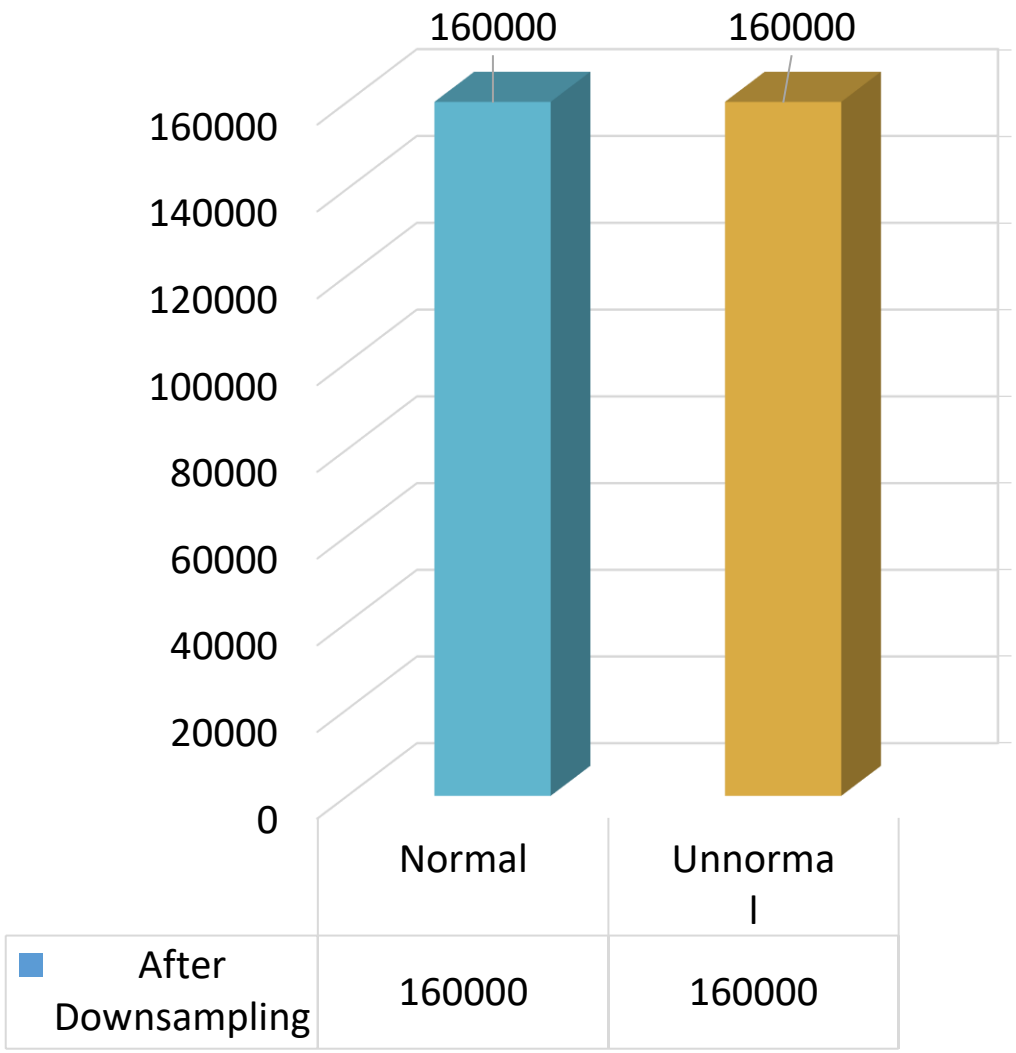
The TON-IoT dataset contains 461,043 samples and 45 features, covering normal and attack events for AI-based evaluation.

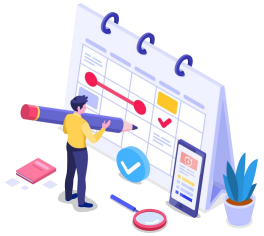




EXPERIMENTAL RESULTS

- **Class Distribution Adjustment:**
  - TON_IoT class imbalance was addressed using Downsampling.
  - Before balancing: "Normal" class = 65%, attacks < 35% → model bias and poor threat detection.

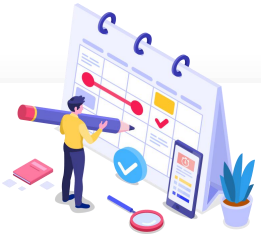




EXPERIMENTAL RESULTS

- **Class Distribution Adjustment:**
  - After balancing: Each class = 160,000 Samples → fair training, improved detection, and better generalization.

EXPERIMENTAL RESULTS



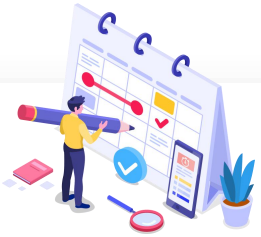
Binary Classification Report (Normal vs. Attack) after Downsampling

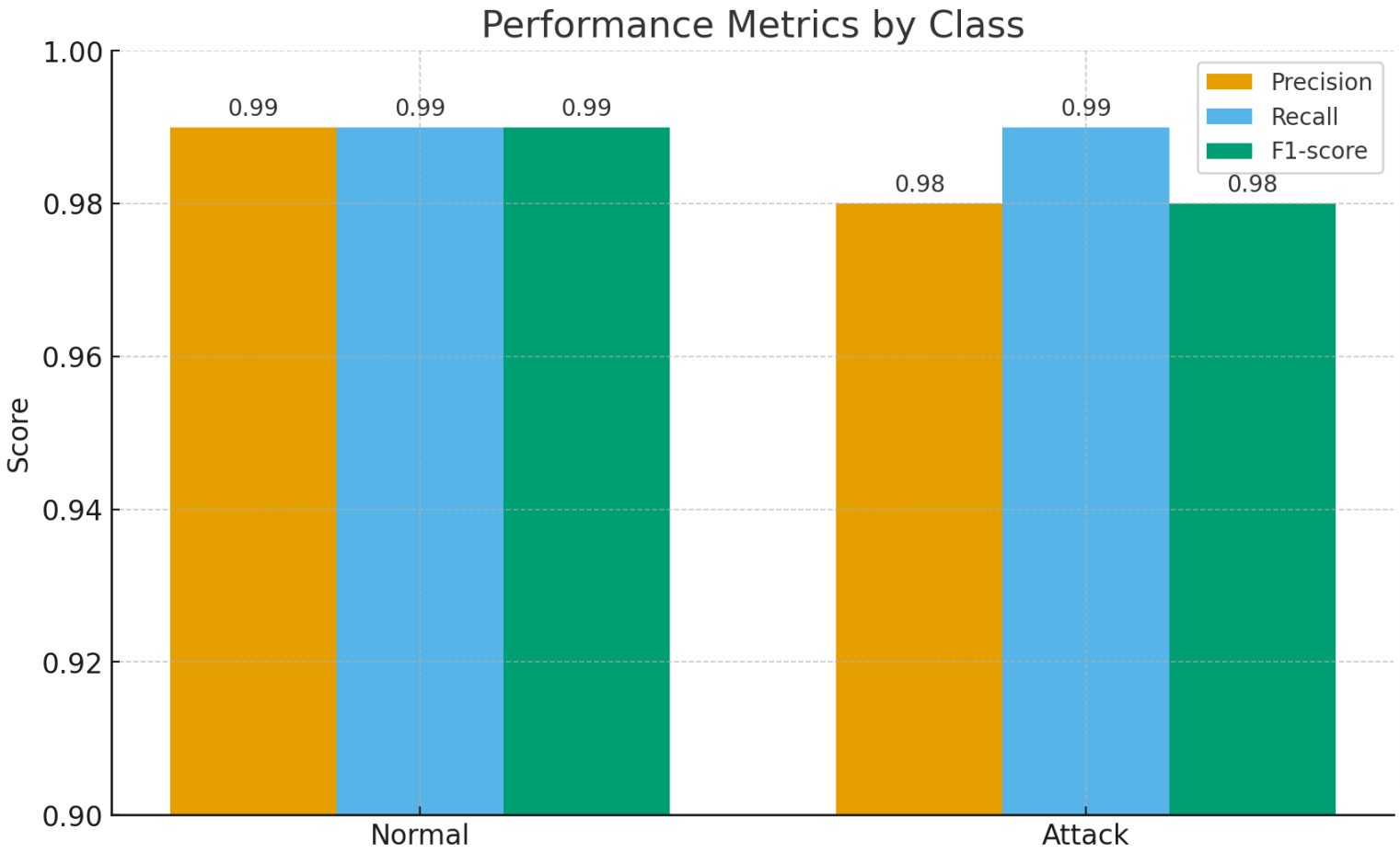
After downsampling, the number of Normal and Attack samples became equal (160k each).

