EXPLORING AI MACHINE LEARNING APPROACHES FOR INTRUSION DETECTION IN IOT SYSTEMS

A. B .Dube
Asst.Professor, Department of Computer Science, Shri Shivaji College of Arts, Commerce and Science, Akola
anitadube108@gmail.com

P. R. Shukla
Asst.Professor, Department of Computer Science, Shri Shivaji College of Arts, Commerce and Science, Akola
poojashuklapriyal@gmail.com

Dr. Prof .V. M. Patil
Ex. HOD & Professor, Department of Computer Science, Shri Shivaji College of Arts, Commerce and Science, Akola
vinmpatil21@yahoo.co.in

Abstract

The exponential growth of the Internet of Things (IoT) has introduced severe security challenges, as conventional security mechanisms are inadequate for resource-constrained devices vulnerable to novel and evolving cyber-attacks. This paper investigates the critical role of Artificial Intelligence (AI) and Machine Learning (ML) in developing efficient and adaptive Intrusion Detection Systems (IDS) for IoT environments. The study presents a comparative analysis of most prominent AI model categories Traditional ML including Supervised, Unsupervised and Hybrid ML. The study concludes that AI-powered IDS are a pivotal innovation for proactive IoT security, capable of learning complex attack patterns and ensuring robust protection for interconnected smart devices.

Keywords*: Supervised Learning, Unsupervised Learning, Hybrid Model, IoT, IDS*

1. Introduction

The proliferation of the Internet of Things (IoT) has ushered in an era of unprecedented connectivity, embedding intelligence into everyday objects from household appliances and wearable devices to critical industrial control systems and smart city infrastructure. This vast network of interconnected, resource-constrained devices generates, processes, and exchanges immense volumes of data. The inherent constraints of IoT devices—such as limited processing power, memory, and energy capacity—often preclude the implementation of robust, traditional security measures, rendering them vulnerable targets for malicious actors. Consequently, IoT systems have become a prime target for sophisticated cyber-attacks, including botnets, denial-of-service (DoS) attacks, data breaches, and ransomware, which can have severe real-world consequences, from privacy violations to physical disruption and economic damage.

Traditional signature-based Intrusion Detection Systems (IDS), which rely on predefined patterns of known threats, are fundamentally ill-equipped to protect IoT ecosystems. Their inability to detect novel, zero-day attacks, coupled with their high computational and storage overhead, makes them unsuitable for the unique architecture of IoT networks (Kikissagbe, 13(18) 2024). This critical security gap has catalysed the urgent need for more intelligent, adaptive, and efficient security solutions. In this context, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative paradigms for constructing next-generation IDS. By learning complex patterns and anomalies from network traffic and device behaviour data, AI techniques can proactively identify both known and previously unseen threats with high accuracy and minimal human intervention.

This paper aims to comprehensively explore and evaluate the application of various AI techniques for intrusion detection in IoT systems. We will investigate mainly on the suitability of different ML models, including supervised learning algorithms for classifying known attacks, unsupervised learning for detecting novel anomalies and reinforcement learning model for adaptive learning and real time decision making. The core research problem we address is the challenge of achieving high detection accuracy and real-time performance within the stringent computational constraints of IoT environments.

2. IoT Security Landscape

2.1 Overview of IoT architecture (devices, networks, applications)

The Internet of Things (IoT) architecture is generally organized into three key layers: **devices, networks, and applications**. Devices—including sensors, actuators, and smart objects—serve as the foundation by capturing and transmitting real-world data. Networks provide the communication channel between devices and central systems through technologies such as Wi-Fi, Bluetooth, Zigbee, cellular networks, and 5G (Mrabet, 2020). Applications then utilize and analyze this data to deliver valuable services across domains like smart homes, healthcare, transportation, industry, and agriculture. Together, these layers enable seamless connectivity, intelligent decision-making, and effective deployment of IoT solutions.

