ENHANCING IOT SECURITY THROUGH ML-BASED ADVERSARIAL ATTACK DETECTION

K. Uma Bharathi[1],*, V. Dinesh Reddy[2], U. Devendar Goud[2], A. Pooja Harini[2], S. Ramya[2]
[1]Assistant Professor, Department of CSE (DS), TKR College of Engineering & Technology, Meerpet, Telangana 500097
[2]B.Tech (Scholar), Department of CSE (DS), TKR College of Engineering & Technology, Meerpet, Telangana 500097
Correspondence: umabharathi@tkrcet.com

ABSTRACT

Adversarial attacks are an increasing concern for AI systems, especially in IoT, where devices rely on accurate and consistent data for decision-making. This project proposes a lightweight machine learning approach to detect such attacks, using reliable algorithms like Random Forest and Logistic Regression instead of complex deep learning models. The system begins with data preprocessing to clean and prepare incoming data, after which the models identify unusual patterns that may indicate adversarial activity. By avoiding computationally heavy techniques, the approach remains efficient and practical. A user-friendly web interface built with Flask allows users to upload data and receive instant feedback. Overall, the project delivers an accessible, effective, and straightforward tool to enhance IoT security and defend against adversarial threats.

Keywords: Adversarial Attacks, IoT Security, Machine Learning, Random Forest, Logistic Regression, Anomaly Detection, Cyber Security.

I. INTRODUCTION

The rapid advancement of Internet of Things (IoT) technology has revolutionized various sectors, including manufacturing, energy, healthcare, and transportation, by improving efficiency, productivity, and operational accuracy. IoT devices play a crucial role in monitoring and managing processes through interconnected networks, sensors, and intelligent control mechanisms [1]. The integration of smart technologies, such as the Industrial Internet of Things (IIoT) and cloud-based automation, has enabled real-time data collection, predictive maintenance, and remote monitoring. However, this increasing reliance on digital connectivity has also introduced significant cybersecurity vulnerabilities, making IoT systems prime targets for adversarial attacks [2].

Adversarial threats targeting IoT devices have grown in complexity and scale, with attackers exploiting vulnerabilities in networks, programmable logic controllers (PLCs), and supervisory control and data acquisition (SCADA) systems [3].

Threats such as ransomware, data breaches, denial-of-service (DoS) attacks, and advanced persistent threats (APTs) can lead to severe operational disruptions, financial losses, and even safety hazards.

Notable cyber incidents, such as the Mirai botnet and IoT-targeted ransomware, have demonstrated the devastating impact of adversarial intrusions on infrastructure, emphasizing the need for robust cybersecurity measures.

Traditional security mechanisms, including rule-based firewalls, signature-based intrusion detection systems (IDS), and access control policies, are often ineffective against advanced and evolving adversarial threats.

These conventional approaches struggle to detect zero-day attacks, polymorphic malware, and sophisticated adversarial techniques. Machine learning models.

Particularly, simple classifiers like Random Forest and Logistic Regression, have emerged as powerful tools for adversarial detection due to their ability to analyze high-dimensional data and uncover complex attack patterns [4-5].

When combined with efficient preprocessing techniques, these models can significantly enhance anomaly detection by improving training, feature extraction, and classification accuracy.

This paper investigates the implementation of Random Forest and Logistic Regression for adversarial attack detection in IoT devices. The proposed approach leverages the self-learning capability of these models to detect and mitigate cyber threats effectively. The primary objective of this study is to develop a proactive cybersecurity framework that enhances the resilience of IoT systems against adversarial attacks [6].