The model achieved Accuracy = 98.79% with Macro/Weighted F1 ≈ 98.64%.

Class	Precision	Recall	F1-score	Support
Normal	0.99	0.99	0.99	160,000
Attack	0.98	0.99	0.98	160,000
Overall Accuracy			98.79%	320,000
Macro Avg	98.50%	98.79%	98.64%	
Weighted Avg	98.50%	98.79%	98.64%	

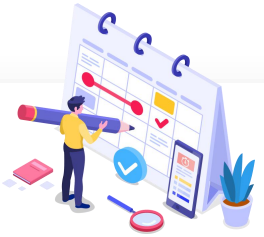
EXPERIMENTAL RESULTS



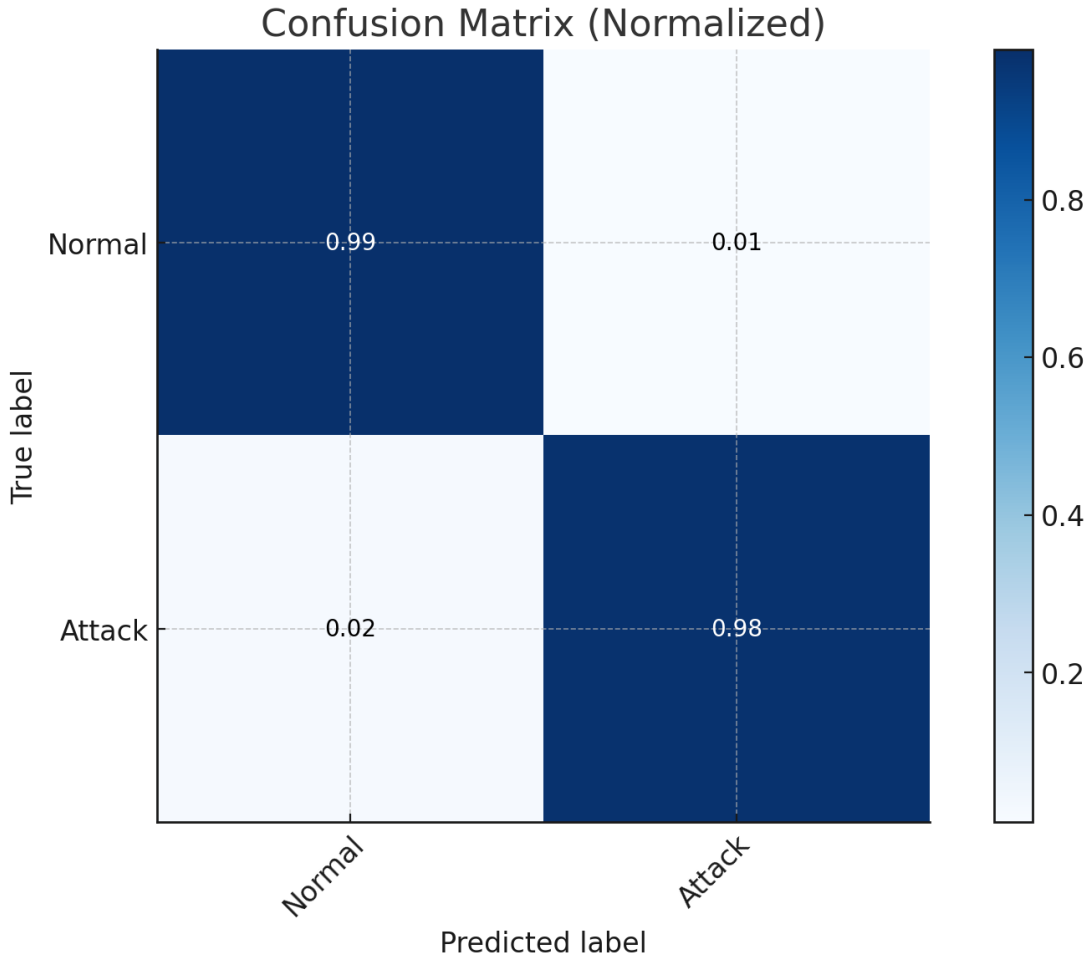


Binary Classification Report (Normal vs. Attack) after Downsampling

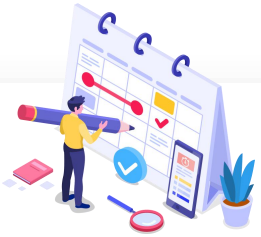
EXPERIMENTAL RESULTS



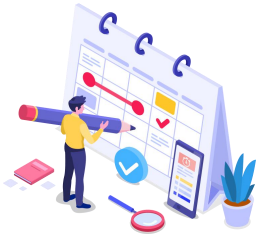
Binary Classification Report (Normal vs. Attack) after Downsampling



