Global Journal of Computer and Engineering Technology (GJCET)

Vol. 5, No. 1, January, 2025 ISSN: 2767-1933

MACHINE LEARNING BASED CYBER ATTACK DETECTION IN INTERNET OF THINGS AND INDUSTRIAL INTERNET OF THINGS

Bhushan Kumar Kashyap, Atal Bihari Vajpayee Vishwavidyalaya, India (bkk.csa@gmail.com)
Tarun Dhar Diwan, Atal Bihari Vajpayee Vishwavidyalaya, India (tarunctech@gmail.com)
H. S. Hota, Atal Bihari Vajpayee Vishwavidyalaya, India (proffhota@gamil.com)

ABSTRACT

The Industrial Internet of Things (IIoT) is transforming industrial operations through enhanced connectivity, automation and data-driven decision-making. However, this increased interconnectivity significantly broadens the attack surface, exposing critical infrastructures to sophisticated and evolving cyber threats. Addressing these security challenges, this study proposes a high-performance cyber-attack detection framework leveraging machine learning techniques. The proposed model integrates optimized feature selection strategies with advanced classification algorithms to enhance detection accuracy while reducing false alarm rates. Experimental evaluation is conducted on the Edge-IIoTset dataset, a comprehensive benchmark representing diverse IIoT attack scenarios. Comparative performance analysis demonstrates that the proposed framework consistently outperforms existing intrusion detection solutions across key metrics, including accuracy, precision, recall and F1-score. The results highlight the framework's scalability, robustness and adaptability for real-world deployment in industrial environments. This work contributes to the development of proactive, reliable and future-ready cybersecurity measures essential for safeguarding mission-critical IIoT infrastructures.

Keywords: Industrial Internet of Things (IIoT), Cyber Security, Machine Learning, Intrusion Detection, Edge-IIoTset.

1. INTRODUCTION

The Industrial Internet of Things (IIoT) represents a paradigm shift in industrial automation, enabling the convergence of operational technology (OT) and information technology (IT) through the integration of smart sensors, actuators, cyber–physical systems and advanced communication networks (Ferrag et al., 2022). This convergence facilitates real-time monitoring, predictive maintenance, process optimisation and data-driven decision-making in critical infrastructure sectors such as manufacturing, energy, transportation and healthcare. The economic and operational benefits of IIoT are substantial, with global adoption projected to grow exponentially in the coming years (The Industrial Internet Reference Architecture, n.d.). Due to hyper connectivity and vast use of IIoT devices, significant cybersecurity vulnerability introduced. IIoT deployments differ from traditional IT systems and are characterised by heterogeneous devices with varying computational capacities, resource constraints and diverse communication protocols including Message Queuing Telemetry Transport (MQTT), Constrained Application Protocol (CoAP), Modbus and Open Platform Communications – Unified Architecture (OPC-UA) (Sisinni et al., 2018). This heterogeneity, combined with the physical accessibility of devices and often insufficient security configurations, creates an expanded attack surface exploitable by malicious actors. Cyber threats such as Distributed Denial-of-Service (DDoS), man-in-the-middle (MITM), injection attacks, information gathering and malware can lead to production downtime, financial loss, safety risks and compromise of sensitive operational data (W. Z. Khan et al., 2020). Traditional Intrusion Detection Systems (IDSs), particularly signature-based and static anomaly detection approaches, are inadequate in IIoT contexts due to their limited adaptability to evolving attack patterns, inability to handle large-scale heterogeneous data streams, and performance degradation under resource-constrained conditions(Jose et al., 2018; Samita, 2024). Signature-based IDS can detect known threats but struggles with novel attacks and requires frequent updates. At the same time, anomaly-based approaches offer better adaptability but often suffer higher false-positive rates and demand significant computational resources (Jose et al., 2018). Consequently, there is an imperative need for intelligent, adaptive and scalable intrusion detection frameworks capable of identifying both known and zero-day threats in real time while maintaining operational efficiency (Rahman et al., 2025). To address these challenges, this study presents a machine learning–based cyber-attack detection framework explicitly designed for IIoT environments and evaluated using the Edge-IIoTset dataset — a comprehensive, realistic cybersecurity dataset tailored for IoT and IIoT attack scenarios (Ferrag et al., 2022). The proposed approach leverages advanced feature selection, optimised model architectures and hybrid classification strategies to enhance detection accuracy while minimising false alarms. By incorporating a wide range of attack categories and realistic network behaviours, the framework ensures practical applicability for modern industrial systems.

2. LITERATURE REVIEW

Aragonés et al. (2025) stated that the Industrial Internet of Things has become a critical enabler of Industry 4.0, facilitating real-time connectivity between industrial assets, cyber–physical systems and cloud services. That is the reason attackers target that environment.

Cyber-Attack Landscape in IIoT: The Industrial Internet of Things (IIoT) facilitates seamless integration between industrial devices, cyber–physical systems and cloud-based analytics platforms (Aragonés et al., 2025). This integration enables predictive maintenance, process optimisation and intelligent decision-making across manufacturing, energy, transportation and healthcare sectors. The proliferation of interconnected devices and heterogeneous communication protocols has expanded the attack surface, making IIoT environments highly vulnerable to cyber threats (Ammar et al., 2018; Tabaa et al., 2020). Distributed Denial-of-Service (DDoS) attacks can overwhelm network bandwidth and computational resources. In contrast, man-in-the-middle (MITM) attacks compromise confidentiality and integrity by intercepting and altering communications (Conti et al., 2018). Injection attacks exploit protocol vulnerabilities to insert malicious commands, reconnaissance techniques gather network intelligence to facilitate further exploitation, and malware infections disrupt operations or steal sensitive data (Restuccia et al., 2018). The consequences of such attacks can include production downtime, safety risks and substantial economic losses.