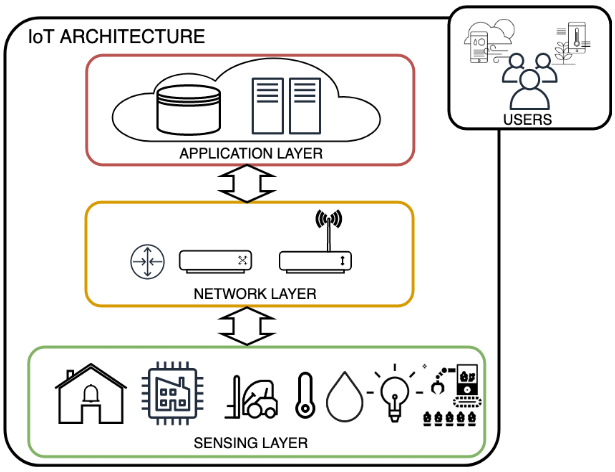

Fig. IoT architecture consisting of sensing, network, and application layers.

Note. Adapted from "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," by Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., & Ayyash, M., 2015, IEEE Communications Surveys & Tutorials, 17(4), 2347–2376.

2.2 Common vulnerabilities and threats in IoT (e.g., DDoS, malware, spoofing, data theft)

1. **Distributed Denial of Service (DDoS) Attacks**
   DDoS attacks (Gelgi, 2024)happen when numerous compromised devices overload a target system, server, or network with massive traffic, exhausting its resources and causing services to become slow or completely inaccessible to legitimate users. One example of a DDoS attack is the 2016 Mirai botnet incident, in which thousands of compromised IoT devices, including cameras and routers, overwhelmed the DNS provider Dyn with heavy traffic, leading to the outage of major websites like Twitter, Netflix, and Reddit for several hours.

2. **Malware and Ransomware**
   These are harmful software designed to disrupt operations, damage systems, or gain unauthorized access. Malware ((n.d.), 2025) includes viruses, worms, and Trojans that can steal information or compromise devices, whereas ransomware ((n.d.) R. , 2025) locks or encrypts files and demands payment for restoration, leading to significant financial and operational impact.

3. **Spoofing and Identity Attacks**
   These attacks ((n.d.) R. , 2025) occur when attackers disguise themselves as legitimate users or devices to gain unauthorized access to sensitive systems or data. By forging IP addresses, emails, or login credentials, they deceive victims and carry out activities such as data theft, fraud, or system exploitation.

4. **Data Theft and Privacy Breaches**
   It happens when unauthorized parties access sensitive information like personal data, financial details, or corporate records. Such incidents can result in identity theft, financial fraud, reputational harm, and a significant loss of trust for individuals as well as organizations (Karim, 2020).

5. **Weak Authentication and Passwords**
   It create serious security vulnerabilities when users depend on simple, reused, or predictable credentials. These weaknesses enable attackers to break into accounts and systems, potentially causing data breaches, identity theft, and unauthorized activities.

3. AI Techniques for Intrusion Detection in IoT

Artificial Intelligence (AI) techniques strengthen intrusion detection in IoT by processing vast amounts of device and network data to uncover suspicious behaviours and anomalies. Methods like machine learning, deep learning, and reinforcement learning support real-time, adaptive, and highly accurate detection of threats beyond the capabilities of traditional approaches.

3.1 Machine Learning (ML) Approaches

Machine Learning (ML) techniques are vital for detecting malicious activities, anomalies, and unauthorized access in IoT environments (Brunel Rolack Kikissagbe, 2024). Given the vast and diverse data generated by IoT networks, ML models can effectively learn attack patterns from historical data and identify previously unseen threats with greater efficiency than traditional rule-based Intrusion Detection Systems (IDS).