EXPERIMENTAL RESULTS

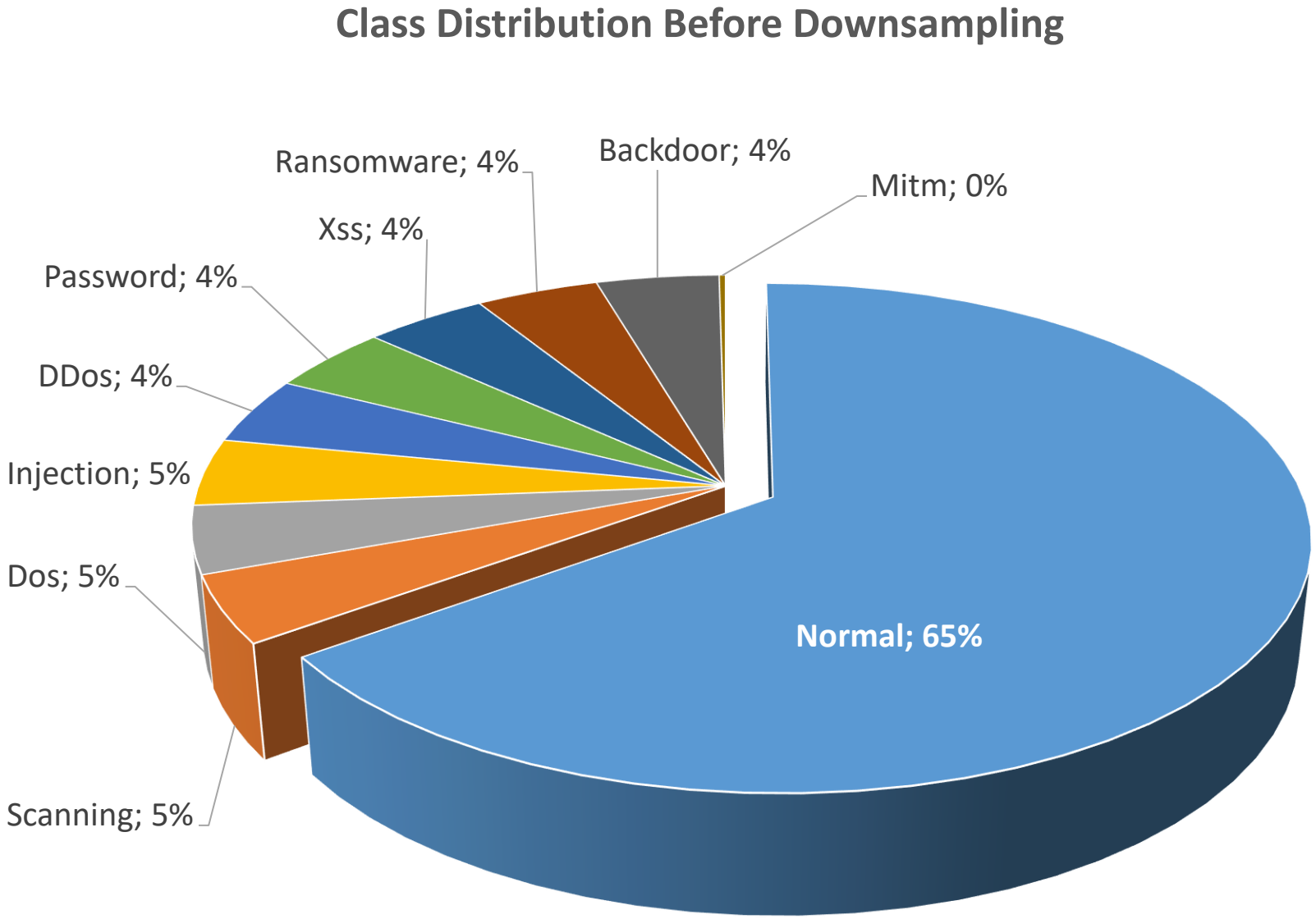


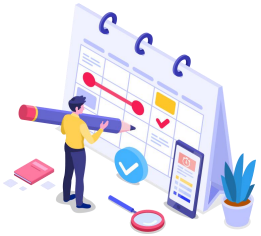
- **Class Distribution Adjustment:**
  - The TON_IoT dataset shows significant imbalance across classes.
  - Before balancing: the Normal class dominates with 65%, while individual attack categories (DoS, DDoS, Scanning, Injection, Password, XSS, Ransomware, Backdoor) range only between 4–5%, and MitM ≈ 0%.
  - This skew causes model bias, reducing fairness and leading to poor detection of minority attacks.



EXPERIMENTAL RESULTS

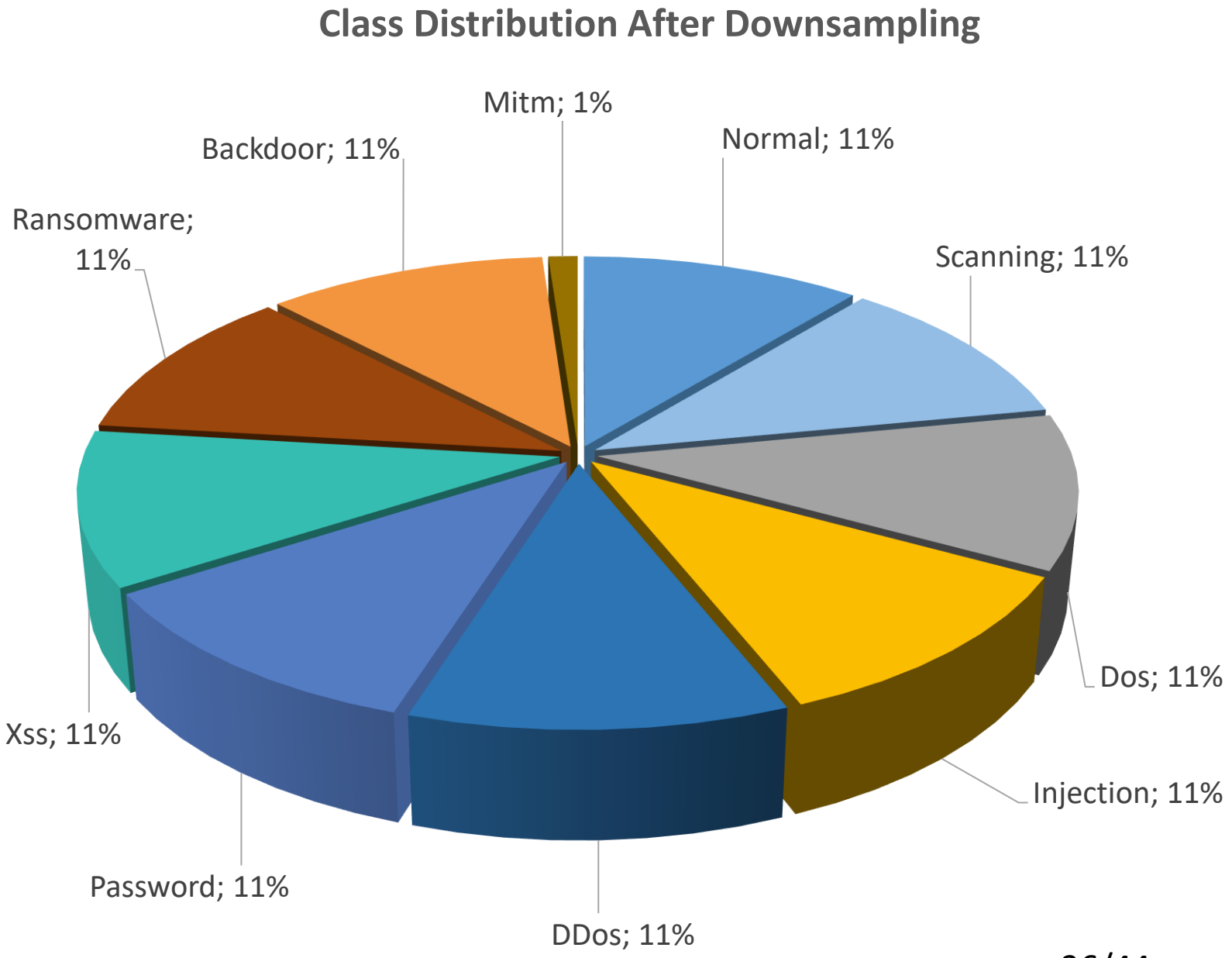
- **Class Distribution Adjustment:**
  - Before balancing: the Normal class dominates with 65%, while individual attack categories (DoS, DDoS, Scanning, Injection, Password, XSS, Ransomware, Backdoor) range only between 4–5%, and MitM ≈ 0%.



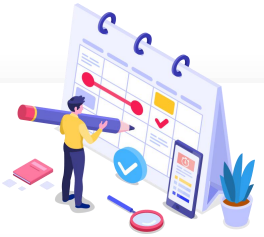


EXPERIMENTAL RESULTS

- **Class Distribution Adjustment:**
  - TON_IoT class imbalance was addressed using Downsampling.
  - After balancing: Each class = 11% → fair training, improved detection, and better generalization.



