International Journal of Safety and Security Engineering
Vol. 15, No. 8, August, 2025, pp. 1733-1744
Journal homepage: http://iieta.org/journals/ijsse

A Review on Intrusion Detection Systems for MQTT in IoT Environments

Kholoud Nasser Al Maawi*, Monir Abdullah

College of Computing and Information Technology, University of Bisha, Bisha 61922, Saudi Arabia

Corresponding Author Email: kholudnasser884@gmail.com

https://doi.org/10.18280/ijsse.150818

Received: 27 July 2025
Revised: 20 August 2025
Accepted: 25 August 2025
Available online: 31 August 2025

Keywords:
ensemble learning, intrusion detection systems, IoT security, machine learning, MQTT security, MQTTset dataset, feature importance, SHAP explainability

ABSTRACT

The rapid expansion of the Internet of Things (IoT) has highlighted critical security holes, particularly in lightweight protocols of message communication such as Message Queuing Telemetry Transport (MQTT). Conventional security measures tend to be ineffective against the constantly evolving cyber world, therefore highlighting the necessity of effective intrusion detection systems (IDS). The present work involves a critical examination of the measures of IDS, particularly designed for MQTT-based IoT applications based on supervised, unsupervised, and ensemble learning (EL) techniques. Out of the 25 papers examined, supervised models achieved accuracies of up to 99%, while ensemble techniques always produced F1-scores above 95%. We discuss the datasets used in the paper for training and validation, reporting that MQTTset dominates as the most frequently used, appearing in 65% of the five-year-old papers. We also highlight the importance of the techniques of feature importance, like Shapley additive explanations (SHAP) and principal component analysis (PCA), in reducing computational overhead while being highly accurate. The review clarifies the merits and drawbacks of current IDS strategies and identifies the main challenges related to the scaling issue, interpretability, and diversity of datasets. The research presents consolidated evidence beneficial in the construction of durable machine learning (ML)-based MQTT IDS and identifies promising directions in which the development should be continued.

1. INTRODUCTION

The Internet of Things (IoT) has become a significant technological and economic catalyst, interlinking billions of devices within consumer, industrial, and service sectors [1]. Although this connectivity offers the potential for enhanced efficiency and innovation, it simultaneously raises significant issues related to security, privacy, and system resilience. The increasing incidence of security vulnerabilities and reported cyberattacks targeting IoT-enabled devices emphasizes the pressing need for effective protection measures [1]. Communication among IoT devices predominantly occurs via application-layer protocols, which can be broadly categorized into request/response (e.g., HTTP, CoAP) and publish/subscribe (e.g., MQTT, AMQP) [2]. Notably, Message Queuing Telemetry Transport (MQTT), initially created by IBM in 1999, has gained widespread adoption owing to its lightweight architecture, small header size, and effectiveness in resource-constrained scenarios [2]. However, despite these advantages, MQTT is deficient in built-in security mechanisms, rendering it susceptible to various attacks such as denial-of-service, spoofing, and message flooding [3, 4]. The weakness has triggered researchers to design intrusion detection systems (IDS) specifically for IoT networks that are focused on MQTT. IDS are devised with the purpose of tracing host or network activities while identifying potential malicious or anomalous activities. The systems are usually of host-based, network-based, or hybrid type and apply signature-based, anomaly-based, or hybrid detection approaches [5]. When applied to IoT and MQTT environments, IDS assumes critical importance due to the escalating diversity and complexity of assaults aimed at compromising confidentiality, integrity, and availability. Recent research indicates that machine learning (ML) and ensemble learning (EL) approaches yield commendable results, often achieving accuracies exceeding 95% on datasets tailored for MQTT applications [6]. Nevertheless, persistent challenges exist, including issues related to dataset imbalance, scalability across varied IoT environments, and the interpretability of complex ML-oriented IDS. This questionnaire offers an in-depth overview of IDS strategies devised for MQTT-based IoT, summarized as supervised, unsupervised, and EL techniques. It also explores the most frequently used datasets in this field, specifically MQTTset, responsible for 65% of the current studies, and the importance of feature analysis as well as explainability techniques like Shapley additive explanations (SHAP) and principal component analysis (PCA) in enhancing efficiency and interpretability. Through the aggregation of current research, this paper recognizes open opportunities and offers insights in terms of directions that can inform the development of more robust IDS for MQTT networks. The rest of this paper is structured as follows: Section 2 overviews MQTT-based IDS research and compares supervised, unsupervised, and ensemble techniques. Section 3 summarizes ML strategies and considers deployment trade-offs. Section 4 overviews MQTT datasets as well as attributes. Section 5 details feature analysis as well as explainability techniques used in MQTT. Section 6 considers open opportunities as well as future study, and Section 7 concludes.

2. LITERATURE REVIEW

Security challenges in IoT environments are a major topic of interest to researchers. Research has focused on traditional security mechanisms, IDSs, and solutions specifically designed for MQTT-based IoT networks. This section provides a comprehensive review of the previous literature, with a critical analysis of the methodologies, results, and limitations of current approaches.

A. Traditional security mechanisms in IoT

The IoT has revolutionized how devices communicate and interact, enabling widespread applications such as smart homes and industrial automation. As these networks expand, significant security challenges have emerged.

Traditional security mechanisms rely primarily on centralized trust models and password-based authentication. However, these approaches are becoming insufficient to address IoT vulnerabilities. Furthermore, limited resources and a lack of standards further exacerbate these vulnerabilities, necessitating the need for more robust solutions [7, 8].

Many traditional IoT security architectures operate on a centralized trust model, relying on a central entity (such as a central authentication authority) to manage identity verification and authorization processes. However, this approach faces problems, most notably the presence of a "single point of failure." A breach of this central entity can compromise the entire network, as in automotive networks, where a breach of the trusted entity leads to widespread security vulnerabilities [7].

