**An ML-Driven Approach for Detecting Threats in Internet of Things Networks**

**K Madhu Sudhan Reddy1, Bodeddula Siva priya 2**

1 Assistant professor, Dept of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India, Email: mca.madhureddy@gmail.com

2Post Graduate, Dept of MCA, Annamacharya Institute of Technology & Sciences, Tirupati, AP, India, Email: bodeddulasivapriya2@gmail.com

**Abstract:**

The Fourth Industrial Revolution has led to the widespread usage of the Internet of Things (IoT) in several industries, including smart agriculture, emergency response systems, and healthcare. These IoT ecosystems are made up of networked devices that work together to gather, transmit, and analyse data using edge computing, gateway systems, cloud infrastructure, and AI-enhanced analytics. These sophisticated networks do, however, have significant security vulnerabilities, including data breaches, unauthorised access attempts, and distributed denial-of-service (DDoS) assaults. To safeguard these environments, robust detection frameworks utilising machine learning, behavioural analysis, and continuous monitoring are required. This paper examines key security concerns in IoT deployments and evaluates various threat detection methodologies. We investigate network behaviour modelling, traffic pattern analysis, and specialist intrusion detection systems designed for Internet of Things environments. The study evaluates machine learning models using the CIC-IDS-2017 and CIC-IDS-2018 benchmark security datasets. Our findings demonstrate the critical role AI-powered security solutions play in identifying and preventing emerging risks in IoT deployments.

**Keywords:** Security, Threat Detection System, Anomaly Recognition, Intrusion Detection Frameworks, Network Behaviour Analysis

**1. INTRODUCTION**

The technological transformation driven by the Fourth Industrial Revolution has fundamentally altered business operations, industrial processes, and human-technology interaction paradigms. Among the pivotal technologies enabling this transformation, the IoT stands alongside AI, cloud infrastructure, and automated process systems as a foundational element. IoT technologies provide transformative capabilities across healthcare delivery, agricultural operations, urban infrastructure, and crisis management frameworks through simplified data acquisition, ongoing monitoring, and automated responses. IoT ecosystems comprise interconnected smart devices exchanging information through network infrastructure to collect, transmit, and analyze physical world data. Key architectural components include peripheral sensing devices, network gateways, cloud-based processing systems, analytical frameworks, and user engagement interfaces. These systems empower individuals and organizations to leverage real-time insights for improved decision-making processes. Despite these advantages, the exponential growth in connected devices introduces substantial security challenges. Many of the billions of networked IoT devices operate with inadequate security protections, creating extensive attack surfaces. Malicious actors exploit these vulnerabilities to deploy various attack vectors including malware distribution, coordinated service disruption campaigns, and unauthorized data extraction. Beyond compromising sensitive information, IoT security breaches can severely impact both consumer and enterprise applications. The interconnected nature of IoT environments means attackers can target multiple systems simultaneously, potentially causing widespread disruption. Addressing these evolving security challenges requires sophisticated threat detection approaches. Conventional security measures such as network firewalls and signature-based intrusion detection systems provide insufficient protection against the evolving threat landscape affecting IoT environments. Research indicates that advanced detection mechanisms incorporating artificial intelligence and machine learning demonstrate superior capability in identifying patterns and anomalies across IoT network traffic. Security models developed using datasets like CIC-IDS-2017 and CIC-IDS-2018 provide valuable frameworks for identifying malicious activities and preventing security incidents.The continuing expansion of IoT implementations necessitates robust security architectures capable of withstanding sophisticated cyber threats. Security researchers and industry practitioners are exploring innovative approaches including distributed ledger authentication systems, enhanced communication protocols, and intelligence-driven threat assessment frameworks. This paper provides comprehensive analysis of IoT security vulnerabilities, detection methodologies, and emerging protection strategies. The goal of this project is to aid in the creation of more intelligent and robust security solutions for IoT ecosystems by investigating real-world use cases and machine learning security frameworks.

**2. METHODS AND MATERIALS**

**A. Dataset**



**Figure 1. Dataset for CIC-IDS-2017**



**Figure 2. Dataset for CIC-IDS-2018**

This research utilizes the publicly accessible CIC-IDS-2017 and CIC-IDS-2018 datasets, which serve as benchmarks for evaluating intrusion detection systems within IoT environments. These comprehensive datasets contain authentic network traffic captures representing various security scenarios, including distributed service disruption attempts, credential compromise attempts, botnet activities, and reconnaissance operations through port scanning.The Canadian Institute for Cybersecurity developed the CIC-IDS-2017 collection, which provides categorized traffic captures distinguishing between normal network operations and various attack patterns. The CIC-IDS-2018 dataset expands upon this foundation by incorporating additional attack methodologies, enhancing its suitability for developing and evaluating machine learning security models.These datasets provide extensive network flow characteristics including packet dimensions, connection duration, endpoint identification, protocol information, and irregularity indicators—all essential parameters for developing effective security solutions.