EXPERIMENTAL RESULTS



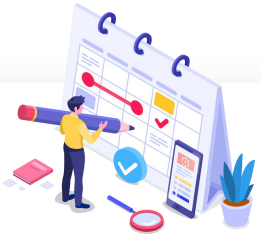
Multi-Class Classification Report on TON-IoT (after Downsampling)

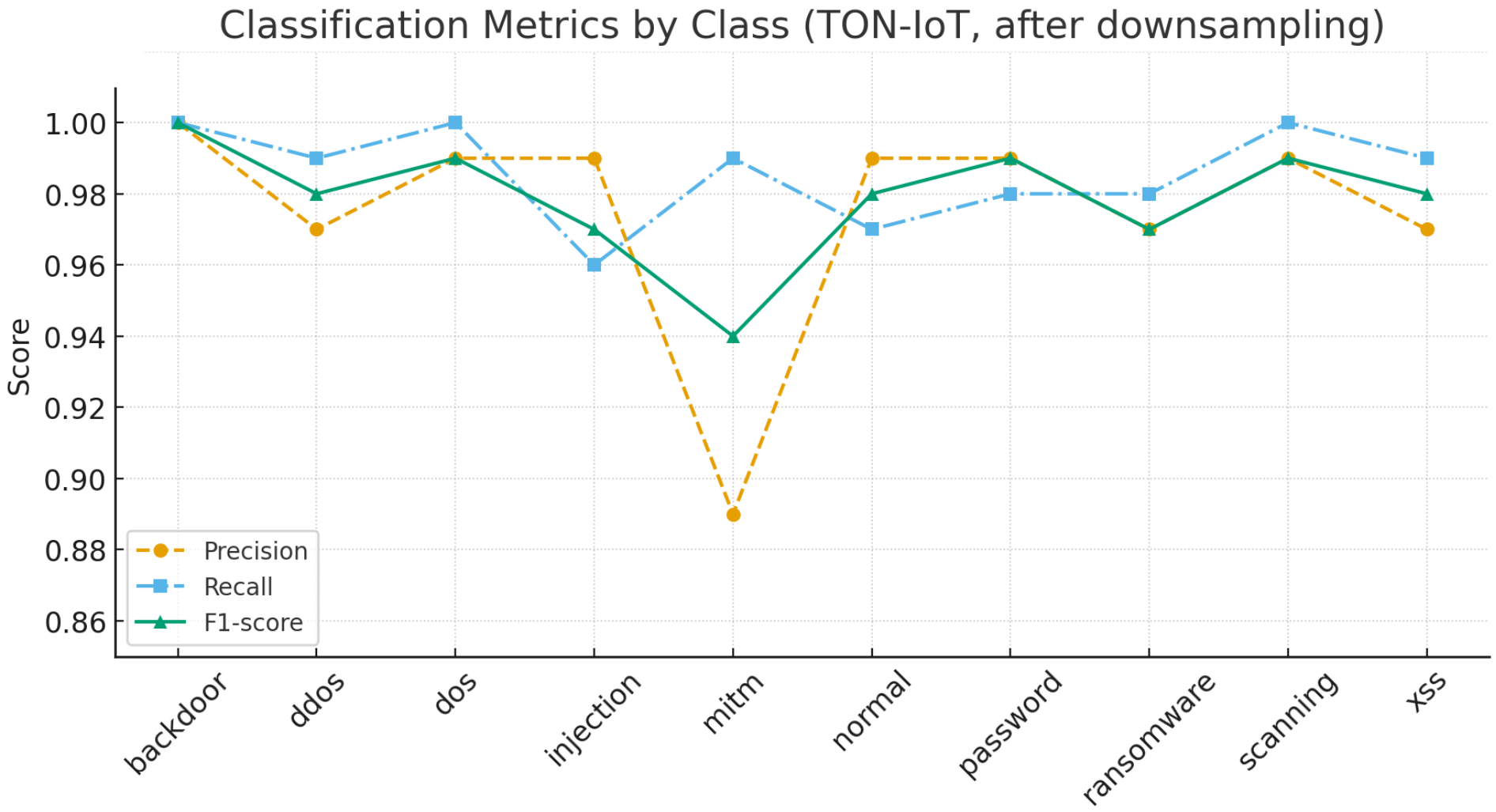
Class Distribution Adjustment:

Balanced per-class sampling (~50,715 each; MitM 4,610) delivers **Accuracy 98.57%**, **Macro-F1 98.01%**, **Weighted-F1 98.45%** with strong per-class scores and a precision dip on **MitM (0.88)**.

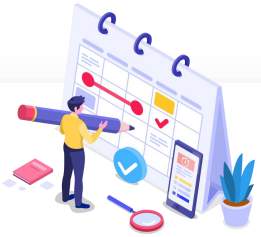
Class	Precision	Recall	F1-score	Support
backdoor	1.00	1.00	1.00	50,715
ddos	0.97	0.99	0.98	50,715
dos	0.99	1.00	0.99	50,715
injection	0.99	0.96	0.97	50,715
mitm	0.89	0.99	0.94	4,610
normal	0.99	0.97	0.98	50,715
password	0.99	0.98	0.99	50,715
ransomware	0.97	0.98	0.97	50,715
Scanning	0.99	1.00	0.99	50,715
Xss	0.97	0.99	0.98	50,715
Overall Accuracy			98.57%	461045
Macro Avg	97.50%	98.59%	98.01%	461045
Weighted Avg	98.35%	98.57%	98.45%	461045