Machine Learning Based Intrusion Detection for IIoT: Machine learning (ML) has been increasingly applied to intrusion detection systems (IDS) due to its capacity to model complex patterns and adapt to evolving threats (Apruzzese et al., 2018). Supervised machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting have demonstrated strong detection performance for known attack patterns (Ferrag et al., 2020; Jyothsna & Prasad, 200 C.E.). But they have a limited ability to detect zero day attacks. Unsupervised methods, including clustering algorithms like k-means and DBSCAN, can identify anomalies without prior attack signatures but often suffer from high false-positive rates in noisy IIoT environments (Haque et al., 2021). Deep learning techniques, such as Convolutional Neural Networks (CNN) (Shone et al., 2018), Long Short-Term Memory (LSTM) networks (Noh & Moon, 2023) and autoencoders (Shone et al., 2018b) have shown promise in capturing spatial and temporal dependencies within IIoT traffic data. Hybrid approaches that integrate anomaly-based and signature-based detection can improve coverage for known and novel threats (Yao et al., 2019), yet they often incur high computational costs, challenging deployment on resource-constrained IIoT devices.

IIoT Cybersecurity Datasets: The availability of high-quality datasets decides the performance of machine learning based Intrusion Detection System. Datasets like NSL-KDD and UNSW-NB15(Z. I. Khan et al., 2024) were primarily designed for traditional IT networks; they lack coverage of IIoT specific devices and protocols. The CICIDS2017 dataset (Tareq et al., 2022) includes a wide range of traffic and attack types but it falls short when capturing the complex and varied nature of Industrial IoT (IIoT) systems. More recent datasets, like TON_IoT (Ferrag et al., 2022) focus more directly on IoT and IIoT traffic. Yet, they still face challenges, particularly in covering different communication protocols and providing a broader spectrum of attack scenarios. To overcome these shortcomings the Edge-IIoTset dataset has been developed offering realistic IIoT traffic data from heterogeneous devices, multiple protocols and various cyberattack scenarios, thereby enabling more robust evaluation of intrusion detection models (Xu et al., 2014).

2.1. Research Gap: While machine learning has brought big improvements to intrusion detection in Industrial IoT several hurdles still remain. A lot of research relies on datasets that don't fully reflect the complexity of real-world IIoT environments, which makes it hard to apply the findings broadly. Handling large volumes of fast-moving IIoT traffic is another challenge, as scalability and real-time processing are not always reliable. On top of that, spotting zero-day attacks is still a major weakness, especially when detection methods rely only on supervised learning. Additionally, only a limited number of studies evaluate intrusion detection solutions in realistic IIoT testbeds that incorporate heterogeneous devices and protocol diversity (Ammar et al., 2018; Xu et al., 2014). This study addresses these gaps by leveraging the Edge-IIoTset dataset to design and evaluate a scalable, accurate and practical Machine Learning based intrusion detection framework for securing Industrial IoT infrastructures.

3. THEORETICAL FRAMEWORK

This study is built on the core ideas of Industrial Internet of Things (IIoT) architecture, cybersecurity principles, different types of cyber-attacks and machine learning models for intrusion detection.

Industrial Internet of Things (IIoT) Architecture: The IIoT architecture integrates physical industrial processes with advanced computing, networking, and data analytics technologies, enabling automation, operational efficiency, and real-time decision making (Xu et al., 2014). In general, the IIoT architecture is composed of multiple layers, each playing a specific role:

i. **Perception Layer** – Comprising IoT and IIoT devices such as temperature sensors, pH meters, ultrasonic detectors, and actuators, this layer is responsible for physical data collection and initial signal processing(Ammar et al., 2018).

ii. **Network Layer** – Ensures secure and reliable data transmission using industrial communication protocols such as MQTT, CoAP, Modbus TCP/IP, and OPC-UA(Tabaa et al., 2020).
iii. **Edge Computing Layer** – Performs computation close to the data source, reducing latency and network load, which is critical for time-sensitive industrial operations(Gope & Sikdar, 2019).
iv. **Fog Computing Layer** – Serves as an intermediate processing layer between edge devices and the cloud, enabling distributed computation and scalability(Restuccia et al., 2018).
v. **Cloud Computing Layer** – Provides large-scale storage, data analytics, and centralized management capabilities.

Application Layer – Hosts industrial applications such as predictive maintenance, process optimization, and quality monitoring(Stallings, 2009).

The distributed and heterogeneous nature of this architecture improves operational flexibility but also broadens the attack surface, making IIoT systems highly vulnerable to cyber threats.

Cybersecurity Requirements in IIoT: Security in IIoT systems is defined by a set of core objectives that ensure data confidentiality, operational reliability, and system resilience(Conti et al., 2018):

i. **Confidentiality** – Protecting sensitive operational data from unauthorized disclosure.
ii. **Integrity** – Ensuring that transmitted or stored data remains accurate and unaltered during its lifecycle.
iii. **Availability** – Guaranteeing that systems and services remain operational and accessible despite cyberattacks or component failures.
iv. **Authentication** – Verifying the legitimacy of devices, users, and processes before granting access to resources.
v. **Non-repudiation** – Preventing entities from denying their actions or transactions within the system(Apruzzese et al., 2018).

Failure to maintain these objectives can lead to disruptions in production, financial losses, reputational damage, and safety hazards in mission-critical operations.

Cyber-Attack Taxonomy in IIoT: IIoT systems are vulnerable to a range of cyber threats, which can be categorized into five broad types (Ferrag et al., 2022):

i. **Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS)** – Overloading network or computational resources to disrupt services (Jyothsna & Prasad, 200 C.E.).
ii. **Man-in-the-Middle (MITM) Attacks** – Intercepting and altering communications between legitimate parties.
iii. **Injection Attacks** – Exploiting vulnerabilities in industrial protocols to inject malicious control commands or data.
iv. **Information Gathering** – Performing reconnaissance to map system architecture, identify vulnerabilities, and plan future attacks.
v. **Malware Attacks** – Deploy malicious software to damage system, steal sensitive data or disrupt normal operations (Alam et al., 2024).

Understanding this taxonomy is critical for designing detection models that can identify a diverse range of attack vectors with minimal false positives.

Machine Learning for Intrusion Detection in IIoT: Machine learning (ML) offers a data driven approach to intrusion detection by automatically identifying patterns and anomalies in network traffic (Shone et al., 2018). In the IIoT context, Machine Learning techniques can be classified as:

i. **Supervised Learning** – Relies on labeled datasets to train models like Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting classifier. Effective for known attack types but often less capable against zero-day threats (Noh & Moon, 2023).
ii. **Unsupervised Learning** – Detects anomalies without prior knowledge of attack signatures using clustering algorithms such as k-means and DBSCAN. Useful for zero-day detection but may generate high false alarm rates (Niyaz et al., 2015).
iii. **Deep Learning** – Employs architectures like Convolutional Neural Networks (CNN) for spatial pattern recognition and Long Short-Term Memory (LSTM) networks for temporal sequence modelling, offering improved feature extraction and detection accuracy (Ferrag et al., 2022; Khacha et al., 2022).