On the other hand, some devices rely on passwords for identity verification, but this approach suffers from numerous vulnerabilities. Devices that rely on weak or default passwords can be easily compromised through guessing, dictionary attacks, and social engineering [7, 8]. Password management issues, such as reuse across multiple accounts or insecure storage, also increase the likelihood of password theft [7].

Furthermore, wireless networks face increasing threats, such as fake access point (AP) attacks, where attackers create fake AP to intercept authentication data and steal credentials [7].

A significant number of IoT devices have limited computational power and energy resources, which makes the implementation of conventional cryptographic systems impractical. To address this challenge, lightweight cryptographic approaches have been developed specifically for resource-constrained IoT platforms. While these methods can reduce computational overhead, they may not always provide sufficient protection against more advanced threats [9, 10]. Elliptic curve cryptography (ECC) is an effective solution that balances security and computational efficiency. It combines centralized management with direct authentication between devices, enhancing communication security and reducing the burden on resource-limited devices [11].

In addition, the lack of unified security standards increases the vulnerability of IoTs. Many devices rely on outdated or proprietary security mechanisms, making them vulnerable to cyberattacks. The diversity of IoT device manufacturers also leads to fragmented security implementations, making it difficult to establish comprehensive security policies that protect the entire system [8].

In general, traditional security mechanisms such as centralized trust models and password authentication struggle to address modern IoT threats, requiring the adoption of more advanced solutions such as zero-trust architectures (ZTA) and advanced authentication techniques such as radio fingerprinting (RFF) and ECC, which can effectively improve IoT security.

B. IDSs on IoTs

Since traditional security mechanisms struggle to protect IoT networks from advanced cyber threats, IDSs have become an important defence tool. These systems monitor network traffic, analyze communication patterns, and detect suspicious activity that may indicate potential attacks.

Unlike static security methods, IDSs can identify known and unknown threats by relying on signature detection techniques, anomaly detection, and ML models.

Alsoufi et al. [12] demonstrated that the use of deep learning techniques, such as convolutional neural networks (CNNs), deep belief networks (DBNs), autoencoders (AEs), and long short-term memory (LSTMs), can achieve accurate results in intrusion detection. The results showed that the CNN model achieved an accuracy ranging from 76.76% to 99.88%, while the DBN achieved an accuracy between 97% and 97.21%. The AE model achieved accuracy between 80% and 99.81%, while the LSTM achieved accuracy ranging from 79.58% to 98%.

Wei et al. [13] demonstrated that a hybrid detection model based on autoencoders (AHDM) was developed to improve attack detection in small data samples. The model demonstrated superiority over DNN, AFE + DNN, and ACID models in terms of accuracy and lower false alarm rates.

Logeswari et al. [14] demonstrated the successful use of the quantum swarm optimization (QIPSO) algorithm for feature selection in IDS systems, which improved the accuracy of intrusion detection. The system was integrated with ANFIS to improve feature selection, and capsule networks (CapsNets) and recurrent neural networks (RNNs) were used to achieve a high accuracy of 98.83% on ToN-IoT data and 98.96% on BoT-IoT data.

Alrayes et al. [15] demonstrated an IDS based on a denoised autoencoder (DAE) was developed, which enhanced detection accuracy by reconstructing the input data and filtering out noise. The system demonstrated an accuracy of 99.991% when tested on CICIDS 2017 data, confirming its effectiveness in detecting sophisticated attacks.

Finally, a study by Sana et al. [16] reviewed the use of ML and deep learning techniques to improve IDS performance in IoT networks. Models powered by Random Forest (RF) and Bootstrap Trees achieved a high accuracy of 99.90%, while the LSTM model achieved an accuracy of 99.97%. The vision transformers (ViT) model demonstrated 100% accuracy in tests, making it a promising option for enhancing IoT security.

C. Existing IDS solutions for MQTT in IoT

With the increasing reliance of IoT networks on the MQTT protocol for data transmission, protecting these networks from cyberattacks has become a major challenge. IDSs designed for MQTT-based IoT networks detect malicious activity while maintaining efficiency and speed.

Ullah et al. [17] presented a TNN-IDS model based on a Transformer Neural Network to improve intrusion detection performance. The study used the extra tree classifier (ETC) algorithm to extract important features that help detect malicious activity.

Three performance optimization algorithms (SGD, RMSProp, and Adam) were tested with different batch sizes, and the model was trained for six rounds using the Sparse Categorical Cross-Entropy algorithm to calculate losses. The model achieved 99.9% accuracy on the MQTT-IoTIDS2020 dataset, outperforming traditional methods. However, the model requires significant computing resources, which may hinder its use in resource-constrained IoT environments.

Alaiz-Moreton et al. [18] proposed a hybrid method for selecting critical features to improve attack detection in MQTT-based IoT networks. The study used the mRMR technique for selecting critical features in the first stage, and then used algorithms such as SVM, Decision Tree (DT), and RFs to improve detection accuracy.

A custom MQTT dataset was created for this research, containing three CSV files dedicated to denial of service (DoS), man-in-the-middle (MitM), and Intrusion attacks. The results showed that the RF algorithm achieved 99.38% accuracy for DoS attacks, while the XGBoost algorithm performed best overall

Hanif and Ilyas [19] focused on detecting DoS and brute force attacks in MQTT-based IoT networks using ML algorithms such as RF, DT, KNN, and XGBoost. Using the MQTTset dataset, the model achieved an accuracy of 95.38% when using EL techniques such as stacking, voting, and bagging.

Mosaiyebzadeh et al. [20] presented an IDS based on deep learning techniques (NIDS) to detect attacks on the MQTT protocol. Trained on the MQTT-IoT-IDS2020 dataset, the model focused on identifying MQTT brute-force threats. The model achieved an average accuracy of 97.09% and an F1 score of 98.33%. However, the study indicated a need to improve the quality of the training data.