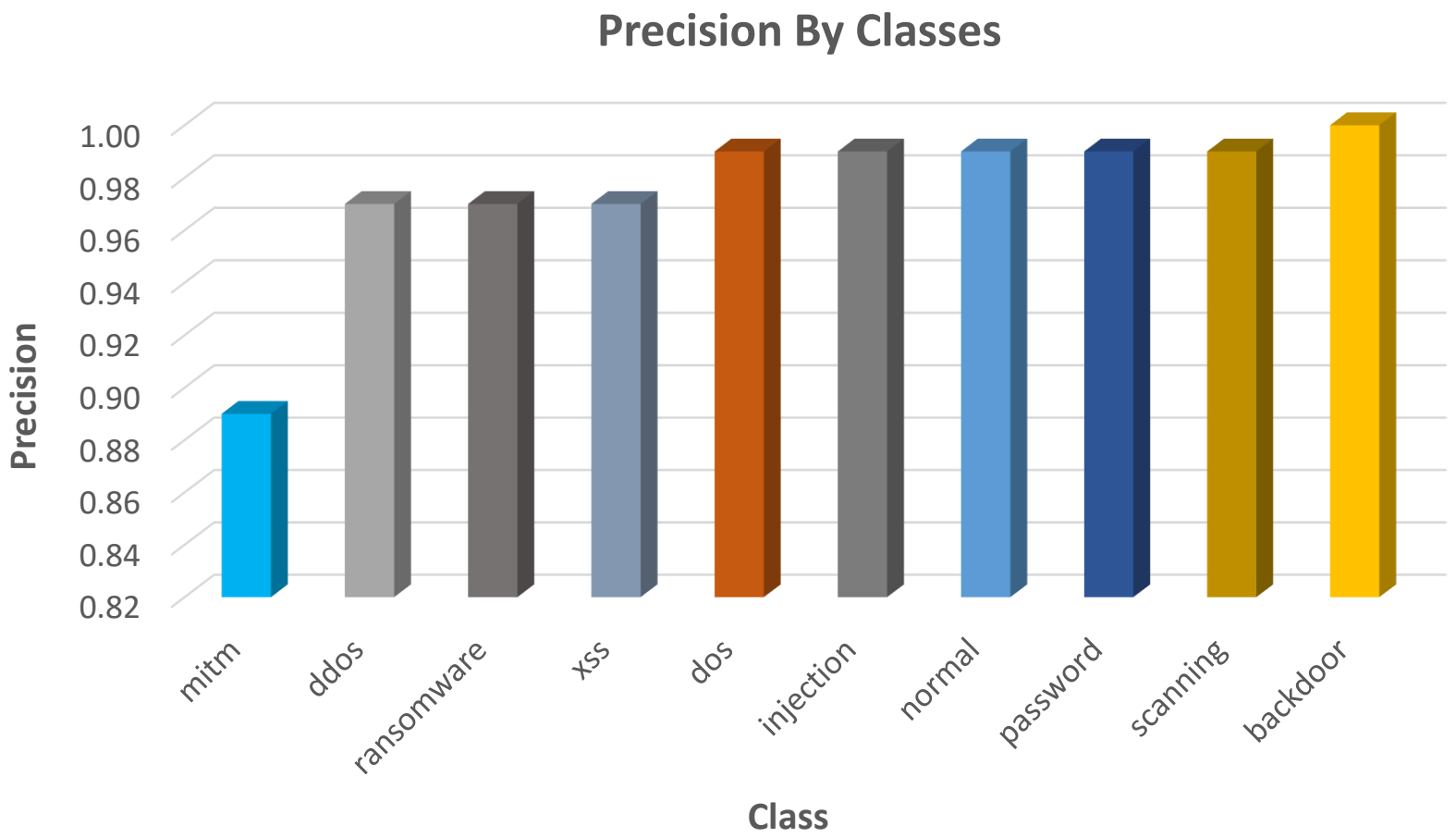
EXPERIMENTAL RESULTS



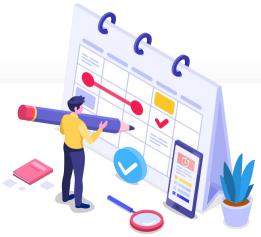


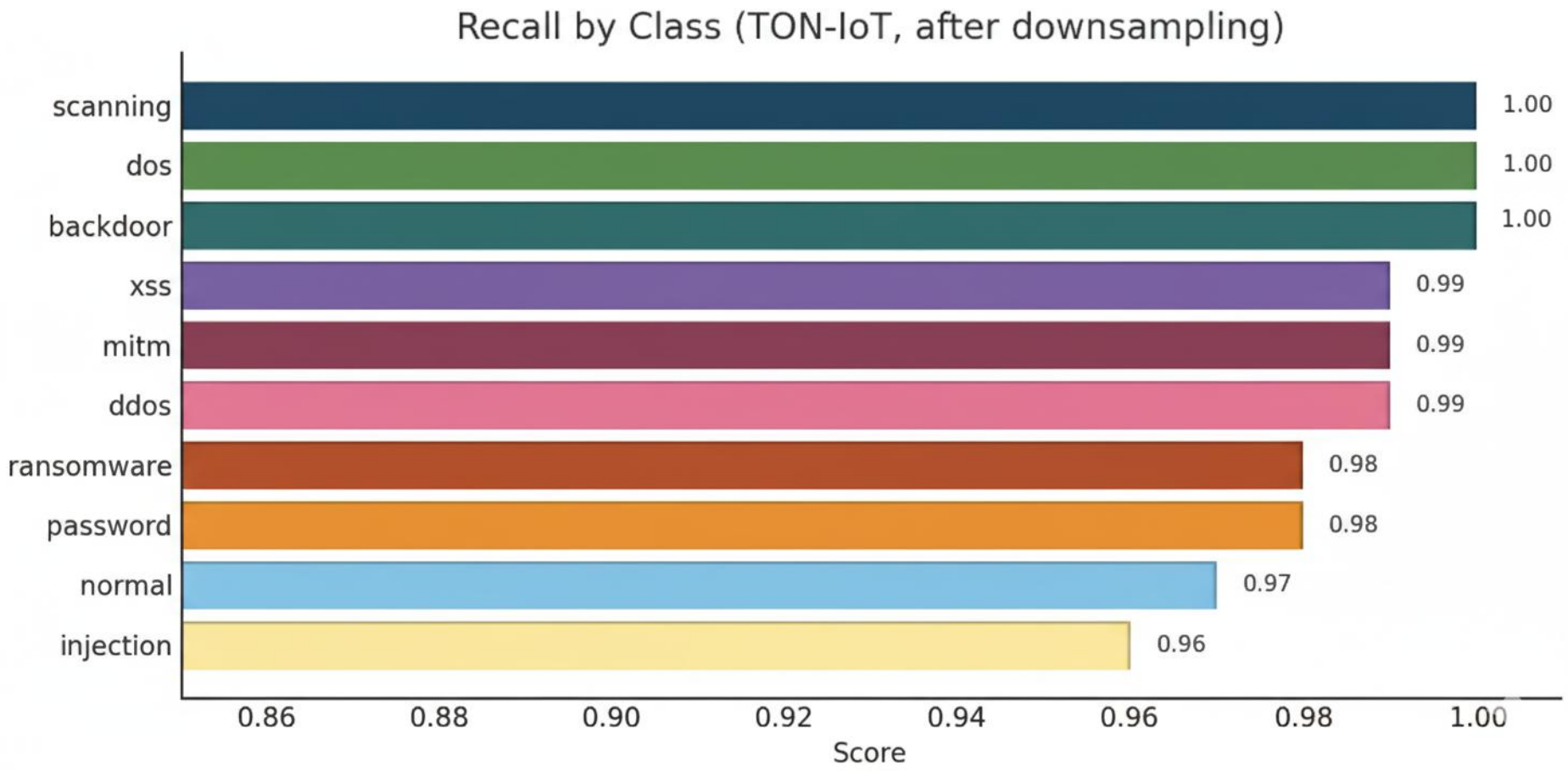
EXPERIMENTAL RESULTS

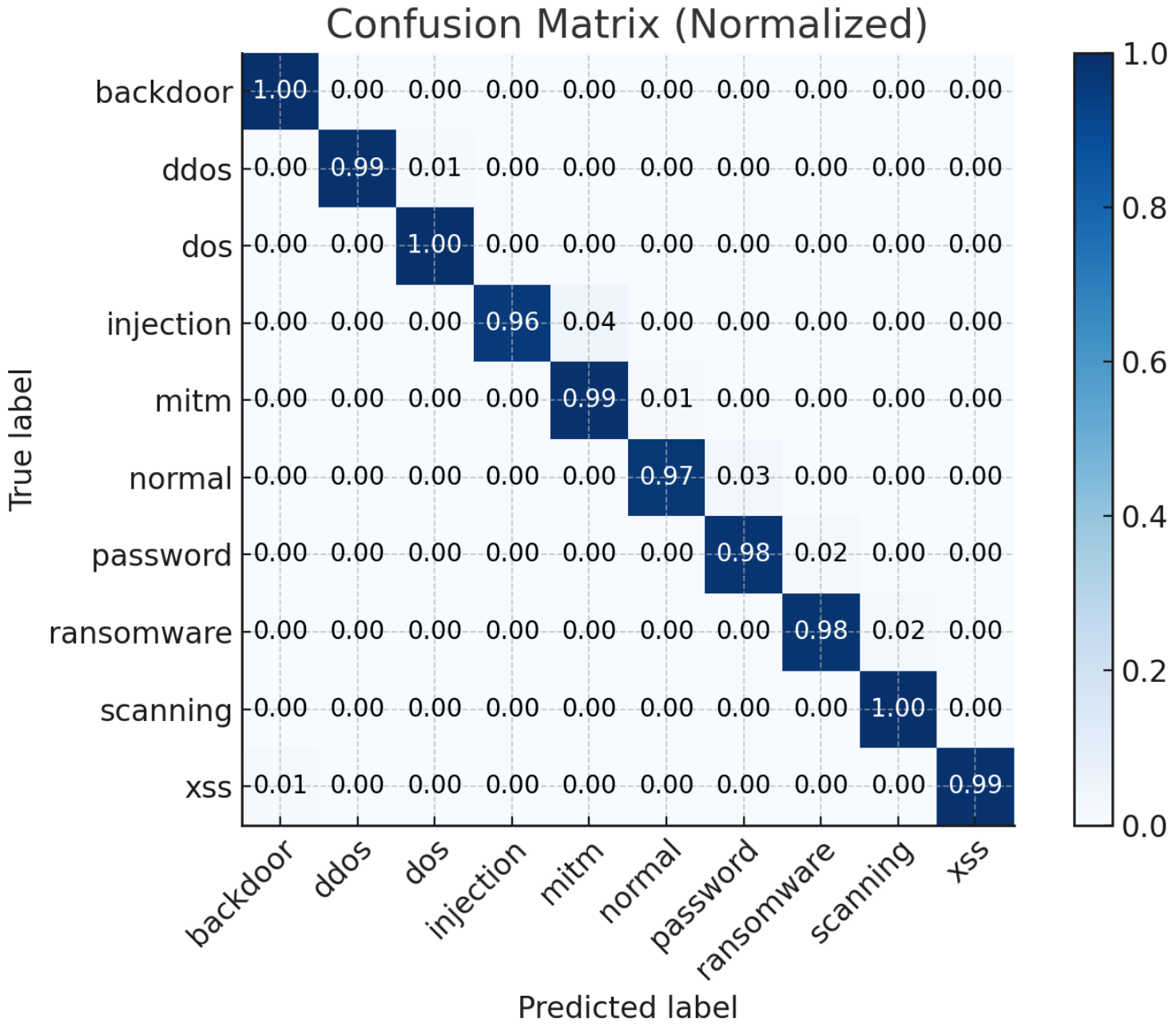


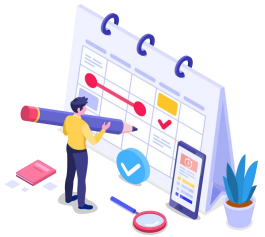


EXPERIMENTAL RESULTS



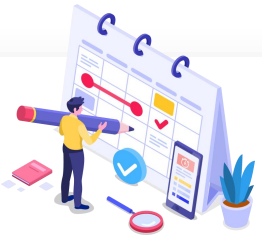




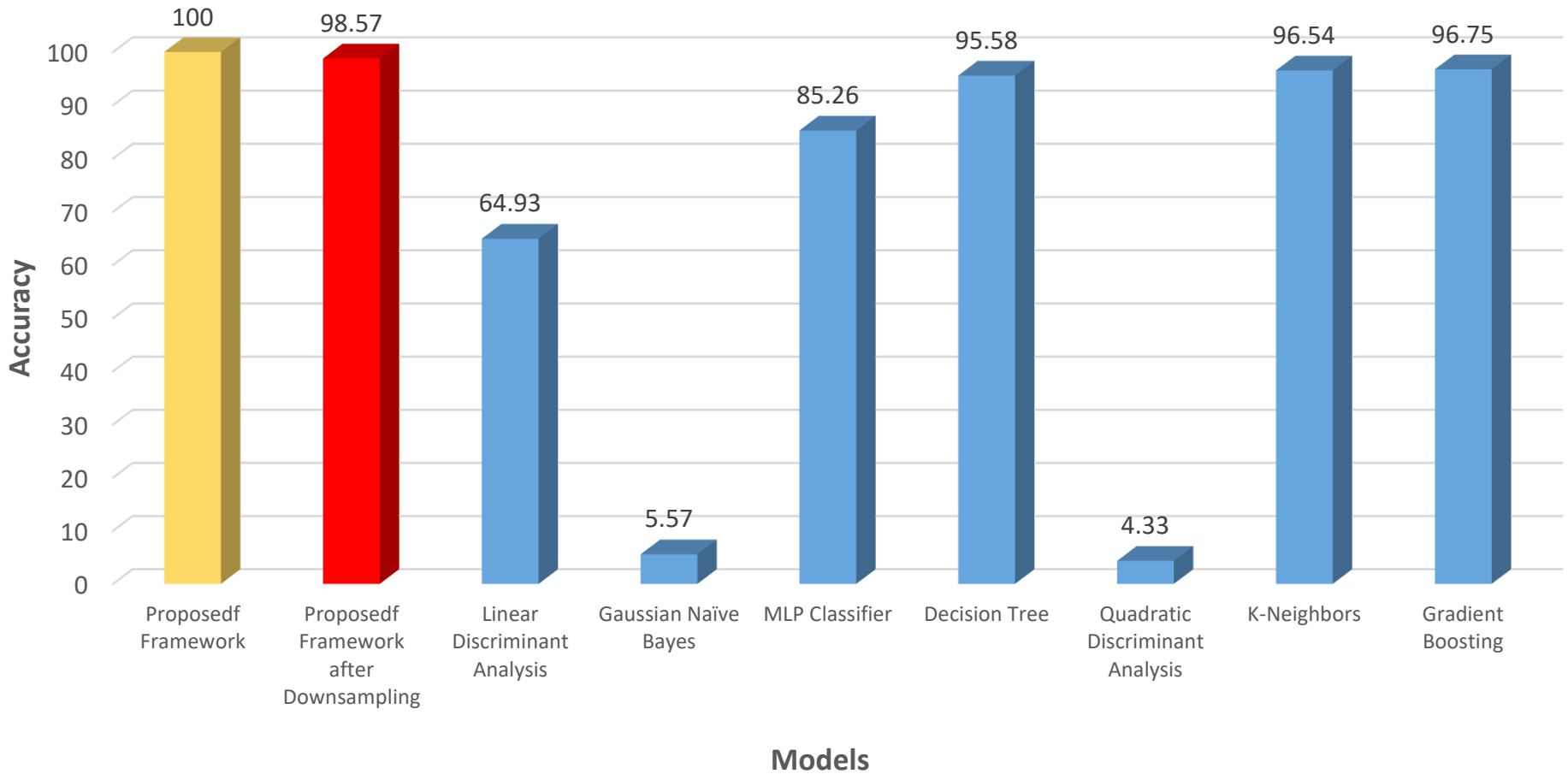


EXPERIMENTAL RESULTS

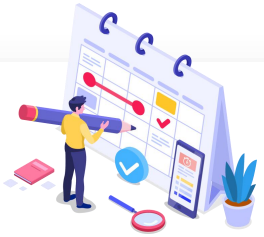
EXPERIMENTAL RESULTS



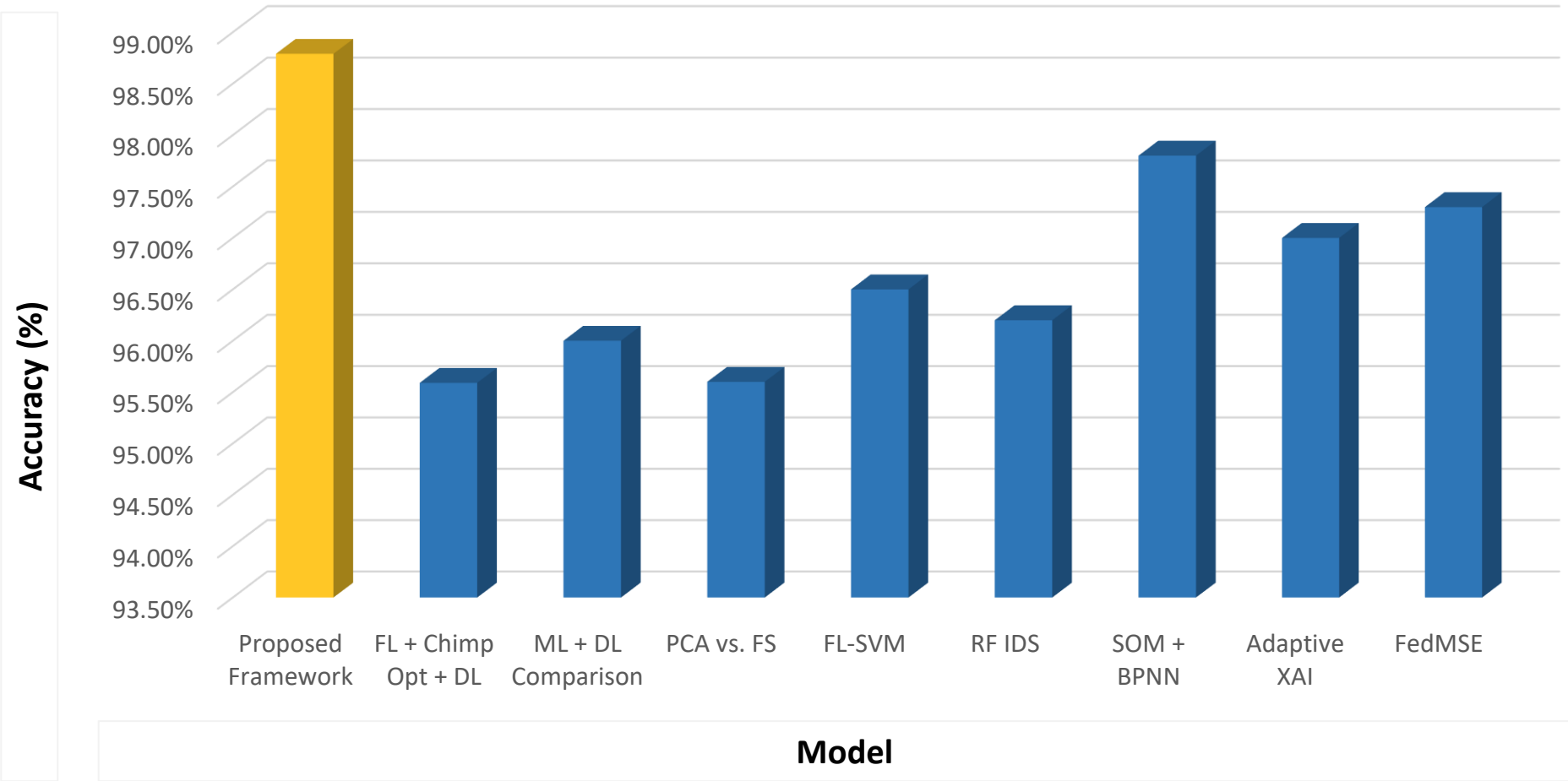
Proposed Framework vs. Other Classifiers



EXPERIMENTAL RESULTS



Model Accuracy Comparison with Prior Research



Conclusions, Recommendations, and Future Work



Conclusion

The framework applies correlation analysis, SelectKBest, and tuned Random Forest.

Results outperformed models like Gradient Boosting and Decision Trees.



A proposed IoT intrusion detection framework achieved 100% accuracy, and 98.57% after balancing.

It also demonstrated scalability potential and highlighted the importance of lightweight, real-time solutions for practical IoT deployment.