Hybrid Approaches – Combine anomaly-based and signature-based detection to improve both detection accuracy and coverage, albeit at the cost of higher computational complexity (Fawcett & Provost, 1999).

4. DATA DESCRIPTION

The Edge-IIoTset dataset (Ferrag et al., 2022) is a comprehensive and realistic cybersecurity dataset specially developed for IoT and IIoT applications, capturing network traffic from heterogeneous devices operating under both normal and attack conditions. It contains 1,909,671 records with 63 attributes encompassing packet level metadata, protocol-specific features and attack annotations, covering a wide range of industrial communication protocols such as MQTT, Modbus TCP/IP, HTTP, DNS and ICMP. Features of the dataset contains various protocol layers including ARP parameters (arp.opcode, arp.hw.size), ICMP fields (icmp.checksum, icmp.seq_le,

icmp.unused), HTTP characteristics (http.content_length, http.response, http.request.method, http.referer, http.request.version), TCP/UDP attributes (tcp.ack, tcp.seq, tcp.len, udp.stream, udp.time_delta), DNS queries (dns.qry.name, dns.qry.type, dns.qry.name.len, dns.qry.qu), MQTT control and topic details (mqtt.conack.flags, mqtt.topic, mqtt.ver, mqtt.proto_len) and Modbus TCP identifiers (mbtcp.len, mbtcp.trans_id, mbtcp.unit_id). Each instance in the dataset is labeled with a binary variable (Attack_label) to indicate whether the traffic is normal or malicious, as well as a multiclass variable (Attack_type) specifying the exact nature of the traffic. The dataset includes one normal class and 14 attack categories: DDoS_UDP, DDoS_ICMP, SQL Injection, DDoS_TCP, Vulnerability Scanner, Password, DDoS_HTTP, Uploading, Backdoor, Port Scanning, XSS, Ransomware, Fingerprinting, and MITM. The diversity of protocols, features, and attack types makes Edge-IIoTset a robust benchmark for developing and evaluating intrusion detection models in IIoT environments.

5. METHODOLOGY

This study proposed a comprehensive experimental framework for detection and classification of cyber attacks in Industrial Internet of Things (IIoT) environments using the publicly available Edge-IIoTset cybersecurity dataset. The methodology encompasses systematic steps including dataset acquisition, preprocessing, feature engineering, data balancing, model implementation and performance evaluation as illustrated in the experimental workflow (Figure 1).

Dataset Acquisition: This study utilized the Edge-IIoTset cybersecurity dataset, a state-of-the-art benchmark dataset publicly available on the Kaggle platform, specially designed for IoT and Industrial Internet of Things (IIoT) security research. This dataset comprises an extensive 1,909,671 meticulously captured network traffic records encompassing 63 rich and diverse features extracted from real world inspired packet captures within simulated IIoT network environments.

Removal of Irrelevant and Redundant Features: To enhance computational efficiency and prevent overfitting, non-informative and redundant attributes were systematically removed. These included raw metadata fields such as frame.time, IP addresses (ip.src_host, ip.dst_host) and port numbers (tcp.srcport, tcp.dstport, udp.port) as well as protocol-specific payloads like http.file_data, tcp.payload and mqtt.msg. Such features, while unique per record, do not contribute meaningfully to generalized attack detection and may introduce model bias. This step reduced the original 63 features to a refined set of 48 variables, focusing on relevant statistical and protocol-level indicators.

Handling Missing and Duplicate Records: Data integrity was ensured by removing all rows containing missing values (NaN) and eliminating duplicate records. This rigorous cleaning process guaranteed that only complete, unique entries remained in the dataset effectively preventing both noise related inconsistencies and data leakage during model training. Post cleaning verification confirmed the absence of missing values (0 null entries).

Randomization through Dataset Shuffling: To mitigate any unintended temporal or sequential dependencies the dataset was randomly shuffled. This step ensured that model training was not influenced by inherent ordering in the dataset, enabling fair and unbiased learning across the entire spectrum of benign and attack records.

Class Distribution Assessment: An analysis of the Attack_type column revealed a highly imbalanced distribution, with the Normal class comprising approximately 1.36 million instances, while rare attack types such as MITM accounted for as few as 358 records. This imbalance highlights the challenge of minority class detection and informed the need for applying advanced class balancing strategies in subsequent stages.

Encoding Categorical Attributes and Data Type Conversion: Categorical variables, including network protocol fields (http.request.method, mqtt.protoname, etc.) and the target label Attack_type were transformed into numerical form using Label Encoding. Additionally, any object-type fields containing numerical values were coerced into numeric format to ensure dataset uniformity. The encoded target variable, Attack_type_encoded, enabled direct compatibility with machine learning and deep learning algorithms while preserving the original categorical semantics.

Final Dataset Characteristics: Following preprocessing, the dataset comprised 1,909,671 records and 48 fully numeric features, alongside an encoded target label. All attributes were in float64 or int64 format with no missing values. The dataset contained 15 distinct traffic classes one benign and fourteen malicious capturing a diverse range of attack vectors, including DDoS, SQL Injection, Ransomware and MITM.

After preprocessing, the dataset was split into training and testing sets using a 70:30 ratio, ensuring that 70% of the data was used for training the model, while the remaining 30% was reserved for evaluating model performance on unseen data.

6. MODEL DEVELOPMENT

To build an effective and comparative framework for cyber-attack detection in Industrial IoT (IIoT) environments, six supervised learning models were developed. These include both traditional machine learning algorithms and deep learning architectures. Each model was selected based on its suitability for handling high-dimensional, multi-class classification problems as represented in the Edge-IIoTset dataset. The following subsections elaborate on the individual models and their development strategies (Figure 1).