3.1.1 Supervised learning (Decision Trees, Random Forests, SVM, KNN)

Supervised learning is a widely applied machine learning approach for intrusion detection (Liu, 2019)in IoT systems. It relies on labelled datasets that include both normal and malicious traffic patterns. By learning from these predefined examples, the model gains the ability to classify new IoT traffic as either benign or malicious. Commonly employed supervised algorithms in this domain include Decision Trees, Random Forests, Support Vector Machines (SVM), and K-Nearest Neighbours (KNN) (Naveen Saran, 2023).

Table 1.1 Comparison of Supervised Learning Method

Working Principle	Use in IoT	Advantages	Limitations
Decision Trees (DT) A tree-like model that splits data based on feature values, forming decision rules at each node.	Suitable for detecting attacks due to their interpretability and fast classification.	Easy to implement, human-readable, and computationally lightweight—ideal for resource-constrained IoT devices.	Prone to overfitting with noisy or complex data.
Random Forest (RF) An ensemble of multiple decision trees, where each tree votes on the classification result.	Provides robust intrusion detection by handling high-dimensional traffic data effectively	High accuracy, reduced overfitting compared to a single decision tree, and scalable to large datasets.	More computationally intensive than individual decision trees, which can challenge low-power IoT devices.
Support Vector Machines (SVM) (Agrawal, 2021) Separates classes by finding the optimal hyperplane in a high-dimensional space.	Effective for binary classification tasks, such as distinguishing between normal and malicious traffic.	High accuracy with small to medium datasets, effective against well-defined attack classes.	Performance decreases with very large datasets, and parameter tuning
K-Nearest Neighbours (KNN) Classifies new data points based on the majority class of their nearest neighbours in feature space.	Useful for anomaly detection in IoT traffic patterns due to its non-parametric nature.	Simple, easy to implement, and effective for small datasets.	Computationally expensive for real-time IoT intrusion detection as it requires distance calculation for every new data point.

3.1.2 Unsupervised learning (Clustering, PCA, anomaly detection)

Unsupervised learning is highly valuable for intrusion detection in IoT systems, where labelled attack data is often limited, incomplete, or unavailable. Rather than depending on predefined categories, these techniques analyse IoT traffic to reveal hidden patterns, group similar behaviours, and identify abnormal activities that may indicate potential intrusions. Common approaches include clustering, dimensionality reduction (PCA), and anomaly detection.

Technique & Working Principle	Techniques Used	Use in IoT	Advantages	Limitations
Clustering Techniques Groups data points with similar characteristics into clusters, while points that do not fit well in any cluster may be treated as anomalies.	K-Means DBSCAN (Density-Based Spatial Clustering of Applications with Noise)	Helps identify unusual traffic flows or previously unseen attack patterns.	Effective for detecting unknown/zero-day attacks.	Sensitive to parameter settings (e.g., number of clusters in K-Means) and may struggle with high-dimensional data.
Principal Component Analysis (PCA) **:** A dimensionality reduction technique that transforms high-dimensional IoT traffic data into a smaller set of principal components while preserving variance.	Covariance Matrix Method Singular Value Decomposition (SVD) Kernel PCA (KPCA)[4]	Helps simplify large traffic datasets, highlighting unusual variations that may indicate malicious behavior.	Reduces computational complexity, useful for lightweight IoT environments.	Assumes linear relationships; may lose important non-linear attack patterns.
Anomaly Detection: Models the normal behavior of IoT traffic and flags deviations as potential intrusions.	Statistical methods (e.g., Gaussian models) Distance-based approaches (e.g., nearest-neighbor outlier detection) Density-based approaches (e.g., Local Outlier Factor – LOF)	Suitable for detecting rare or novel intrusions without requiring labeled attack data.	Effective against zero-day and evolving threats.	Risk of high false positives, as unusual but legitimate IoT behaviors may be misclassified.