**B. Data Preprocessing**

The raw datasets used for IoT threat detection typically contain inconsistencies, missing elements, and disproportionate class distributions that can significantly impact machine learning model effectiveness. Therefore, systematic data preparation is essential to transform raw information into structured formats suitable for analytical processing. The initial preprocessing phase focuses on data cleansing to eliminate duplicate records and extraneous network flows, ensuring dataset consistency. Missing values are either removed when non-critical or completed through appropriate interpolation techniques. Since IoT traffic captures contain both normal and malicious activities, eliminating noise and redundant attributes that don't contribute to detection accuracy is crucial. Following data cleansing, feature selection identifies the most relevant network attributes for model training. IoT traffic data encompasses numerous parameters including packet dimensions, flow duration, protocol information, and destination ports. Since not all attributes contribute equally to detection performance, selecting the most significant features reduces computational requirements while enhancing model accuracy. Recursive feature removal, correlation analysis, and information gain evaluation are statistical techniques that assist in identifying critical features while eliminating superfluous or irrelevant components. After feature selection, numerical data undergoes normalization and scaling to ensure each attribute contributes proportionally to the learning process. Network parameters such as packet size and duration often exhibit widely varying ranges, potentially creating learning bias. Standardization techniques including Z-score normalization and range scaling bring numerical attributes within comparable distributions, preventing models from overemphasizing features with larger magnitude values. Categorical elements such as protocol types and service identifiers require conversion to numerical representations before processing by machine learning algorithms. Encoding approaches such as one-hot encoding or label transformation address this requirement. One-hot encoding proves particularly effective for non-ordinal categorical values by representing each category as a binary vector, ensuring algorithms correctly interpret non-numeric information when analyzing protocol behaviors. Addressing class imbalance is critical since real-world IoT security datasets typically contain significantly fewer attack instances compared to normal traffic. This disproportion can cause algorithms to favor the predominant class. Attack sample representation is artificially increased by the Synthetic Minority Oversampling Technique (SMOTE) and related techniques, whilst undersampling techniques can decrease typical traffic samples to produce more balanced training sets. These techniques enhance model performance in detecting security threats even when they represent a small percentage of overall traffic. Finally, the preprocessed dataset undergoes division into training and evaluation segments, typically allocating 70-80% for model development and 20-30% for performance assessment. This approach ensures models learn effectively from historical data while being evaluated against previously unseen network traffic. The accuracy, generalisability, and efficacy of machine learning security frameworks in Internet of Things contexts are greatly improved by appropriate preprocessing.

**C. System Implementation**

The IoT threat detection system follows a phased implementation approach, combining continuous network monitoring capabilities with machine learning security models. The system architecture facilitates malicious activity identification through comprehensive traffic analysis and anomaly detection methodologies. Implementation encompasses data acquisition, model development, real-time monitoring, and threat classification processes to ensure effective and scalable security coverage.The implementation begins with data collection from IoT network environments, capturing detailed packet information including source and destination addressing, protocol specifications, packet dimensions, temporal data, and connection duration. The CIC-IDS-2017 and CIC-IDS-2018 datasets provide labeled traffic captures essential for training and validating threat detection models. Collected data undergoes preprocessing including cleansing, normalization, and encoding before entering the machine learning pipeline.Following data preparation, appropriate machine learning algorithms are selected and trained. Various classification approaches including ensemble methods (Random Forest), support vector techniques, decision tree algorithms, and neural network architectures undergo evaluation to determine optimal threat detection performance. Feature extraction methods identify critical attributes for classification, enhancing model accuracy. The prepared dataset undergoes division into training and testing segments, with hyperparameter optimization performed to maximize detection effectiveness.After training completion, the system deploys into production environments for continuous IoT network monitoring. An intrusion detection module captures incoming and outgoing communications from connected devices, with the trained model processing this information in real-time to identify patterns and classify network activities. When potential threats are detected, the system generates security alerts enabling immediate response from security teams. Visualization interfaces provide comprehensive threat information and activity logs for administrative oversight.To maintain effectiveness against evolving threats, the system incorporates periodic model updates with emerging attack patterns and retraining processes using current datasets. Threat intelligence feeds and automated logging mechanisms enhance adaptability to new security challenges. Additionally, the implementation includes secure communication protocols and encryption systems to protect the integrity of the detection framework itself. A thorough performance review comparing actual security incidents to F1-score metrics, recall, accuracy, and precision is part of the final deployment. These measurements guide continuous refinement to minimize false positives while maintaining high detection rates. The deployed solution provides an efficient, scalable, and intelligence-driven approach to securing IoT environments against network-based attacks and unauthorized access attempts.