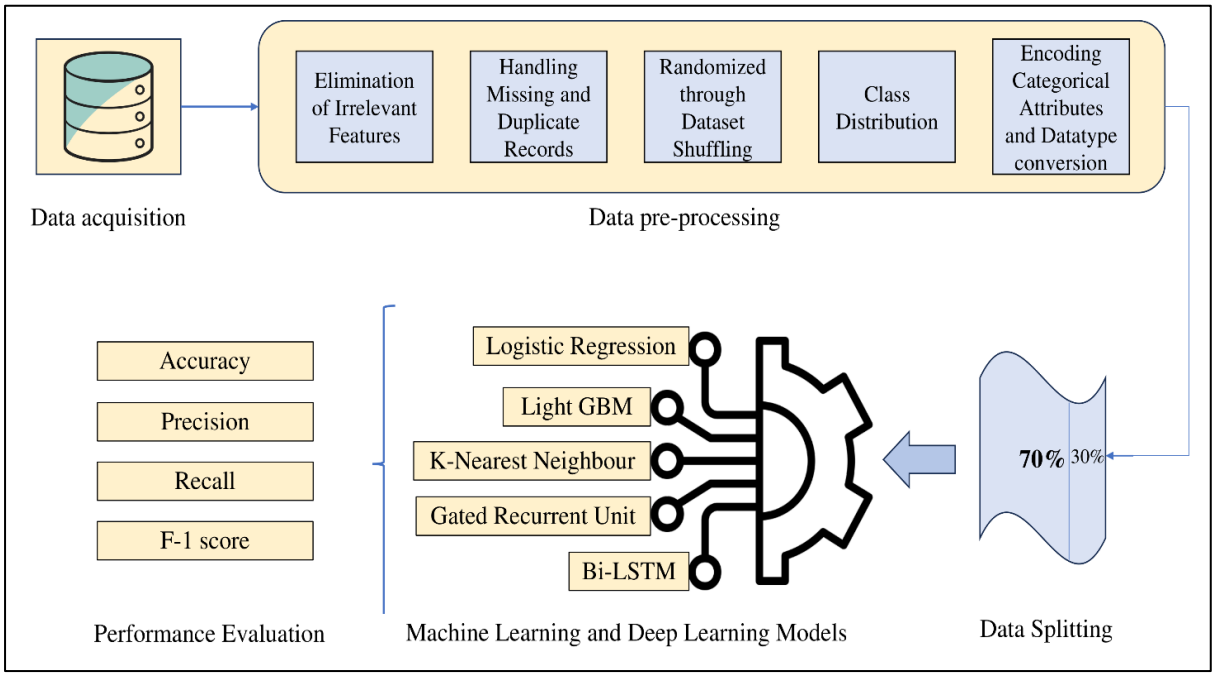

Figure 1: Workflow of the Proposed Machine Learning Model

Logistic Regression (LR): Logistic Regression was chosen as the baseline model because it is simple, easy to understand, and works well for problems where things can be clearly separated into categories. To handle multiple categories, the model was set up to classify using a method that considers all categories together. Before training, the data features were standardized using a common technique to put all variables on a similar scale, which helps the model learn faster. The training process was given more iterations, up to 1000, to make sure the model fully optimized its performance, especially since the dataset was large and complex.

K-Nearest Neighbors (KNN): K-Nearest Neighbors (KNN) is a straightforward algorithm used to understand how data points group together locally. It works by looking at the closest neighbors around a point—measured using a distance like Euclidean distance—and assigning the most common class among those neighbors to that point. In this case, we chose to look at the 5 nearest neighbors after testing to find the best fit. Since KNN is sensitive to differences in feature scales, we first normalized the data to keep everything consistent. Although simple, KNN provided a helpful baseline to see how well the data forms natural clusters.

Light Gradient Boosting Machine (LightGBM): LightGBM was chosen because it's fast and handles large, organized datasets very well. It uses a smart way of building decision trees, growing them leaf by leaf, which helps the model learn quicker and more accurately. For this study, we started with the default settings to see how well it performs right out of the box. LightGBM is especially handy because it can work directly with categorical data and missing values, and it can train using multiple processors at the same time. It also helps us understand which features are most important for making predictions, adding transparency to the model.

Bidirectional Long Short-Term Memory (Bi-LSTM): Even though the data didn't involve time or sequences, we decided to try a Bidirectional Long Short-Term Memory (Bi-LSTM) network to see if it could capture relationships between features looking both forwards and backwards. We reshaped the input to make it look like a sequence for the model. The Bi-LSTM had one layer with 64 units, followed by dropout and dense layers to help with learning and avoid overfitting. We trained the network using the Adam optimizer and a loss function designed for multi-class problems. Surprisingly, even though the data wasn't time-based, the Bi-LSTM performed well by using information from both directions, which helped it get a better overall understanding of the input features.

Gated Recurrent Unit (GRU): Although the data didn't have a time-based or sequence structure, we tested a Gated Recurrent Unit (GRU) network to see if it can capture dependencies that behave like sequences. The data was reshaped to resemble sequential input for the model. The GRU had one layer with 64 units, followed by dropout and dense layers to help with learning and avoid overfitting. It was trained using the Adam optimizer with a loss function suited for multi-class classification. Even without explicit time-series data, the GRU performed well because its efficient gating system helps it keep important information and forget irrelevant details, while also being faster and less complex than traditional LSTMs.

7. EVALUATION METRICS FOR MODELS

The trained models were rigorously evaluated on a reserved test dataset to assess their effectiveness in detecting and classifying cyberattacks within the IIoT environment. To ensure comprehensive evaluation, several performance metrics were considered, each addressing specific aspects of model prediction quality. These include Accuracy, Precision, Recall, F1-score, Confusion Matrix, and Receiver Operating Characteristic – Area Under Curve (ROC-AUC). The inclusion of these metrics provides a robust assessment framework, especially important when working with imbalanced multiclass datasets common in cybersecurity applications (Figure 1).

Accuracy: Accuracy is the most intuitive metric, representing the proportion of correct predictions out of the total predictions made:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} (\text{Rainio et al.}, 2024) \quad \ldots\ldots\ldots\ldots\ldots\ldots (i)$$

While useful as an overall indicator, accuracy can be misleading in imbalanced datasets, as it may favor the dominant (normal) class and ignore the misclassification of minority classes, such as rare attack types(Buda et al., 2018).