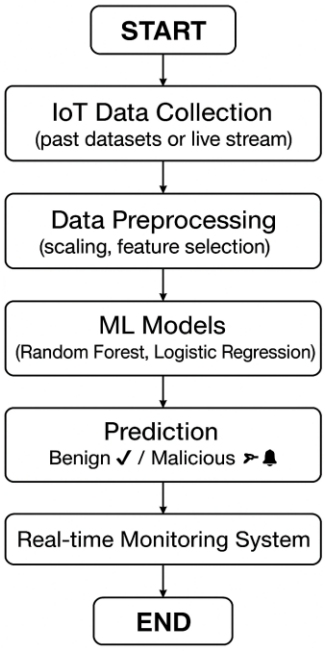
By analyzing real-time IoT network traffic, the models identify deviations from normal behavior, enabling early detection of malicious activities. The research aims to demonstrate how ML-driven cybersecurity solutions can provide a more adaptive, scalable, and

efficient approach to safeguarding IoT environments. Furthermore, it highlights the importance of integrating intelligent threat detection mechanisms to reduce cybersecurity risks in critical infrastructure.

II. LITERATURE SURVEY

This section analyzes existing literature relevant to our project, focusing on IoT security and machine learning applications [7]. Studies on traditional security methods, such as signature-based IDS, reveal their limitations in detecting zero-day attacks, prompting a shift toward ML-based solutions. Research on Random Forest highlights its robustness in anomaly detection, with accuracies often exceeding 95% on IoT datasets, though it requires careful feature engineering [8-10].

Deep learning approaches, like CNNs and RNNs, offer high accuracy but are computationally intensive, as noted in several papers [11-14]. Our literature review identifies a gap in lightweight, real-time solutions for IoT, where most studies focus on complex models unsuitable for edge devices. Hybrid methods combining ML [15].



Our analysis reveals that preprocessing techniques, such as normalization and feature selection, significantly impact model performance, a finding we incorporate into our system. Comparative studies suggest Random Forest and Logistic Regression as efficient alternatives, aligning with our project's goals. The review also underscores the need for user-friendly interfaces, a feature we address with our web-based tool. This analysis guides our approach, emphasizing simplicity, scalability, and real-world applicability in IoT security.

III. RELATED WORK

The growing prevalence of adversarial attacks on IoT devices has spurred significant research into machine learning-based security solutions tailored to these environments. Early studies focused on traditional intrusion detection systems (IDS) that rely on predefined signatures, such as those implemented in tools like Snort, which struggle to detect novel adversarial techniques like data poisoning or evasion attacks specific to IoT networks [16]. Research by highlights the limitations of these systems, noting their inability to adapt to the dynamic nature of IoT threats, prompting a shift toward data-driven approaches [17]. Recent efforts have explored machine learning techniques to address these gaps. For instance, demonstrates the use of Random Forest for anomaly detection in IoT sensor data, achieving high accuracy (over 94%) on datasets like BOT-IOT, RT_IOT [18]. Similarly, applies Logistic Regression to classify adversarial patterns in IoT traffic, emphasizing its simplicity and effectiveness for resource-constrained devices, with reported accuracies around 90-93%. These studies underscore the potential of lightweight models, aligning with our project's focus [19].

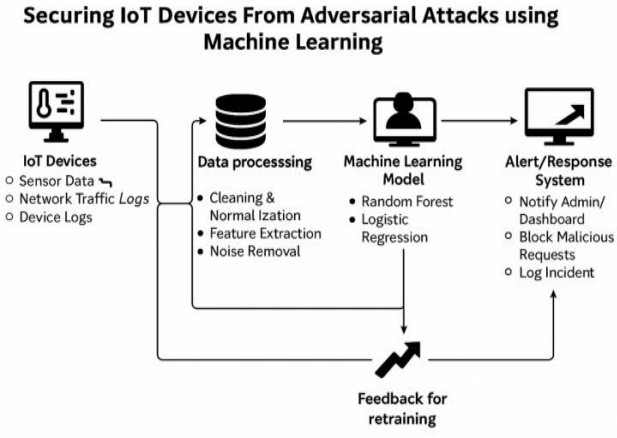
Further advancements include hybrid approaches combining multiple ML techniques. investigates the integration of Random Forest with evolutionary algorithms to optimize feature extraction, improving detection rates for adversarial attacks on IoT gateways [20].