Omotosho et al. [21] reviewed centralized and federated learning methods for detecting attacks in MQTT-based IoT networks by implementing six ML models: DT, RF, logistic regression (LR), linear discriminant analysis (LDA), Gaussian Naive Bayes (GNB), and deep neural network (DNN). Using the MQTT-IoT-IDS2020 dataset, the RF algorithm achieved 99% accuracy in detecting DoS and MitM attacks.

Alasmari et al. [22] examined ML-based IDS to protect smart-home IoT devices from MQTT attacks. The authors evaluated a total of 22 ML models and concluded that the generalized linear model (GLM) was among the top-performing classifiers. On the MQTT-IoT-IDS2020 dataset, the GLM model attained a superior accuracy of 100% following random over-sampling to address class imbalance.

Siddharthan et al. [23] introduced a new dataset called SEN-MQTTset, which contains features specific to the MQTT protocol. The research proposed an intelligent IDS leveraging the best ML models and dataset optimization techniques. The system achieved an accuracy exceeding 99% in detecting attacks.

Khan et al. [24] proposed a DNN-based model for detecting attacks on the MQTT protocol. Using the MQTT-IoT-IDS2020 dataset, the model achieved 97.13% accuracy in classifying multiple attacks.

Table 1. A comparative analysis of various MQTT-based IDS solutions

Ref.	Problem	Dataset	Methods	Results	Strengths	Limitations
[17]	Detecting MQTT-based attacks with deep feature learning	MQTT-IoT-IDS2020	Transformer NN + Extra Tree Classifier	99.9% Accuracy	High accuracy, advanced deep feature extraction	High computational cost, less suitable for IoT edge devices
[18]	Multiclass detection of DoS, MitM, and Intrusion attacks	Custom MQTT dataset	mRMR + SVM, DT, RF, GRU RNN	99.4% Accuracy	Feature optimization improved classification	Limited dataset generalization
[19]	Detecting DoS and brute-force attacks	MQTTset	RF, DT, KNN, XGBoost + Ensembles	95.38% Accuracy	Ensemble improved detection over base models	Moderate accuracy, performance sensitive to features
[20]	Detecting brute-force and flooding attacks	MQTT-IoT-IDS2020	DNN	97.09% Accuracy, F1 = 98.33%	High detection of brute-force attacks	The quality of training data affects performance
[21]	Generalization for minority-class attacks	MQTT-IoT-IDS2020	Federated + Centralized ML (RF, DT, LR, LDA, DNN)	Up to 100% Accuracy	Strong performance across models	Federated models increase complexity
[22]	Protecting smart-home IoT devices	MQTT-IoT-IDS2020	GLM + Oversampling + AutoML	100% Accuracy	A balanced dataset improved detection	Overfitting risk due to oversampling
[23]	Detecting attacks in IoT-MQTT with a new dataset	SEN-MQTTset	Optimized Ensemble + Statistical Feature Gen	> 99% Accuracy	Dataset tailored to MQTT-specific features	New dataset requires validation
[24]	Detecting multi-label MQTT intrusions	MQTT-IoT-IDS2020 + custom	DNN with flow features	99.92% Accuracy	Handles multi-class scenarios	Deep models require high resources
[25]	Identifying unknown/novel attacks	MQTT-IoT-IDS2020 + custom	GAN-AE (GAN + Autoencoder)	97% Accuracy	Effective for anomaly detection	GAN models are harder to train, resource-intensive
[26]	Improving MQTT IDS with ensemble methods	MQTTset	Bagging, Boosting, Stacking Ensembles	F1 = 95%, MCC > 90%	Consistently strong performance	Higher resource consumption

Boppana et al. [25] presented a model called GAN-AE that combines generative adversarial networks (GANs) and autoencoders (AEs) to improve malicious activity detection in MQTT-enabled IoT networks. The model was trained and validated with two distinct datasets: MQTT-IoT-IDS2020, an openly available dataset of flow-based network features, and a private dataset created in prior work to simulate real MQTT network configurations. The model achieved 97% accuracy on the two different datasets.

Zeghida et al. [26] investigated the application of EL methods for improving IDS accuracy in MQTT security. With the use of the MQTTset dataset, the research demonstrated that EL could greatly enhance the performance of IDS, with an F1-score of 95% and an MCC of over 90%.

To present a systematic comparison of these solutions, Table 1 overviews important details of each study. The table presents a brief analysis of current IDS solutions, indicating their approach and efficiency in securing MQTT communication.

These studies exhibit clear patterns of MQTT-specific IDS studies. Supervised learning (SL) models (e.g., RF, SVM, DNN) are continuously high in accuracy, in many cases above 95%, but they are overwhelmingly reliant on labeled data, which is expensive and often imbalanced. Unsupervised methods (e.g., GAN-AE) hold promises for the detection of unknown and novel attacks but are difficult to train and require higher computational power. Ensemble methods (e.g., bagging, boosting, stacking) generally deliver the most stable results, with F1-scores above 95% in several cases, but at the expense of increased computational overhead, which may limit deployment in resource-constrained IoT environments.

In short, although there have been advances in optimizing IDS performance for MQTT-IoT, future work is desirable to focus on lightweight ensemble models, on generating standardized and balanced data, and on the inclusion of explainable AI (XAI) methods for optimizing trustworthiness and interpretability without sacrificing efficiency.

3. MACHINE LEARNING APPROACHES FOR MQTT IDS

ML is a branch of artificial intelligence that focuses on developing algorithms capable of analyzing data, detecting patterns, and making decisions without the need for direct programming. ML allows systems to automatically learn and improve performance over time based on experience.