- Ensure easy setup and integration with existing IoT systems.
- Real-time detection and instant alerts for suspicious activity.
- Regular model updates to handle evolving attacks.
- Use lightweight, fast, and energy-efficient models for resource-limited devices.
- Comply with data protection standards (e.g., GDPR).
- Foster collaboration between academia and industry for unified standards.
- Simple integration with current monitoring platforms.

Future Work

- Conduct real-world trials in smart cities, hospitals, and factories to test scalability.
- Address energy and resource limitations of IoT devices.
- Incorporate adaptive learning for continuous model evolution.
- Strengthen privacy and security with global unified standards.
- Reduce network latency and improve cloud performance.
- Explore solutions for protocol standardization across diverse IoT devices.



Outcome of The Thesis

G. Dahy, A. Alamri, and H. N. Elmahdy, “IoT Intrusion Detection Framework Using SelectKBest and Hyperparameter–Tuned Random Forest,” in Proceedings of the 9th World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4 2025), London, UK, Aug. 2025.



CERTIFICATE

This is to certify that

Ghada Dahy, Abdullah Alamri, Hesham N. Elmahdy

has contributed a paper titled

IoT Intrusion Detection Framwork Using SelectKBest and Hyperparameter-Tuned Random Forest

in the

9th WORLD CONFERENCE ON

SMART TRENDS IN SYSTEMS, SECURITY AND SUSTAINABILITY (WORLDS4 2025)

London, UK

The paper has also been selected for publication in the (WORLDS4) Conference as per fulfilment of guidelines issued by Springer.

SIMON JAMES FONG
International Conference Chair

9th EDITION WORLDS4 2025

AMIT JOSHI
International Conference Chair

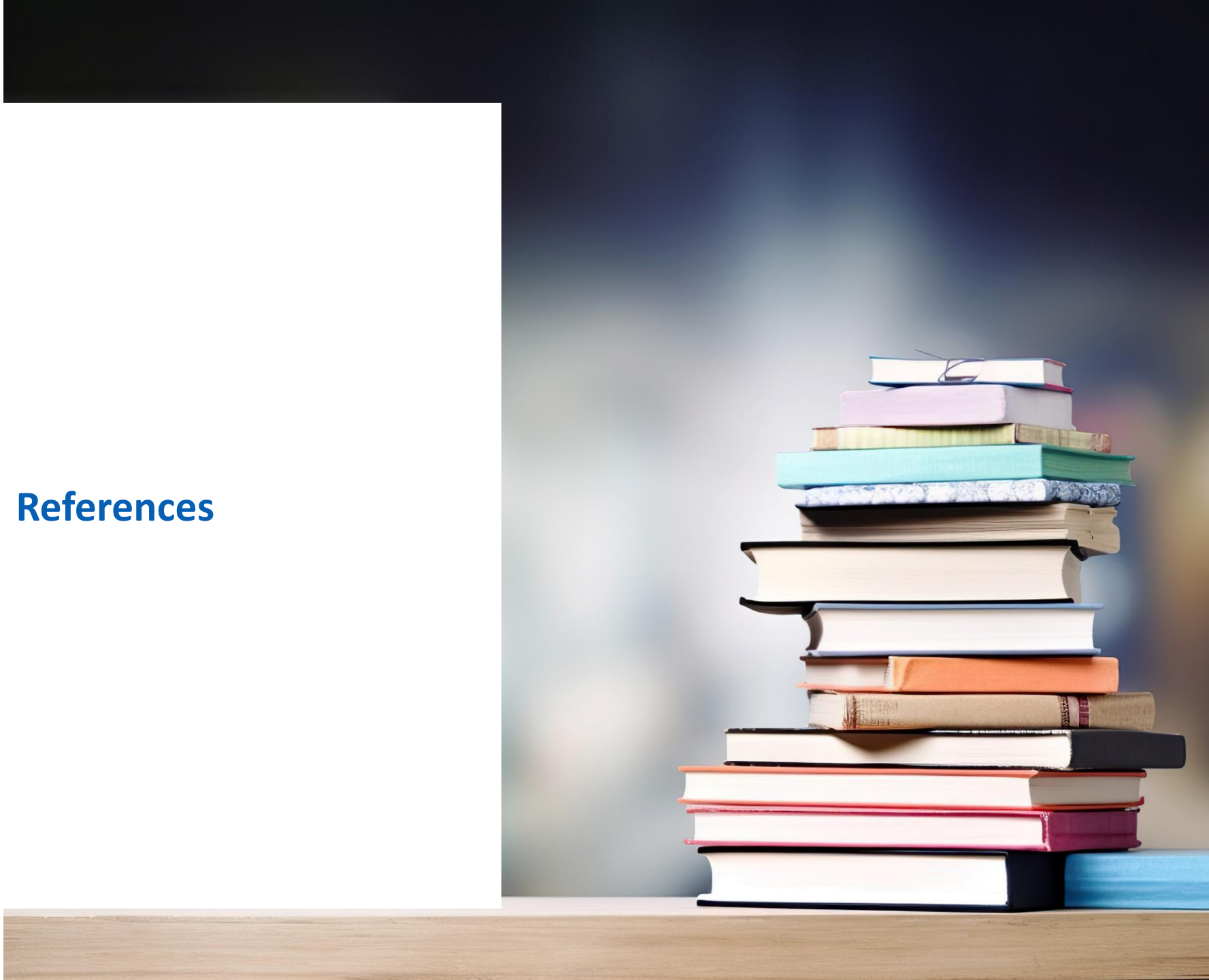
NILANJAN DEY
TPC Chair

19-21 AUGUST, 2025
DIGITAL PLATFORM : ZOOM

DHARM SINGH JAT
TPC Chair

Supporting Local Partner | International Organizing Partner | International Managing Partner | Publication Partner