Precision, Recall, and F1-Score: To mitigate the limitations of accuracy and provide a more class-sensitive evaluation, the following metrics were employed:

Precision evaluates the accuracy of positive predictions. It is the ratio of true positives to all predicted positives.

$$\text{Precision} = (\text{TP})/(\text{TP} + \text{FP}) (\text{Rainio et al.}, 2024) \quad \ldots\ldots\ldots\ldots\ldots\ldots (ii)$$

Recall (or Sensitivity) measures the ability to correctly identify all relevant instances (i.e., true positives).

$$\text{Recall} = (\text{TP})/(\text{TP} + \text{FN}) (\text{Rainio et al.}, 2024) \quad \ldots\ldots\ldots\ldots\ldots\ldots (iii)$$

F1-score is the harmonic mean of Precision and Recall, balancing both concerns in a single metric.

$$\text{F1-Score} = 2(\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) (\text{Rainio et al.}, 2024) \quad \ldots\ldots\ldots\ldots\ldots\ldots (iv)$$

These metrics were computed for each class and averaged using:

Macro average, treating all classes equally.

Weighted average, considering the class distribution. Such detailed class-wise evaluations are essential in intrusion detection systems, where false negatives (missed attacks) can have serious security implications(Weiss, n.d.).

Confusion Matrix: To analyze the prediction behavior of models at the class level, confusion matrices were constructed. A confusion matrix helps visualize the number of correct and incorrect predictions for each class, making it easier to identify which types of attacks are often misclassified. This form of error analysis is critical for refining model performance in real-world deployment where detection of minority attack types is as crucial as detecting dominant ones(Fawcett & Provost, 1999).

ROC Curve and AUC Score: The Receiver Operating Characteristic (ROC) curve and its Area Under the Curve (AUC) were used for threshold-independent model evaluation. The ROC curve plots the True Positive Rate against the False Positive Rate, while the AUC summarizes this curve into a single value ranging from 0.5 (no discrimination) to 1.0 (perfect classification). For multiclass settings, the One-vs-Rest strategy was adopted to compute individual ROC curves for each class, and the average AUC was reported(Hanley & McNeil, n.d.). The ROC-AUC metric is especially useful for evaluating models in cybersecurity domains due to its ability to reflect a model's sensitivity to classification thresholds across various classes, even in the presence of class imbalance(Saito & Rehmsmeier, 2015).

8. RESULT AND DISCUSSION

The comparative performance analysis of five algorithms—Logistic Regression, Light Gradient Boosting Machine (Light GBM), K-Nearest Neighbour (KNN), Gated Recurrent Unit (GRU), and Bi-directional Long Short-Term Memory (Bi-LSTM)—on the Edge-IIoTset dataset demonstrates significant performance variations across cyberattack categories. Logistic Regression achieved an overall accuracy of 91%, delivering high precision and recall for attacks such as Backdoor (99% precision, 97% recall), DDoS ICMP (100%, 97%), and MITM (97%, 100%), but showing notable weaknesses in detecting SQL Injection (77%, 81%), Uploading (88%, 76%), and Vulnerability Scanner (83%, 91%). Light GBM emerged as the top-performing traditional algorithm, achieving a flawless 100% accuracy, precision, and recall across all attack types, indicating exceptional classification capability and generalization. KNN recorded an overall accuracy of 88%, excelling in Fingerprinting (98%, 99%)

and SQL Injection (100%, 100%) but struggling significantly with Uploading (72%, 64%) and Normal traffic (79%, 78%). GRU achieved 99% accuracy, maintaining perfect detection rates for the majority of attacks, with only minor performance drops in Vulnerability Scanner (100% precision, 82% recall) and XSS (85% precision, 100% recall). Bi-LSTM attained an outstanding 100% accuracy, with near-perfect precision and recall for all classes, showing minimal reductions in MITM (99% recall), Normal traffic (99%, 99%), and Vulnerability Scanner (100% precision, 90% recall) (Table 1). Overall, the results confirm that while conventional models like Logistic Regression and KNN provide competitive performance for specific attack types, advanced deep learning models such as GRU and Bi-LSTM, along with the LightGBM ensemble method, consistently outperform them—achieving near-perfect detection rates and demonstrating superior robustness, scalability, and applicability for real-time IIoT cyberattack detection. The classification report of the models presented in Table 1 was derived using Equations (i)–(iv).

Table 1: Classification Report of Models

Algorithm	Accuracy	Target	Precision	Recall	F1-score
Logistic Regression	91	Backdoor	99	97	98
		DDoS HTTP	96	98	97
		DDoS ICMP	100	97	99
		DDoS TCP	90	99	95
		Fingerprinting	86	85	86
		MITM	97	100	99
		Normal	100	78	88
		Password	87	95	91
		Port Scanning	93	99	96
		Ransomware	90	89	89
		SQL Injection	77	81	79
		Uploading	88	76	81
		Vulnerability Scanner	83	91	87
		XSS	83	91	87
Light GMB	100	Backdoor	100	100	100
		DDoS HTTP	100	100	100
		DDoS ICMP	100	99	100
		DDoS TCP	100	100	100
		Fingerprinting	99	100	99
		MITM	100	100	100
		Normal	100	100	100
		Password	100	100	100
		Port Scanning	100	100	100
		Ransomware	100	100	100
		SQL Injection	100	100	100
		Uploading	100	100	100
		Vulnerability Scanner	100	100	100
		XSS	100	100	100
K-Nearest Neighbour	88	Backdoor	88	93	90
		DDoS HTTP	97	84	90
		DDoS ICMP	100	98	99
		DDoS TCP	94	100	97
		Fingerprinting	98	99	98
		MITM	100	100	100
		Normal	79	78	79
		Password	80	81	80
		Port Scanning	94	97	95
		Ransomware	90	94	92
		SQL Injection	100	100	100
		Uploading	72	64	68
		Vulnerability Scanner	100	100	100

		XSS	84	80	82
GRU	99	Backdoor	100	100	100
		DDoS HTTP	100	100	100
		DDoS ICMP	100	100	100
		DDoS TCP	100	100	100
		Fingerprinting	100	100	100
		MITM	100	99	100
		Normal	99	99	99
		Password	100	100	100
		Port Scanning	100	100	100
		Ransomware	100	100	100
		SQL Injection	100	100	100
		Uploading	100	100	100
		Vulnerability Scanner	100	82	90
		XSS	85	100	92
Bi-LSTM	100	Backdoor	100	100	100
		DDoS HTTP	100	100	100
		DDoS ICMP	100	100	100
		DDoS TCP	100	100	100
		Fingerprinting	100	100	100
		MITM	100	99	99
		Normal	99	99	99
		Password	100	100	100
		Port Scanning	100	100	100
		Ransomware	100	100	100
		SQL Injection	100	100	100
		Uploading	100	100	100
		Vulnerability Scanner	100	99	90
		XSS	98	100	99