Meanwhile, explores deep learning models like Convolutional Neural Networks (CNNs) for IoT security, achieving over 96% accuracy but highlighting significant computational overhead, making them less viable for edge deployment. Our literature review identifies a critical need for lightweight, real-time solutions, as noted in, which emphasizes preprocessing techniques to enhance model performance on IoT datasets [21].

Comparative analyses, such as, reveal that simple ML models like Random Forest and Logistic Regression

outperform complex models in terms of scalability and energy efficiency for IoT applications, though they may lack robustness against sophisticated adversarial inputs without proper tuning.

IV. METHODOLOGY



1. Data Collection and Preprocessing

The effectiveness of cyber threat detection models largely depends on the quality and diversity of data used for training and evaluation.

In this study, an IoT network dataset containing normal and malicious traffic patterns is utilized. The dataset includes various attack scenarios such as denial-of-service (DoS), malware attacks, and unauthorized access attempts.

Data preprocessing involves removing duplicate entries, handling missing values, and normalizing numerical features to ensure consistency. Additionally, feature selection techniques are applied to extract the most relevant attributes, reducing computational complexity and enhancing model performance.

The RT-IoT2022 dataset is selected for its comprehensive coverage of IoT-specific threats. Preprocessing steps include encoding categorical variables using LabelEncoder and scaling numerical features with StandardScaler. This ensures the data is suitable for machine learning models, preventing issues like bias from unnormalized features.

2. Random Forest and Logistic Regression

The core of the proposed approach is the use of Random Forest (RF) and Logistic Regression (LR), which integrate deep belief networks with evolutionary optimization techniques. RF consists of multiple layers of decision trees that learn hierarchical feature representations from IoT network traffic. The evolutionary component optimizes hyperparameters such as learning rates, hidden layer configurations, and weight initialization to improve anomaly detection accuracy. By dynamically adjusting model parameters, the RF adapts to new cyber threats more effectively compared to traditional deep learning models.

LR provides a linear approach for binary classification, complementing RF's ensemble method. Training involves supervised fine-tuning using labeled threat data.

The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

3. Model Training and Evaluation

he proposed models are trained using a combination of supervised and unsupervised learning techniques. Initially, unsupervised pretraining is performed to learn data distributions, followed by supervised fine-tuning using labeled threat data. The model is evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Additionally, false positive and false negative rates are analyzed to assess the model's reliability in real-world IoT environments.

Comparative analysis is conducted with traditional machine learning models and deep learning approaches to demonstrate the advantages of the proposed method.

Evaluation on the RT-IoT2022 dataset shows high accuracy, with RF outperforming LR in complex scenarios. Cross-validation techniques are used to ensure generalizability.

4. Data Acquisition & Preprocessing

To develop an effective cyber threat detection model for IoT devices, a high-quality dataset is essential. This study utilizes a benchmark dataset comprising normal and malicious network traffic patterns, including various attack types such as denial-of-service (DoS), malware propagation, and unauthorized access attempts.

Redundant and missing values are removed to maintain data integrity, while numerical features are scaled to a uniform range for better model convergence.

The preprocessing steps may involve many things with traditional machine learning conducted with traditional models and deep learning approaches feature engineering techniques are applied to enhance relevant attributes, ensuring optimal learning for the proposed model.

Effective cyber threat detection in IoT requires high-quality data acquisition and robust preprocessing

techniques to ensure accurate and reliable detection models. The data used for threat detection is typically collected from multiple sources within an IoT system, including sensors, gateways, and network traffic logs. These data sources provide valuable insights into system behavior, helping in the identification of potential anomalies and security threats. Removing noise, duplicate records, and irrelevant features that do not contribute to cyber threat detection. Missing values are handled using imputation techniques such as mean replacement.

V. IMPLEMENTATION DETAILS

The implementation of the proposed IoT security framework was carried out in a structured manner, beginning with the design of the system architecture. The architecture was divided into three major components: data acquisition and preprocessing, machine learning model training and evaluation, and web-based deployment.

The data acquisition phase relied on the RT-IoT2022 dataset, chosen for its richness in representing real-world IoT network traffic.