4. Comparison of AI Techniques

Aspect	**Supervised Learning** (Anshita Singh, 2025)	**Unsupervised Learning** (Anshita Singh, 2025)
Definition (Anshita Singh, 2025)	Learns from labelled datasets containing both normal and malicious traffic.	Learns from unlabelled data by identifying hidden structures and unusual behaviours.
Data Requirement (Ansam Khraisat, 2019)	Requires large labelled datasets (normal vs. attack traffic).	Works well when labelled data is scarce or unavailable.
Common Algorithms (João Vitorino, 2022)	Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbours (KNN).	Clustering (K-Means, DBSCAN), Dimensionality Reduction (PCA), Anomaly Detection methods.
Detection Capability (Anshita Singh, 2025)	Effective for known attacks since models are trained on historical attack patterns.	Effective for unknown or zero-day attacks, as it detects anomalies beyond known patterns.
Accuracy (João Vitorino, 2022)	Generally high accuracy with sufficient labelled training data.	Accuracy may vary; risk of false positives due to normal but uncommon behaviour being flagged as attacks.
Computational Cost (Anshita Singh, 2025)	Can be computationally expensive depending on dataset size, but optimized algorithms (e.g., Random Forest) work well.	Often lighter in training but may struggle with high-dimensional IoT data.
Adaptability	Limited adaptability to evolving threats, requires retraining with new labelled data.	More adaptable to dynamic IoT environments, can detect emerging attack patterns.
Use Cases in IoT (João Vitorino, 2022)	Smart homes, healthcare, and industrial IoT where labelled datasets are available.	Large-scale IoT networks where attack data is incomplete, and anomaly detection is crucial.
Limitations	Needs labelled data (which is costly to obtain) and may fail with novel attacks.	May generate false alarms and lacks precise classification of attack types.

5. Key Challenges:

Supervised Learning Challenges

Supervised learning requires large amounts of labelled data, which is often impractical in IoT environments. Some key challenges include:

- **Scarcity of Labelled Data:** Collecting and labelling IoT traffic data (normal vs malicious) is costly and time-consuming.
- **Evolving Threats:** Models trained on past attacks may fail to detect novel or zero-day intrusions.
- **Overfitting Risk:** Models like Decision Trees and KNN can overfit training data, reducing generalization on unseen IoT traffic.
- **High Computational Cost:** Training supervised models (e.g., SVM, Random Forest) on large-scale IoT datasets can be resource-intensive, which is problematic for resource-constrained IoT devices.
- **Class Imbalance:** IoT datasets often contain far more benign traffic than attack samples, leading to biased models that under-detect intrusions (Musthafa, 2024).
- **Scalability Issues:** As IoT networks grow, retraining supervised models to handle new devices and traffic patterns becomes challenging.

Unsupervised Learning Challenges

Unsupervised learning overcomes the need for labelled data but brings its own set of limitations:

- **High False Positives:** Legitimate but uncommon IoT behaviours may be misclassified as attacks (Adnan, 2021).
- **Lack of Attack Classification:** Unlike supervised models, most unsupervised approaches only flag anomalies without identifying the exact attack type.
- **Parameter Sensitivity:** Algorithms like K-Means or DBSCAN require careful parameter tuning (e.g., number of clusters, density thresholds) to work effectively.
- **Scalability:** Processing massive IoT traffic in real time using clustering or anomaly detection methods can be computationally demanding (Khan).
- **Interpretability Issues:** Results from PCA or clustering are often difficult to interpret, reducing trust in IoT security decisions.
- **Handling High-Dimensional Data:** IoT traffic data is often complex and multi-dimensional, which can degrade the performance of clustering or anomaly detection methods (Khan).

7. Conclusion:

AI-based IDS represent a fundamental shift from reactive signature matching to proactive, intelligent threat detection. While Deep Learning models show superior performance in handling complex and novel attacks, their computational cost and lack of explainability remain significant hurdles.