References



References

1. [1] Karras, Aristeidis, Anastasios Giannaros, Christos Karras, Leonidas Theodorakopoulos, Charalampos S. Mammassis, Georgios A. Krimpas and Sotiris Sioutas, "TinyML Algorithms for Big Data Management in Large-Scale IoT Systems," Future Internet, vol.16, no. 2, pp. 42, 2024.
2. [2] Pandey, Shraiyash and Bharat Bhushan. "Recent Lightweight Cryptography (LWC) Based Security Advances for Resource-Constrained IoT Networks," Wireless Networks, vol.30, no. 4, pp. 2988, (2024).
3. [3] Vlajic, Natalija and Daiwei Zhou, "IoT as a Land of Opportunity for DDoS Hackers," Computer, vol. 51, no. 7, pp.: 26–34, (2018).
4. [4] Narkedimilli, S., Makam, S., Sriram, A. V., Mallellu, S. P., Sathvik, M. S. V. P. J., & Venkatesha Prasad, R. "Enhancing IoT Network Security through Adaptive Curriculum Learning and XAI," arXiv preprint arXiv:2501.11618, (2025).
5. [5] Li, J., M. S. Othman, H. Chen, and L. M. Yusuf. "Optimizing IoT Intrusion Detection System: Feature Selection versus Feature Extraction in Machine Learning," Journal of Big Data, vol. 11, article no. 36, (2024).
6. [6] Jamshidi, S., Nafi, K. W., Nikanjam, A., and Khomh, F. "Evaluating Machine Learning-Driven Intrusion Detection Systems in IoT: Performance and Energy Consumption," arXiv preprint arXiv:2504.09634, (2025).
7. [7] AlHamouz, Z., & Abu-Shareha, M. "A Hybrid Intrusion Detection System Combining Self-Organizing Map and Backpropagation Neural Network for IoT Networks," Journal of Network and Computer Applications, vol. 224, p. 103964, (2024).
8. [8] Narayan, S., Kumar, R., and Sharma, P. "IIDS: Design of Intelligent Intrusion Detection System for Internet-of-Things Applications," arXiv preprint arXiv:2308.00943, (2023).
9. [9] Devine, M., Ardakani, S. P., Al-Khafajiy, M., & James, Y. "Federated Machine Learning to Enable Intrusion Detection Systems in IoT Networks," Electronics, vol. 14, no. 6, p. 1176, (2025).
10. [10] Nguyen, V. T., and Beuran, R. "FedMSE: Semi-supervised Federated Learning Approach for IoT Network Intrusion Detection," Computers & Security, vol. 151, article no. 104337, (2025).
11. [11] Karunamurthy, A., Vijayan, K., Kshirsagar, P. R., and Tan, K. T. "An Optimal Federated Learning-Based Intrusion Detection for IoT Environment," Scientific Reports, vol. 15, article no. 104337, (2025).
12. [12] UNSW Canberra Cyber. "ToN_IoT Datasets." UNSW Research. Accessed April 21, 2025. https://research.unsw.edu.au/projects/toniot-datasets

Outcome of The Thesis

We further applied the proposed framework on another benchmark dataset (RT-IoT2022) and developed a separate research paper based on the obtained results. The paper is currently undergoing final arrangements for publication.



Optimized Machine Learning Framework for Intrusion Detection in IoT Networks Using the RT-IoT2022 Dataset

Ghada Dahy, Abdullah Alamri, and Hesham N. Elmahdy

Department of Information Technology, Faculty of Computers and Artificial Intelligence, Cairo University, Egypt

***Abstract*— Soaring out of hand, the Internet of Things (IoT) has led to new levels of connectivity accompanied by a massive cyber-security threat. The nature of the IoT traffic; dynamic, heterogeneous and high volume can be particularly difficult to deal with using Traditional Intrusion Detection Systems (IDS). This paper presents an IDS structure propped on machine learning that will be used to find and categorize malicious activities in Internet of Things (IoT) setups utilizing an openly accessible RT-IoT2022 data. The methodology is combining extensive data preprocessing, exploratory data analysis (EDA), feature scaling, and usage of several supervised learning algorithms, including K-Nearest Neighbors (KNN), Decision Tree, and Random Forest. Model configurations optimization was done through hyperparameter tuning using GridSearchCV. The results of experiments show that the optimized Random Forest classifier had an optimal performance, with the almost perfect accuracy and F1-scores in various classes of attacks. The paper sheds light on the usefulness of ensemble techniques and hyperparametrization to the precision, the scalability and the generalizability of IDS. The conclusions can also be used to formulate the light but effective IDS systems appropriate in the IoT implementations in the real world.**

Keywords—RT-IoT2022, Machine Learning, Internet of Things (IoT), Random Forest, Decision Tree, K-Nearest Neighbors, Hyperparameter Optimization, Cybersecurity, Intrusion Detection System (IDS),Network Traffic Classification.

I. Introduction

The modern age of global integration relies heavily on the super fast growth of the Internet of Things (IoT) that practically transformed the modern way of life by incorporating smart technologies into our homes, workplaces, industry, and cities [1]-[3]. Though such integration has helped in the rise of notable possibilities regarding convenience, automation, and efficiency in operations, it has also come along with plenty of security concerns [4] to [6]. Most of the IoT gadgets are created with less security in mind, thus making them prone to abuse and putting data confidentiality, system integrity, and the overall net… security at risk [7]--[9]. These concerns are ev… by the heterogeneity and size of IoT, th… to roll out highly sophisticate… (IDS) that would be … environments …

…namic IoT landscapes.

…eprocessing and Exploratory Data Analysis (EDA)

In a bid to create the optimal intrusion detection system on IoT networks, a precise data preprocessing and exploratory data analysis (EDA) workflow was applied within the RT-IoT2022 dataset. The data was retrieved through UCI

spread of malware, and individual phishing exploitation. These risks reveal the insufficiency and lack of effectiveness of traditional IDS solutions, which are sadly, poorly suited towards addressing the shifting topologies and changing threat vectors that are observed in IoT networks [17].

To solve these problems, researchers developing IDS have resorted to machine learning (ML) and deep learning (DL) to compute intelligent, adaptive ones capable of automatically detecting known and new attack patterns [18]-[20]. As an example, [21] demonstrates that the DL architecture can be used effectively in detecting large-scale intrusions of the network type, and [22] presents a summary of DL-based IDSs designed to target IoT, suggesting that long-standing challenges of this framework remain that include a call to scalable models, imbalance in DL data, and detecting IoT intrusions in real-time.

The necessity of increasing the IDS of IoT networks is pointed out by significant and dangerous security incidents throughout the past few years [23]. Considering there are billions of devices currently running worldwide, the impact of breaches may range as far as theft of personal data and significant nationwide infrastructure malfunctions [24]. The infamous 2016 Mirai botnet that was able to take key internet services down at the cost of vulnerability in connected devices is an example of the risks involved in the lack of IoT security [25]. As more and more important sectors, like healthcare, energy, and transportation are affected by IoT, the necessity of intrusion detection mechanisms that can be deemed as resilient and scalable emerges as more imperative [26].

Contemporary IDS strategies increasingly inte… and DL algorithms trained on network … between benign activity and … Breakthroughs … paradi…

Thank you..