The term "learning" in ML refers to the process of exploring different representations of data to determine the most appropriate one based on available information. The term "machine" refers to the use of mathematical and logical algorithms to perform these processes automatically, enabling automatic decision-making and pattern recognition [27]. ML has become an important component of enhancing network security, particularly in improving IDSs and overcoming the limitations of traditional rule-based methods [28, 29]. Using traditional ML algorithms and deep learning models, it is possible to analyze large amounts of network traffic data, improve anomaly detection, and more accurately identify cyber threats with greater flexibility [30]. They enable the automation of threat analysis, pattern identification, and detection of previously unknown attacks and hence are an integral component in today's cybersecurity architectures [29]. ML-based intrusion detection solutions in the MQTT environment are divided into three main categories: SL, unsupervised learning, and EL.

A. Supervised learning approaches

SL is an ML method that relies on labeled data to train models for classification and prediction tasks. In this method, a dataset containing pairs of inputs and outputs is provided, allowing the model to learn the relationship between these variables with the help of a supervisor. The goal is to develop a system capable of accurately predicting outputs when confronted with new, unknown data [31]. SL methods include a group of algorithms such as regression models, DTs, support vector machines, and k-nearest neighbor (k-NN) algorithms. These methods are widely used in fields such as image recognition, natural language processing, and fraud detection [32, 33]. Among the common algorithms, the most prominent are NB, RF, and neural networks, the former and latter being preferred for their high accuracy. The performance of these models is evaluated using multiple benchmarks to ensure their reliability [32, 33]. By identifying patterns within labelled data, SL enables predictive modelling that supports informed decision-making across diverse fields [33].

For example, Anthi et al. [34] proposed a three-layer IDS for smart home IoT devices, using SL techniques to analyze device behavior and detect malicious packets in real time. Ashraf et al. [35] built an advanced security system based on multiple algorithms such as decision trees, support vector machines, multi-layer neural networks, random decision forests, and LR. Their focus on rigorous pre-processing ensured that irrelevant features were removed, thereby enhancing detection accuracy.

In addition, Akintoye et al. [36] contributed to the domain by suggesting a novel IDS framework that combined decision trees, GNB, k-NNs, LR, RF, and support vector machines, including feature selection, dataset resampling, and normalization. The research attained enhanced classification accuracy on NSL-KDD and UNSW-NB15 datasets. Huang et al. [37] compared supervised models for anomaly-based IDS to demonstrate that decision trees, Naive Bayes, and k-NNs had the potential to effectively improve network security by detecting out-of-order traffic patterns. A novel work by Farooq [38] suggested a multi-layer classification mechanism that leverages decision trees, fuzzy logic, and neural networks for automatically uncovering new attack signatures and multi-stage cyberattacks. Collectively, the above works highlight the compelling necessity of SL-based IDS for the betterment of network security in IoT environments.

As opposed to supervised approaches founded on pre-labeled information, unsupervised learning follows a different path by identifying inherent patterns within the data without pre-existing labels.

B. Unsupervised learning approaches

Unsupervised learning is a technique that relies on analyzing unlabelled data to discover hidden patterns and structures in the data. This method relies on exploring internal relationships between data points, making it particularly effective for anomaly detection and data segmentation [39]. Unsupervised learning methods include clustering techniques such as K-means and hierarchical clustering, as well as dimensionality reduction methods such as PCA and factor analysis. These methods are used in various fields, such as computer vision, speech recognition, and natural language processing [40, 41].

Bhadauria and Mohanty [42] presented a hybrid framework for intrusion detection that combined signature-based detection with anomaly detection. DT and NB algorithms were used to classify known attacks, while clustering algorithms such as DBSCAN and Isolation Forest were used to detect anomalies, improving detection rates and reducing false alarms.

Wang et al. [43] proposed UTEN-IDS, an unsupervised EL system aimed at IoT environments. Based on autoencoders and the Isolation Forest algorithm, UTEN-IDS efficiently detected anomalies in MQTT traffic with improved performance in identifying new attacks.

Building on these methods, Jha et al. [44] suggested an immune system-based IDS that emulates the adaptive behaviour of human T-cells and B-cells. This system was found to detect unknown and known threats with high accuracy and a very low false alarm rate when tested with the KDD99 dataset.

Additionally, Alom and Taha [45] utilized unsupervised deep learning methods, specifically Auto Encoders, Restricted Boltzmann Machines, and k-means clustering to realize feature extraction and dimensionality reduction with detection accuracy ranging from 91.86% to 92.12%.

Also, Idrissi et al. [46] designed EdgeIDS, an unsupervised host IDS using a GAN, specifically for resource-constrained IoT devices. EdgeIDS reported ROC-AUC scores as high as 0.99 on the MQTTset dataset, thereby proving suitable for effective anomaly detection in practical scenarios.

C. Ensemble learning method

EL is a method that combines multiple ML models to improve the accuracy and reliability of the predictive system. Unlike depending on a single individual classifier, EL aggregates the positive aspects of heterogeneous models to avoid errors and improve generalization [47]. This method is well-suited to minimizing the imperfections typical of single ML models, like higher variance, bias, and poor generalization ability. EL fortifies the global predictive capability of a system through the aggregation of multiple weak learners [48].

The principal objective of EL is to reduce prediction error and enhance the stability of models. This is done by taking advantage of the collective decision-making strength of various models, which ensures that the mistakes made by individual classifiers are made up for by the advantages provided by others [49]. EL is extensively utilized in real-world applications to increase both classification accuracy and reliability and is an important tool in fraud detection, medical diagnosis, and cybersecurity [50].

The success of EL lies in its three primaries: bagging, boosting, and stacking. Bagging, or Bootstrap Aggregating, is the process of training multiple instances of the same model on various subsets of the data to achieve variance reduction and stability enhancement. RF is one of the key methods in the bagging family that combines decision trees to enhance predictive capability [51]. Boosting, on the other hand, is the sequential training of models, with each subsequent model attempting to correct the errors of the previous one. Techniques like AdaBoost, Gradient Boosting, and XGBoost are great instances of this strategy that enhance model prediction using iterative learning procedures [52]. Stacking employs a meta-learning strategy, where various base learners are individually trained, and the last model combines their output to maximize overall predictive performance [53].