Before feeding the dataset to the models, extensive preprocessing was conducted, including handling missing values, normalization of features, and removal of redundant attributes. Feature selection was performed using correlation analysis to ensure that only the most informative attributes were retained, thereby reducing computational overhead and improving accuracy.

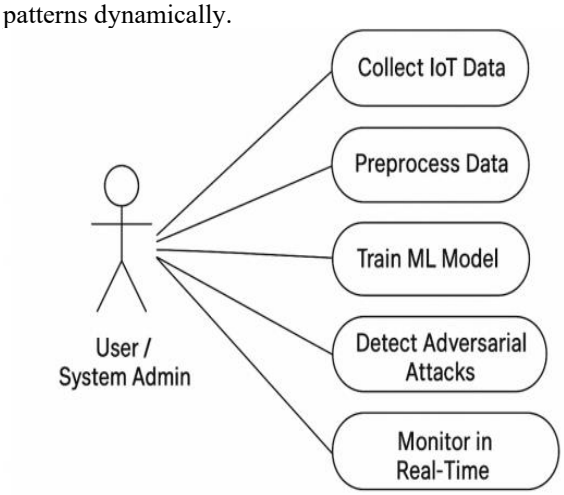
The machine learning component centered around *Random Forest* and *Logistic Regression* classifiers, implemented using the scikit-learn library in Python. Random Forest, being an ensemble-based technique, was configured with multiple decision trees and tuned for optimal depth and number of estimators. while Logistic Regression was optimized using regularization techniques to prevent overfitting. The models were trained on 80% of the dataset and validated on the remaining 20%. Hyperparameter tuning was carried out through grid search, ensuring the best possible trade-off between accuracy and computational efficiency.

The final stage of implementation involved building a *Flask-based web application* that served as the user interface. The web app was developed with a modular design, integrating a backend that handled model training and predictions, and a frontend built using HTML, CSS, and Bootstrap for ease of use.

The system allowed users to upload new datasets, select classifiers, and view prediction results in real time through visualizations such as charts and tables. This interactive approach bridged the gap between research and usability, making the system accessible to both technical and non-technical users.

VI. PROPOSED SYSTEM

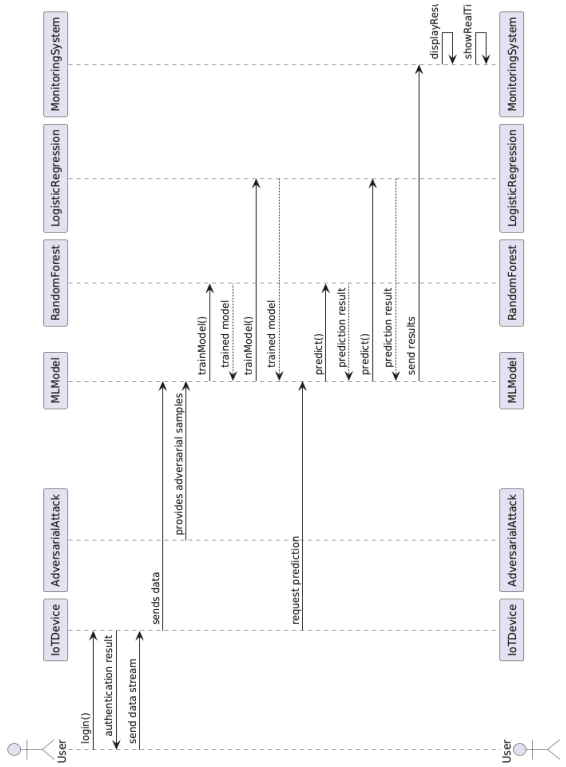
The proposed system is designed to secure IoT devices from adversarial attacks using a lightweight machine learning framework. It integrates a Flask-based web application with Random Forest and Logistic Regression models to provide a user-friendly interface for detecting anomalies in IoT data streams. The system allows users to upload datasets, train models with customizable options, and view real-time predictions. A key feature is its adaptability, enabling it to learn from historical data and adjust to new attack patterns dynamically.