**3. RESULTS AND ANALYSIS:**

**Figure 3. Confusion Matrix**

The confusion matrix provides visual representation of classification performance by comparing actual classifications against predicted outcomes. The matrix comprises four essential measurements:

* **True Positives (TP):** Successfully identified attack instances
* **True Negatives (TN):** Correctly recognized normal traffic patterns
* **False Positives (FP):** Normal activities incorrectly flagged as attacks
* **False Negatives (FN):** Attack activities misclassified as normal traffic

Higher TP and TN values indicate superior model performance, while elevated FP and FN measurements suggest classification challenges requiring additional optimization.



**Figure 4. ROC Curve & AUC Score**

Using the Receiver Operating Characteristic (ROC) curve, performance trade-offs between true positive rates and false positive rates are evaluated across different threshold settings. Superior discrimination between benign and malevolent traffic patterns is indicated by curves that approach the upper-left corner. The Area Under Curve (AUC) statistic is used to quantify overall detection capabilities; values near 1.0 indicate greater performance.

**4. DISCUSSION**

For IoT threat detection scenarios, our research shows that neural network architectures and other deep learning techniques routinely perform better than conventional machine learning algorithms like decision trees.The confusion matrix results clearly illustrate classification accuracy while highlighting specific misclassification patterns. Minimizing false positive rates remains crucial for practical IoT security implementations to prevent alert fatigue among security teams.The ROC curve analysis and AUC metrics confirm our model's capability to distinguish between legitimate and malicious network activities effectively. Our findings emphasize that thorough feature selection and appropriate data preprocessing significantly impact detection performance. Class imbalance remains a persistent challenge, as security incidents typically represent a small fraction of overall network traffic in real-world environments. Implementing synthetic oversampling techniques like SMOTE demonstrates considerable improvement in detection capabilities by creating more balanced training datasets.For practical deployment in production environments, optimizing models for efficient, low-latency detection proves essential to maintain security coverage without degrading IoT operational performance. Further research opportunities exist in exploring federated learning approaches and distributed ledger authentication methods to enhance security while preserving privacy. Continuous model retraining with emerging attack patterns will ensure adaptability against evolving security threats in IoT ecosystems.

**5. CONCLUSION**

This study shows how well machine learning techniques can detect and reduce security risks in Internet of Things networks. Our findings demonstrate that neural network implementations—in particular, deep learning architectures—offer higher detection accuracy when compared to conventional classification algorithms. The study highlights the importance of appropriate class balance techniques, meticulous feature selection, and meticulous data preprocessing in enhancing detection performance. A strong framework for recognising and resolving new security issues is produced by combining regular model updates with ongoing monitoring capabilities. To further fortify IoT security postures against ever-more-sophisticated threats, future research directions should investigate distributed ledger security techniques and adaptive learning methodologies.

**REFERENCES:**

[1]. Abbas, A., Khan, M. A., Latif, S., Ajaz, M., Shah, A. A., & Ahmad, J. (2021). A new ensemble-based intrusion detection system for internet of things. *Arabian Journal for Science and Engineering*, *47*, 1–15.

[2]. Abdulla, A. R., & Jameel, N. G. M. (2023). A review on IoT intrusion detection systems using supervised machine learning: Techniques, datasets, and algorithms. *UHD Journal of Science and Technology*, *7*(1), 53–65. <https://doi.org/10.21928/uhdjst.v7n1y2023.pp53-65>

[3]. Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. *Electronics*, *11*(2), 198. <https://doi.org/10.3390/electronics11020198>

[4]. Albulayhi, K., Abu Al-Haija, Q., Alsuhibany, S. A., Jillepalli, A. A., Ashrafuzzaman, M., & Sheldon, F. T. (2022). Iot intrusion detection using machine learning with a novel high performing feature selection method. *Applied Sciences*, *12*(10), 5015. <https://doi.org/10.3390/app12105015>

[5].Alsahli, M. S., Almasri, M. M., Al-Akhras, M., Al-Issa, A. I., & Alawairdhi, M. (2021). Evaluation of machine learning algorithms for intrusion detection system in WSN. *International Journal of Advanced Computer Science and Applications*, *12*(5). <https://doi.org/10.14569/IJACSA.2021.012057>

[6]. Alsulami, A. A., Abu Al-Haija, Q., Tayeb, A., & Alqahtani, A. (2022). An intrusion detection and classification system for IoT traffic with improved data engineering. *Applied Sciences*, *12*(23), 12336. <https://doi.org/10.3390/app122312336>

[7]. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8