The importance of EL lies in the fact that it can improve the predictive power and stability of the model, especially when individual models struggle to generalize well. It is being extensively applied in ML competitions and real-world settings because of its better performance compared to individual models [54]. As the field of ML continues to expand, ensemble techniques continue to be a significant strategy for creating very accurate and robust predictive models capable of conforming to intricate patterns of data.

All three ML approaches to MQTT IDS have characteristics of each other's strengths and weaknesses. Supervised methods dominate in accuracy, but they are burdened with imbalanced dataset issues and limited flexibility. Unsupervised methods are good at discovering innovative threats, but they are prone to generating false alarms and require complex models. Ensemble models consistently produce the most robust results, with F1-scores in many works higher than 95%, though at a steep computational price, as shown in Table 2.

Table 2. Comparison of ML approaches for MQTT IDS

Approach	Example Algorithms	Strengths	Limitations
Supervised	DT, RF, SVM, k-NN, Neural Networks	High accuracy; reliable for known attacks	Needs labeled data; poor at zero-day detection
Unsupervised	K-means, DBSCAN, PCA, Autoencoders, GANs	Detects novel/zero-day attacks	High false positives; complex to interpret
Ensemble	Bagging (RF), Boosting (XGBoost), Stacking	Stable, robust, best across datasets	Computationally heavy; unsuitable for constrained IoT

Performance of MQTT IDS solutions in the literature is evaluated mainly through standard evaluation criteria borrowed from the confusion matrix. The main criteria are Accuracy, Precision, Recall, F1-score, and Error Rate. These criteria list IDS performance evaluation comprehensively and reflect both detection ability and error biases. Most recent studies emphasize the F1-score for Accuracy because it considers the imbalance of the data and making it a preferred metric in MQTT-based IDS evaluation [55, 56].

4. DATASETS FOR ML-BASED MQTT IDS IN IOT

Datasets play a central role in training and testing ML-driven IDS for MQTT-based IoT networks. They simulate both normal traffic and numerous cyber-attacks, providing the ground truth needed for training and testing detection models [57, 58]. Recent research demonstrates that dataset choice directly affects detection accuracy, generalizability, and computational efficiency. This section reviews the most widely used MQTT datasets.

4.1 MQTT-IoT-IDS2020 dataset

Developed by Hindy et al. [59], the MQTT-IoT-IDS2020 dataset filled the gap in publicly available datasets tailored to MQTT traffic. It simulates benign traffic alongside four attack scenarios: Adversive Sweeping, UDP Sweeping, SSH brute force, and MQTT brute force. Data was collected in a controlled IoT testbed with 12 MQTT sensors, a broker, a simulated camera feed, and an attacker generating malicious traffic [58, 59].

4.2 Custom MQTT dataset

Proposed by Alaiz-Moreton et al. [18], this dataset contains three CSV files targeting different scenarios: DoS (94,625 frames), MitM (110,668 frames), and Intrusion (80,893 frames). Each file includes labeled benign and attack traffic, allowing researchers to evaluate IDS models under controlled MQTT attack conditions.

While the Custom MQTT Dataset focuses on a range of attack scenarios, another dataset, MQTTset, was designed specifically for smart home environments.

4.3 MQTTset dataset

Developed by Vaccari et al. [60], MQTTset was designed for smart home environments. It contains 33 MQTT-specific features from sensors such as temperature, humidity, CO^2, motion, and door status. Eight IoT devices communicated via the Mosquitto MQTT broker, simulating realistic home automation. As researchers attempted to optimize datasets and feature selection more, the SEN-MQTTset dataset was proposed to improve the performance of ML algorithms.

4.4 SEN-MQTTset dataset

Introduced by Siddharthan et al. [23], SEN-MQTTset enhanced MQTTset by expanding attack types and feature richness. It includes scenarios of normal operation, subscriber attacks (e.g., Connect Flooding), and broker attacks. Traffic was collected using Raspberry Pi and NodeMCU devices, with 120 raw features across flow-, TCP-, IP-, and MQTT-specific layers [23].

These datasets play a vital role in the development and evaluation of ML-based IDSs for MQTT networks. providing diverse and realistic attack scenarios that contribute to effective model training and testing, as shown in Table 3.

Table 3. Summary of datasets for MQTT IDS

Dataset	Year	Size / Samples	Attack Types	Strengths	Limitations
MQTT-IoT-IDS2020	2020	Multi-format (packet/flow/session)	Sweeping, Brute Force (SSH, MQTT)	Public, diverse attack categories	Imbalanced, noisy, limited attacks
Custom MQTT	2019	DoS: 94k, MitM: 110k, Intrusion: 80k	DoS, MitM, Intrusion	Clear scenarios, labeled	Small scale, limited diversity
MQTTset	2022	33 MQTT features, smart-home sensors	Multiple MQTT-specific	MQTT-focused, balanced, realistic IoT	Imbalance in cases
SEN-MQTTset	2022	120 features, multi-view	Normal, Subscriber, Broker Attacks	Rich feature set, multi-layer	Smaller, proprietary preprocessing

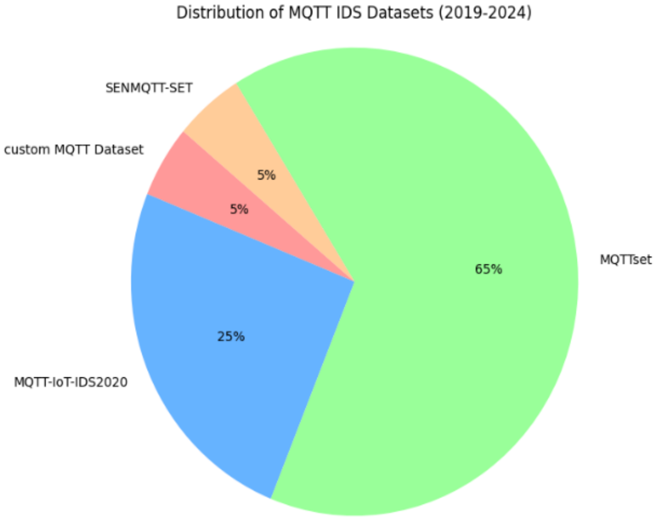

Figure 1. Dataset usage in MQTT IDS research

The above figure discusses the datasets that have been used in this study to evaluate the performance of ML-based MQTT IDS. Figure 1 presents a review of openly accessible datasets that were collected in the last five years. As can be seen, the MQTTset dataset has been used the most, with 65% of the studies. The MQTT-IoT-IDS2020 dataset follows with a 25% utilization rate, and the SENMQTT-SET and certain tailor-made MQTT datasets both have a 5% utilization rate.