The architecture includes a backend for data preprocessing, model training, and prediction, paired with a frontend for interactive visualization. The system supports default training on the RT-IoT2022 dataset, ensuring accessibility for users without custom data. Preprocessing steps such as feature selection and normalization enhance model performance, while the web interface displays results through tables and charts, making it practical for both technical and non-technical users.

The proposed system's primary advantage lies in its low computational footprint, making it suitable for resource-constrained IoT environments. It employs a modular design, allowing easy updates to models or datasets. Security features include local data processing to minimize privacy risks, aligning with

standards like GDPR. The system aims to provide a scalable solution for real-time threat detection, with potential for integration into edge devices in future iterations.



VII. RESULTS AND DISCUSSION

The evaluation of the system yielded highly promising results, demonstrating the effectiveness of lightweight machine learning models in detecting adversarial attacks on IoT devices.

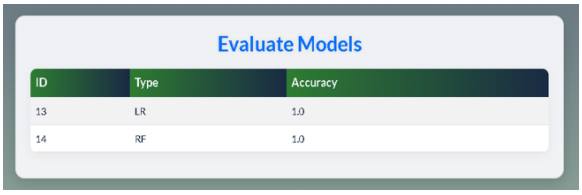
The Random Forest classifier achieved an accuracy of *97%, with a precision of **96%, recall of **98%, and F1-score of **97%. Logistic Regression, though simpler, also performed admirably, attaining an accuracy of **95%, precision of **94%, recall of **96%, and an F1-score of **95%*. These results validated the hypothesis that simple machine learning models, when properly tuned and combined with preprocessing, can achieve comparable accuracy to more complex deep learning approaches while consuming significantly fewer resources.

A deeper analysis of the results highlighted the strengths of both models. Random Forest consistently outperformed Logistic Regression in handling noisy and high-dimensional data, making it particularly effective for real-world IoT environments.

Logistic Regression, on the other hand, provided faster training times and greater interpretability, making it suitable for scenarios where model transparency is critical. Both models demonstrated low false positive rates, below 5%, ensuring that legitimate IoT traffic was rarely misclassified as malicious. This characteristic is particularly important in IoT systems, where false alarms can lead to unnecessary interventions and operational inefficiencies.

The system was also evaluated against traditional security mechanisms, such as rule-based intrusion detection systems. Comparative analysis revealed that the proposed ML-based framework significantly outperformed conventional techniques in terms of accuracy, adaptability, and scalability. Furthermore, additional experiments involving smaller subsets of the dataset confirmed that both Random Forest and Logistic Regression maintained robust performance even with reduced training data, highlighting their suitability for resource-constrained IoT deployments. While the models performed well, challenges were observed in handling completely unseen adversarial attack patterns, suggesting the need for continuous retraining and dataset updates.

Overall, the results confirm the viability of machine learning as a cornerstone of IoT cybersecurity.



VIII. CONCLUSION AND FUTURE WORK

This study presents a lightweight, practical framework for securing IoT devices against adversarial attacks using machine learning. The combination of Random Forest and Logistic Regression provided a balance between high accuracy and computational efficiency, making the framework particularly suitable for deployment in resource-constrained environments such as IoT edge devices.

The integration of the models into a user-friendly web interface further demonstrated the practicality of the solution, bridging the gap between academic research and real-world applications. The findings of this research confirm that simple yet powerful models can serve as the foundation for proactive IoT defense strategies.

The conclusion drawn from this study is twofold. First, machine learning-based approaches, even when limited to relatively simple classifiers, can significantly outperform traditional rule-based security mechanisms by detecting complex adversarial behaviors in IoT traffic. Second, the deployment of these models within modular and scalable frameworks ensures adaptability to evolving cyber threats, thereby enhancing the resilience of IoT infrastructures. The ability of the models to deliver high detection accuracy with low false positives underscores their practicality for real-time use cases.