Traditional Machine Learning offers a more efficient and interpretable solution for well-defined problems. The future lies in hybrid models that combine the strengths of different AI paradigms, coupled with research into explainability, adversarial robustness, and adaptive learning. For successful deployment, the choice of AI technique must be aligned with the specific network environment, security requirements, and available resources. This taxonomy and comparison serve as a foundation for making these critical decisions and guiding future innovation in the field.

Supervised learning is best suited for scenarios where labelled IoT datasets exist, providing high accuracy in detecting known attack types.

Unsupervised learning is more practical for real-world IoT environments with scarce labelled data, offering better adaptability to unknown or evolving threats, though it may suffer from higher false positives.

Supervised Learning**:** Struggles with data labelling, adaptability to new attacks, and computational demands.

Unsupervised Learning**:** Useful for detecting unknown threats but faces challenges with false positives, scalability, and interpretability.

A hybrid approach (semi-supervised or ensemble methods) is often recommended to balance the strengths and limitations of both supervised and unsupervised learning in IoT intrusion detection.

References:

1. (n.d.), A. (2025, August 23). *Understanding malware and its risks*. Retrieved from Acalvio Technologies: https://www.acalvio.com/resources/glossary/understanding-malware-and-its-risks-acalvio/
2. (n.d.), R. (2025, August 23). *Spoofing attacks: The act of disguising a communication or identity*. Retrieved from https://www.rapid7.com/fundamentals/spoofing-attacks/
3. (n.d.), R. (2025, August 23). *Spoofing attacks: The act of disguising a communication or identity so that it appears to be associated with a trusted*. Retrieved from https://www.rapid7.com/fundamentals/spoofing-attacks/
4. Adnan, A. M. (2021). An intrusion detection system for the Internet of Things based on machine learning: Review and challenges. *Symmetry*. doi:https://doi.org/10.3390/sym13061011
5. Agrawal, A. S. (2021). Comparative analysis of SVM kernels and parameters for efficient anomaly detection in IoT. *In 2021 5th International Conference on Information Systems and Computer Networks (ISCON), IEEE*, 1-6.
6. Ansam Khraisat, I. G. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity 2.1*.
7. Anshita Singh, M. C. (2025). Intrusion Detection System: Comparative Analysis of Supervised and Unsupervised Techniques. . *IJRASET*.
8. Brunel Rolack Kikissagbe, M. A. (2024). Machine Learning-Based Intrusion Detection Methods in IoT Systems : A Comprehensive Review. *Electronics*.
9. Gelgi, M. (2024). Systematic literature review of IoT botnet DDoS attacks and detection techniques. *International Journal of Computers and Applications (published on PMC)*.
10. João Vitorino, R. A. (2022). A Comparative Analysis of Machine Learning Techniques for IoT Intrusion Detection. *International Symposium on Foundations and Practice of Security*.
11. Karim, A. A. (2020). Online identity theft, security issues, and reputational damage. *ResearchGate*.
12. Khan, N. A. (n.d.). Explainable AI-based Intrusion Detection System for Industry 5.0: An Overview of the Literature, Associated Challenges, the Existing Solutions, and Potential Research Directions. *arXiv preprint arXiv:2408.03335.*
13. Kikissagbe, B. R. (13(18) 2024). *Machine Learning-Based Intrusion Detection Methods in IoT Systems: A Comprehensive Review.*
14. Liu, H. &. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. *Applied Sciences*, 9(20), 4396.
15. Mrabet, H. B. (2020). A survey of IoT security based on a layered architecture of sensing and data analysis. Sensors. 3625.
16. Musthafa, M. H. (2024). Optimizing IoT intrusion detection using balanced class distribution, feature selection, and ensemble machine learning techniques. *Sensors*. doi:https://doi.org/10.3390/s24134293 MDPI
17. Naveen Saran, N. K. (2023). A comparative study of supervised Machine Learning classifiers for Intrusion Detection in Internet of Things. *Procedia Computer Science*, 2049-2057.