5. FEATURES ANALYSIS FOR MQTT INTRUSION DETECTION

Feature analysis plays an important role in ML-based IDS, particularly for MQTT environments, as it reduces data size, removes redundancy, and highlights the most relevant MQTT traffic features. It entails the selection of a subset of significant features, thereby fostering data organization and computational efficiency without the model performance being compromised [61, 62].

For this purpose, several analysis techniques are used, such as statistical analysis, dimensionality reduction, visualization-based analysis, and feature importance assessment. Table 4 is a comparison of these methods. The following subsections explain these techniques, their applications to IDS, and their relevance for MQTT.

5.1 Statistical analysis

Statistical analysis is a fundamental technique in data science that helps organize and interpret big data effectively. This includes hypothesis testing, regression analysis, and maximum likelihood estimation—techniques originally developed for mechanical calculating machines. With the advancement of modern computing, more sophisticated statistical methods have emerged that require fewer assumptions about the distribution of data, allowing researchers to analyze data and draw meaningful conclusions with greater flexibility [63, 64].

In engineering applications, statistical methods are used to describe data using distribution functions and metrics such as the mean and standard deviation. These tools also help build predictive models using small samples, incorporating confidence intervals to estimate sampling errors [63]. These statistical tools are valuable in IDSs, where they are used to analyze network traffic and detect abnormal patterns that indicate cyber threats. For MQTT IDS, statistical analysis can identify unusual message lengths or irregular publish frequencies. While effective for basic anomaly detection, it is often insufficient for handling complex, multi-stage attacks, requiring more advanced methods.

Table 4. Feature analysis techniques and their applications in IDSs

Technique	Description	Tools/Methods	Application to IDS (MQTT Context)	Strengths	Limitations
Statistical Analysis [63, 64]	Uses statistical methods to describe data behaviour, test hypotheses, and detect unusual trends.	Hypothesis testing, regression analysis, probability distributions, confidence intervals	Detects anomalies in traffic flow, connection duration, or MQTT message size patterns.	Simple and computationally light.	Limited capability for complex/multi-stage MQTT attacks.
Dimensionality Reduction [65, 66]	Reduces the number of input features while preserving important information, enhancing efficiency.	PCA, LDA, t-SNE, UMAP, Autoencoders	Reduces redundancy in datasets like MQTTset or SEN-MQTTset, improving training efficiency.	Reduces overhead, faster IDS training.	May discard subtle MQTT-specific attack indicators.
Visualization-Based Analysis [67, 68]	Uses graphical techniques to identify hidden structures, trends, and anomalies.	Heatmaps, scatter plots, and graph-based visualizations	Visualizes relationships between QoS levels, publish rates, and detected anomalies.	Intuitive for human analysts.	Supportive only; not a direct detection tool.
Feature Importance Analysis [69, 70]	Ranks feature by influence on model predictions, improving transparency and efficiency.	SHAP, Permutation Importance, Gini Importance, LIME	Highlights MQTT-specific features (e.g., CONNECT Flags, ClientID, Packet Length) critical in attack detection.	Enhances interpretability, builds trust in IDS decisions.	Computationally expensive; explanations may be hard for non-experts.

5.2 Dimensionality reduction

Dimensionality reduction is an important process in ML that aims to transform large data into smaller data sets while retaining important information. This process offers benefits such as data compression, reduced storage requirements, and the elimination of redundant features, which improves computational efficiency and model accuracy [65, 66].

Dimensionality reduction techniques include traditional methods such as feature selection, as well as advanced methods such as PCA, LDA, and empirical pattern analysis (EMD). These techniques are widely used in fields such as audio processing, computer vision, and medical image diagnosis. In MQTT IDS, dimensionality reduction helps optimize model training efficiency in large-scale IoT networks [67, 68]. However, it risks discarding subtle interactions that may be relevant for attack detection.

5.3 Visualization-based analysis

Graph visualization-based analysis is a powerful tool for understanding complex data, helping users recognize patterns, detect anomalies, and understand complex relationships between data [71]. This approach leverages human visual perception to simplify the exploration of intricate relationships within large datasets, making it especially useful in fields such as software analysis, gameplay analytics, and cybersecurity.

For MQTT IDS, visualization can show correlations between QoS levels, traffic volume, and attack patterns, supporting analysts in understanding complex traffic. Graph visualization in cybersecurity has been used to review large-scale traffic data [72, 73]. These techniques help improve the understanding of IDSs, enabling security analysts to quickly identify suspicious patterns and take necessary measures to prevent attacks. However, to further refine intrusion detection capabilities, it is essential to quantify the influence of individual features through feature importance analysis.

5.4 Feature importance analysis

Feature importance analysis is a fundamental ML technique that helps determine the impact of each feature on model predictions. This analysis allows for improved model performance while ensuring transparency and ease of understanding, especially in sensitive fields such as healthcare and cybersecurity [69, 70].

In IDSs, feature importance analysis helps improve detection accuracy while reducing computational burden. By identifying the most influential features, the model can effectively distinguish between normal activity and cyberattacks, enhancing efficiency and reliability [74, 75].