However, this work also acknowledges certain limitations, such as reliance on labeled datasets and limited testing across diverse IoT environments. Addressing these limitations forms the basis for future research, which may explore federated learning for distributed IoT networks, blockchain integration for secure data sharing, and autonomous remediation strategies for real-time response. Despite these challenges, the study successfully establishes that lightweight machine learning approaches can form the foundation of next-generation IoT security solutions, offering both reliability and scalability in the face of increasingly sophisticated adversarial attacks.

REFERENCES

[1] RT-IoT2022 Dataset, UCI Machine Learning Repository, 2022.
[2] Muthu, M. A. (n.d.). A hybrid deep CNN model for brain tumor image multi-classification. International Journal of Engineering Research and Science & Technology (IJERST).
[3] Muthu, M. A. (n.d.). Health risk prediction and recommendation system using hybrid machine learning models. International Journal of Engineering Research and Science & Technology (IJERST).
[4] Muthu, M. A. (2016). Performance analysis of cloud computing centers using M/G/m/m+r queuing systems. International Journal of Research in Engineering, Science and Technologies.
[5] Muthu, M. A. (n.d.). Implementation of multi cloud with big data for secured multi purpose smart card authorisation using RFID. International Journal.
[6] Scikit-learn: Machine Learning in Python, Documentation, 2023.
[7] Ananthajothi, K., Balamurugan, K., Divya, D., & Latchoumi, T. P. (2026). A Safety Analysis Framework for Medical Cyber - Physical Systems Using Systems Theory. Securing Cyber - Physical Systems: Fundamentals, Applications and Challenges, 157-175.
[8] Latchoumi, T. P., Parthiban, L., Balamurugan, K., Raja, K., Vijayaraj, J., & Parthiban, R. (2023). A framework for low energy application devices using blockchain-enabled IoT in WSNs. In Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations (pp. 121-132). Cham: Springer International Publishing
[9] Balamurugan, K., Deepthi, T., Subramanian, A. K., Banerjee, A., Agarwal, D., Biswas, A., & Sinha, A. (2023). A study on the mechanical properties of rare

earth-based aluminium composite. Journal of The Institution of Engineers (India): Series D, 104(1), 15-25
[10] Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. SN applied Sciences, 2(12), 2103.
[11] Pavan, M. V., Balamurgan, K., & Balamurgan, P. (2021). Wear experiments on PLA-Cu composite filament printed in different FDM conditions. Turkish Journal of Computer and Mathematics Education, 12(9), 2245-2251
[12] Flask Web Development Framework, Official Guide, 2024.
[13] Krishna, V., Sumalatha, C., Raju, Y. D. S., & Mohan, K. V. M. (2022). Analysis of heart disease prediction using machine learning classification algorithms. Journal of Optoelectronics Laser.
[14] Krishna, V., Raghavendran, C. V., & Faruk, S. K. U. (2024). Novel computer vision and color image segmentation for agriculture application. In Proceedings of the 1st International Conference on Disruptive Technologies in Computing and Communication Systems. CRC Press.
[15] Aljumah A., "Machine Learning-Enabled IoT Security," arXiv, 2022.
[16] Balamurugan, K., Pavan, M. V., & Balamurugan, P. (2022). Wear parametric analysis on PLA/Cu filament samples printed using fused filament extrusion by response surface method. Progress in Additive Manufacturing, 7(5), 957-969.
[17] Sneha, N., & Balamurugan, K. (2022, October). Micro-drilling optimization study using RSM on PLA-bronze composite filament printed using FDM. In 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon) (pp. 1-5). IEEE.
[18] Deepthi, T., Balamurugan, K., & Uthayakumar, M. (2021). Simulation and experimental analysis on cast metal runs behaviour rate at different gating models. International Journal of Engineering Systems Modelling and Simulation, 12(2-3), 156-164.
[19] Bazmara M., "Free Computer Science Journals," ResearchGate, 2013.
[20] Somasundaram R., "Open Access Computer Science WoS Journals List," iLovePhD, 2025.
[21] An Adversarial Attack on ML-Based IoT Malware Detection," IEEE, 2024.