The Receiver Operating Characteristic (ROC) curve analysis was performed for all five evaluated models—Bi-LSTM, GRU, KNN, LightGBM, and Logistic Regression—across multiple IIoT attack categories. Both Bi-LSTM and GRU models achieved near-perfect classification performance, with Area Under the Curve (AUC) values of 1.00 for almost all attack types, and only a marginal drop for the Normal class (AUC = 0.99 in Bi-LSTM). LightGBM similarly exhibited flawless discrimination capability, achieving an AUC of 1.00 across all classes, indicating exceptional separability between benign and malicious traffic. In contrast, KNN showed high but slightly varied performance, with AUC values ranging from 0.91 (Vulnerability Scanner) to 1.00 (multiple classes such as DDoS ICMP, DDoS TCP, Fingerprinting, and MITM), reflecting its sensitivity to feature space distribution. Logistic Regression performed reasonably well but demonstrated visible variations, with several classes such as SQL Injection (AUC = 0.99) and Uploading (AUC = 0.99) slightly lower than the deep learning and LightGBM models. Overall, the ROC curve analysis reinforces that Bi-LSTM, GRU, and LightGBM deliver near-ideal detection performance with almost zero false-positive rates, while traditional models like Logistic Regression and KNN, though effective, show minor degradations in certain attack classes—highlighting the superiority of advanced deep learning and gradient boosting methods for IIoT intrusion detection (Figures 2,3,4,5,6).