Feature importance analysis techniques, including tools such as SHAP, Permutation Importance, and Gini Importance, are widely used to rank features based on their impact on prediction outcomes.

SHAP is one of the most advanced and reliable methods for interpreting ML model decisions.

SHAP is one of the most prominent tools used to interpret model predictions in IDSs. This technique provides a clear explanation of the impact of each feature on the model's results, both at the local level (for each individual case) and at the global level (to analyze the impact of features as a whole) [76, 77].

SHAP was developed based on cooperative game theory, giving it a strong mathematical foundation and an advantage over many other interpretation methods [78]. Comparative studies have shown that SHAP outperforms tools such as permutation importance and local independent model explanation (LIME) in providing clear and reliable interpretations [79].

For MQTT IDS, SHAP identified protocol-specific characteristics such as CONNECT Flags, ClientID, and Packet Length as prominent contributors for the purpose of distinguishing denial-of-service and brute force attacks. This strengthens model transparency, builds trust with security analysts, and supports the development of more effective countermeasures against MQTT-specific threats.

Although feature analysis methods significantly enhance the efficiency and effectiveness of MQTT-based IDS, there are limitations too. Statistical analysis provides simplicity but is incapable of identifying advanced or adaptive threats. Dimensionality reduction techniques like PCA are scalable but lose weak MQTT-specific signs of attacks. Visualization-oriented methods make it easy for analysts to comprehend the behaviour of the traffic, but can't be employed as single-coverage detection tools. Feature importance methods such as SHAP and LIME encourage transparency and trustworthiness, but they generally incur computational overhead and are difficult to interpret unless one is an expert. Therefore, future studies should focus on lightweight and interpretable techniques developed specifically for MQTT settings for a balance of accuracy, efficiency, and usability for practical applications for the IoT.

6. FUTURE DIRECTIONS

Many important trends in the future of MQTT intrusion detection research. More details can be found in Figure 2.

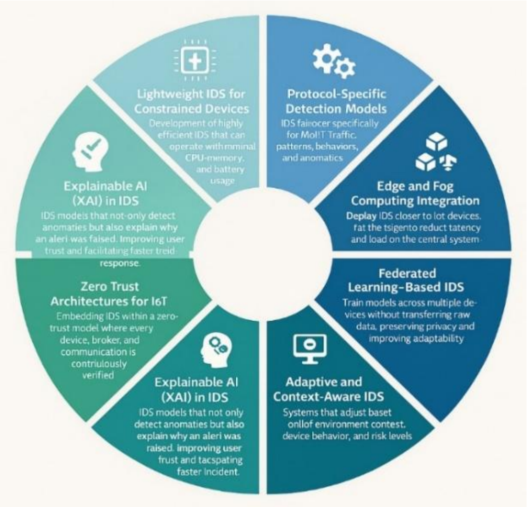
Lightweight IDS for Constrained Devices
Development of highly efficient IDS that can operate with mminal CPU-memory, and battery usage
Protocol-Specific Detection Models
IDS fairocer specifically for Mo!!T Traffic. patterns, behaviors, and anomatics
Edge and Fog Computing Integration
Deplay IDS closer to lot devices. fat the tsigento reduct tatency and load on the central system
Federated Learning–Based IDS
Train models across multiple de-vices without transferring raw data, preserving privacy and improving adaptability
Adaptive and Context-Aware IDS
Systems that adjust baset onllof environment contest. device behavior, and risk levels
Explainable AI (XAI) in IDS
IDS models that not only detect anomaties but also explain why an aleri was raised. improving user frust and tacspating faster Incident.
Zero Trust Architectures for I6T
Embedding IDS within a zero-trust model where every device, broker, and communication is contriulously verified
Explainable AI (XAI) in IDS
IDS models that not-only detect anomaties but also explain why an aleri was faised. Improving user trust and facilitating faster treid response.

Figure 2. Future research trends on MQTT IDS for IoT

Lightweight IDS for Resource-Constrained Devices: As most MQTT-based IoT nodes run on constrained CPU, memory, and battery, the development of lightweight IDS systems maintaining the accuracy of detection with minimal overhead is an essential direction of the research.

Protocol-Specific Detection Models: Traditional IDS typically don't register MQTT-specific abuse patterns like malformed CONNECT messages or abuse of the QoS. Future IDS should include MQTT semantics for increased accuracy of anomaly detection.

Combining Edge and Fog Computing: Deploying IDS closer to the IoT devices through the edge and fog nodes reduces latency, reduces bandwidth usage, and enables near real-time detection of MQTT intrusion without overloading the central servers.

Federated Learning-Based IDS: Federated IDS can simultaneously train models on dozens of IoT devices without sharing raw MQTT traffic. This provides improved data privacy and enables adaptive learning across heterogeneous deployments.

Adaptive and Context-Aware IDS: Future IDS must adapt dynamically to environmental situations, for example, device, level of QoS, or volume of traffic. This prevents IDS from adapting inadequately to changing threats within MQTT networks by dint of static thresholds for detection.

XAI for IDS: Explainability is increasingly needed for building analyst confidence in IDS decisions. SHAP, LIME, and future lightweight explainability tools can help analysts interpret alerts within MQTT configurations and respond more constructively.

ZTA for IoT Devices: Integrating IDS into a ZTA where everything, including MQTT connections and devices, is always authenticated provides higher resistance to unauthorized access, MitM, and spoofing.

7. CONCLUSIONS

As the scale of the IoT continues to increase, the security of MQTT communication has become an increasingly urgent problem. As the incidence and complexity of cyberattacks are on the increase, there is an exponential need for effective, precise, and extensible intrusion detection solutions. This work gave an exhaustive survey of IDS systems designed for MQTT contexts, considering the cases of supervised, unsupervised, and ensemble schemes for learning, and the most employed data and feature analysis techniques.