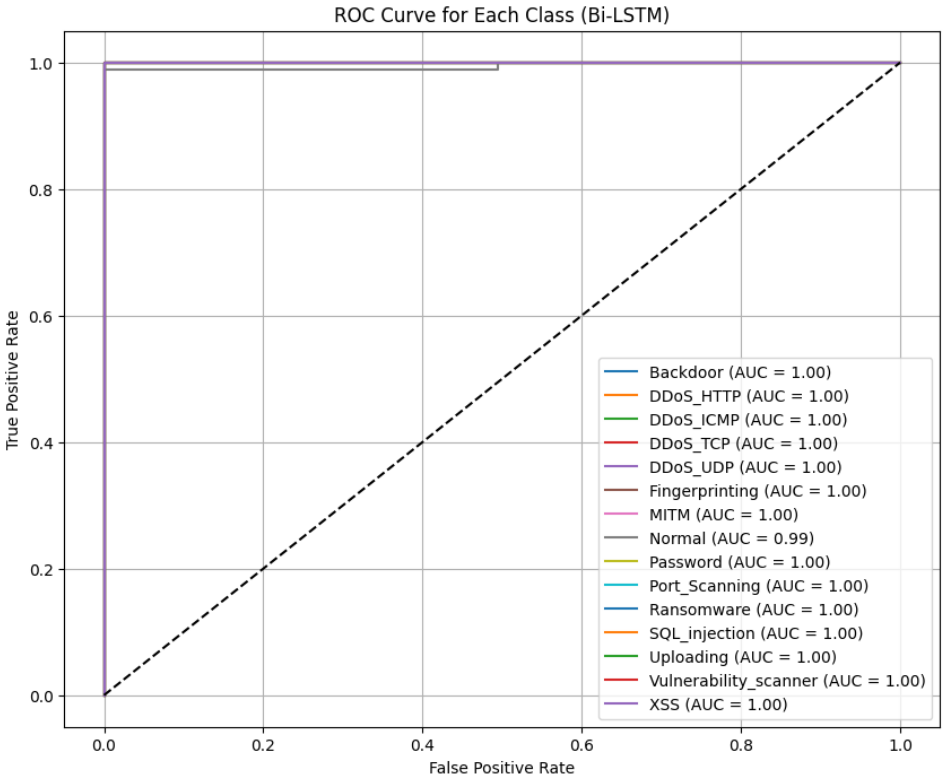

Figure 2: AUC-ROC Curve for Multi-Class IIoT Attack Detection Using Bi-LSTM

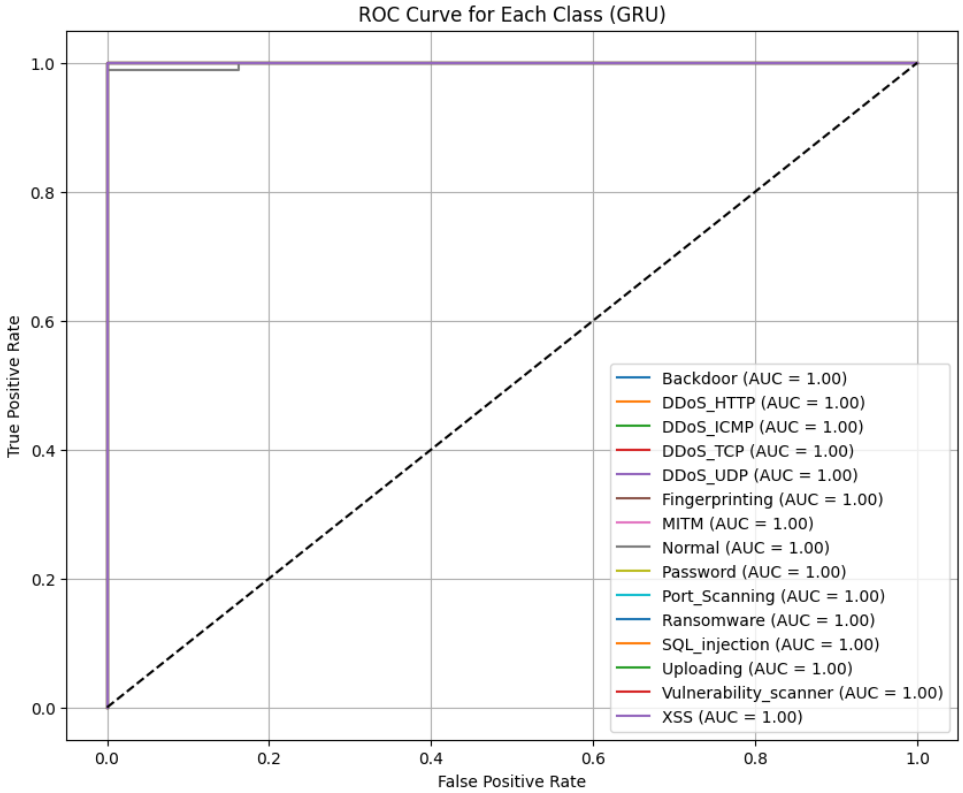

Figure 3: AUC-ROC Curve for Multi-Class IIoT Attack Detection Using GRU

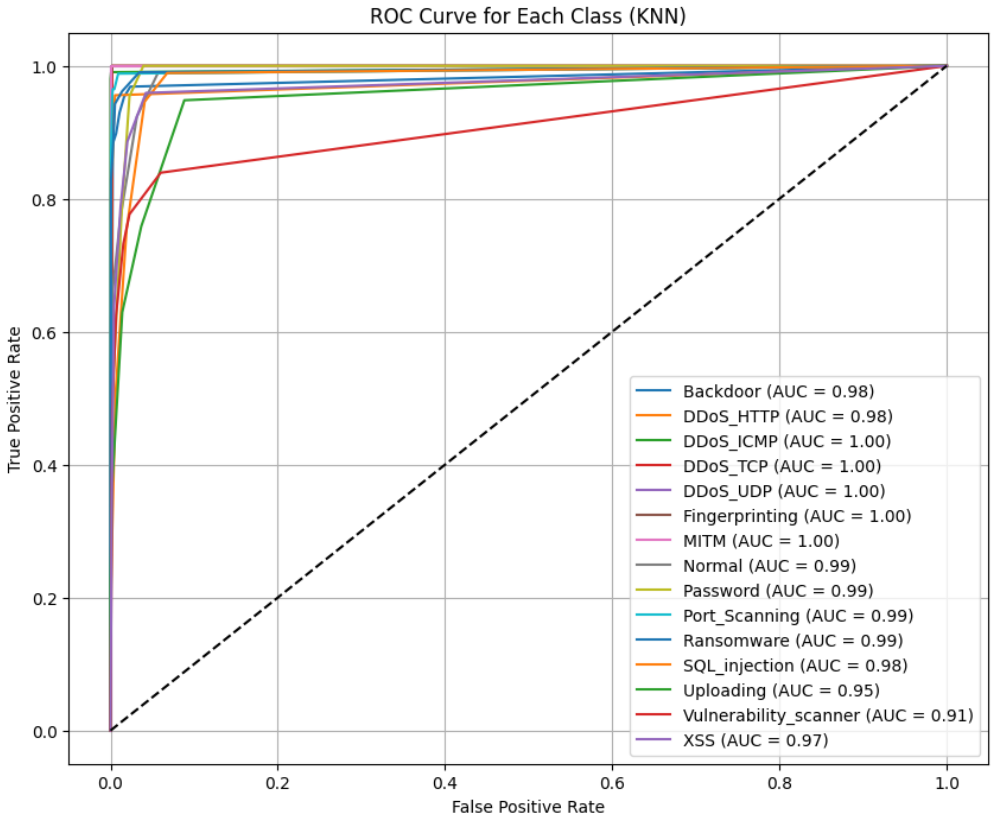

Figure 4: AUC-ROC Curve for Multi-Class IIoT Attack Detection Using K-NN

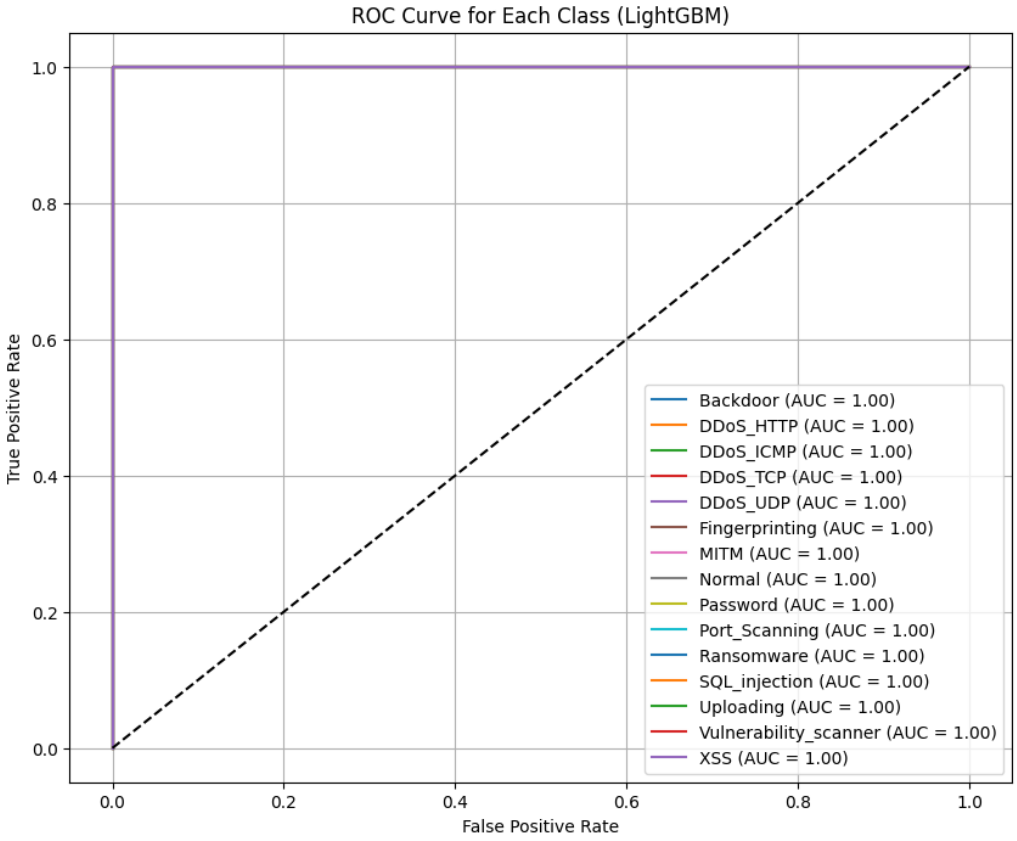

Figure 5: AUC-ROC Curve for Multi-Class IIoT Attack Detection Using LightGBM

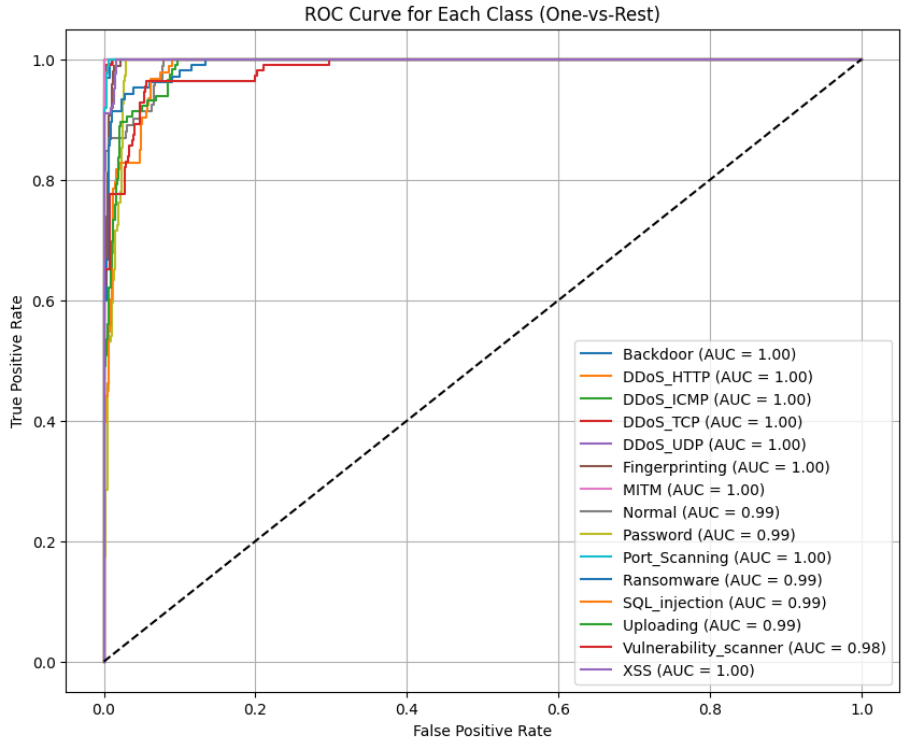

Figure 6: AUC-ROC Curve for Multi-Class IIoT Attack Detection Using Logistic Regression

9. CONCLUSION

The Industrial Internet of Things (IIoT) offers transformative benefits for industrial operations, but its increased connectivity makes it highly susceptible to sophisticated cyber threats. This study proposed and evaluated a high-performance cyber-attack detection framework using multiple machine learning and deep learning models, benchmarked on the Edge-IIoTset dataset. The results demonstrate that advanced models—particularly LightGBM, GRU and Bi-LSTM—significantly outperform traditional algorithms like Logistic Regression and KNN, achieving near-perfect detection rates across a wide range of attack scenarios. LightGBM's flawless accuracy and deep learning models' ability to capture temporal patterns in network traffic highlight their suitability for real-time, large-scale IIoT security applications. The comparative analysis confirms that such approaches offer superior precision, recall, and robustness, ensuring minimal false alarms while maintaining scalability for industrial deployment. In conclusion, this research underscores the importance of adopting advanced AI-driven intrusion detection mechanisms to protect critical IIoT infrastructures. Future work will focus on integrating the proposed framework with adaptive learning capabilities to address emerging attack vectors, enhance resilience, and enable proactive defense strategies in dynamic industrial environments.

REFERENCES

Ammar, M., Russello, G., & Crispo, B. (2018). Internet of Things: A survey on the security of IoT frameworks. Journal of Information Security and Applications, 38, 8–27. https://doi.org/10.1016/j.jisa.2017.11.002

Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. International Conference on Cyber Conflict, CYCON, 2018-May, 371–389. https://doi.org/10.23919/CYCON.2018.8405026

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. https://doi.org/10.1016/j.neunet.2018.07.011

Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. In Future Generation Computer Systems (Vol. 78, pp. 544–546). Elsevier B.V. https://doi.org/10.1016/j.future.2017.07.060

Fawcett, T., & Provost, F. (1999). Analysis and Visualization of Classifier Performance: Comparison Under Imprecise Class and Cost Distributions Analysis and Visualization of Classiier Performance: Comparison under Imprecise Class and Cost Distributions. https://www.researchgate.net/publication/2637219

Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. IEEE Access, 10, 40281–40306. https://doi.org/10.1109/ACCESS.2022.3165809

Gope, P., & Sikdar, B. (2019). An Efficient Privacy-Preserving Authentication Scheme for Energy Internet-Based Vehicle-to-Grid Communication. IEEE Transactions on Smart Grid, 10(6), 6607–6618. https://doi.org/10.1109/TSG.2019.2908698

Hanley, J. A., & McNeil, B. J. (n.d.). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve1.

Jose, S., Malathi, D., Reddy, B., & Jayaseeli, D. (2018). A Survey on Anomaly Based Host Intrusion Detection System. Journal of Physics: Conference Series, 1000(1). https://doi.org/10.1088/1742-6596/1000/1/012049

Rainio, O., Teuho, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-56706-x

Restuccia, F., D'Oro, S., & Melodia, T. (2018). Securing the Internet of Things in the Age of Machine Learning and Software-Defined Networking. IEEE Internet of Things Journal, 5(6), 4829–4842. https://doi.org/10.1109/JIOT.2018.2846040

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE, 10(3). https://doi.org/10.1371/journal.pone.0118432

Samita. (2024). A Review on Intrusion Detection System for IoT based Systems. In SN Computer Science (Vol. 5, Number 4). Springer. https://doi.org/10.1007/s42979-024-02702-x

Stallings, William. (2009). Business data communications. Pearson/Prentice Hall.

Tabaa, M., Monteiro, F., Bensag, H., & Dandache, A. (2020). Green Industrial Internet of Things from a smart industry perspectives. Energy Reports, 6, 430–446. https://doi.org/10.1016/j.egyr.2020.09.022

Weiss, G. M. (n.d.). Mining with Rarity: A Unifying Framework.