Threefold contributions of this work are as follows. First, we systematically compared IDS methodologies and observed that EL methods, especially RF and XGBoost, were consistently more accurate and robust than single-model classifiers. Secondly, we undertook a dataset-oriented analysis and demonstrated that MQTTset became the leading benchmark (applied in 65% of the studies), and SEN-MQTTset proposes MQTT-oriented features that refine detection granularity. Finally, we assessed feature analysis methodologies and observed that SHAP-based interpretability became a prominent tool for the ML-driven IDS interpretation, revealing computational and interpretability hurdles.

Future research directions should address some of the identified gaps. Developing richer and more diverse MQTT datasets is required for more accurately resembling real-world traffic and attack behaviours. Handling dataset imbalance and heterogeneity for IoT implementations remains an urgent need. IDS optimization for resource-limited devices remains the need for lightweight yet effective ML models. Furthermore, federated learning offers one promising direction for distributed, privacy-preserving IDS, and XAI must move toward lightweight, MQTT-aware architectures that make trade-offs between interpretability and efficiency. Through the integration of the literature available and conceptualization of the open problems, it offers practical and theoretical insights for the practitioner and the researcher alike. It highlights the necessity of IDS solutions that are not only precise but understandable, resource-aware, and adaptive within the MQTT-dominated IoT security context.

ACKNOWLEDGMENT

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at the University of Bisha for supporting this work through the Fast-Track Research Support Program.

REFERENCES

[1] Rose, K., Eldridge, S., Chapin, L. (2015). The Internet of Things: An overview. The Internet Society (ISOC), 80(15): 1-53. https://courses.sidnlabs.nl/ssi-2019/slides/lecture3b.pdf.

[2] Elhadi, S., Marzak, A., Sael, N., Merzouk, S. (2018). Comparative study of IoT protocols. Smart Application and Data Analysis for Smart Cities (SADASC'18). https://doi.org/10.2139/ssrn.3186315

[3] Ansari, D.B., Rehman, A.U., Ali, R. (2018). Internet of Things (IoT) protocols: A brief exploration of MQTT and CoAP. International Journal of Computer Applications, 179(27): 9-14. https://doi.org/10.5120/IJCA2018916438

[4] Yassein, M.B., Shatnawi, M.Q., Aljwarneh, S., Al-Hatmi, R. (2017). Internet of Things: Survey and open issues of MQTT protocol. In 2017 International Conference on Engineering & MIS (ICEMIS), Monastir, Tunisia, pp. 1-6. https://doi.org/10.1109/ICEMIS.2017.8273112

[5] Ashoor, A.S., Gore, S. (2011). Importance of intrusion detection system (IDS). International Journal of Scientific and Engineering Research, 2(1): 1-4. https://portal.arid.my/Publications/f3da7cd3-5bab-4294-94d1-6a22c1d4235d.pdf.

[6] Tidjon, L.N., Frappier, M., Mammar, A. (2019). Intrusion detection systems: A cross-domain overview. IEEE Communications Surveys & Tutorials, 21(4): 3639-3681. https://doi.org/10.1109/COMST.2019.2922584

[7] Jing, W., Peng, L., Fu, H., Hu, A. (2024). An authentication mechanism based on zero trust with radio frequency fingerprint for Internet of Things networks. IEEE Internet of Things Journal, 11(13): 23683-23698. https://doi.org/10.1109/JIOT.2024.3385989

[8] Polat, G., Sodah, F. (2019). Security issues in IoT: Challenges and countermeasures. ISACA Journal, 1-7. https://www.isaca.org/resources/isaca-journal/issues/2019/volume-1/security-issues-in-iot-challenges-and-countermeasures.

[9] Amrita, Ekwueme, C.P., Adam, I.H., Dwivedi, A. (2024). Lightweight cryptography for Internet of Things: A review. EAI Endorsed Transactions on Internet of Things, 10(1): 1-9. https://doi.org/10.4108/eetiot.5565

[10] Mehdipour, F. (2020). A review of IoT security challenges and solutions. In 2020 8th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC), Alexandria, Egypt, pp. 1-6. https://doi.org/10.1109/JAC-ECC51597.2020.9355854

[11] Pham, C.D.M., Nguyen, T.L.P., Dang, T.K. (2019). Resource-constrained IoT authentication protocol: An ECC-based hybrid scheme for device-to-server and device-to-device communications. In Future Data and Security Engineering, pp. 446-466. https://doi.org/10.1007/978-3-030-35653-8_30

[12] Alsoufi, M.A., Razak, S., Siraj, M.M., Nafea, I., et al. (2021). Anomaly-based intrusion detection systems in IoT using deep learning: A systematic literature review. Applied sciences, 11(18): 8383. https://doi.org/10.3390/app11188383

[13] Wei, N., Yin, L., Tan, J., Ruan, C., et al. (2023). An autoencoder-based hybrid detection model for intrusion detection with small-sample problem. IEEE Transactions on Network and Service Management, 21(2): 2402-2412. https://doi.org/10.1109/TNSM.2023.3334028

[14] Logeswari, G., Roselind, J.D., Tamilarasi, K., Nivethitha, V. (2025). A comprehensive approach to intrusion detection in IoT environments using hybrid feature selection and multi-stage classification techniques. IEEE Access, 13: 24970-24987. https://doi.org/10.1109/ACCESS.2025.3532895

[15] Alrayes, F.S., Zakariah, M., Amin, S.U., Khan, Z.I., Helal, M. (2024). Intrusion detection in IoT systems using denoising autoencoder. IEEE Access, 12: 122401-122425. https://doi.org/10.1109/ACCESS.2024.3451726

[16] Sana, L., Nazir, M.M., Yang, J., Hussain, L., et al. (2024). Securing the IoT cyber environment: Enhancing intrusion anomaly detection with vision transformers. IEEE Access, 12: 82443-82468. https://doi.org/10.1109/ACCESS.2024.3404778

[17] Ullah, S., Ahmad, J., Khan, M.A., Alshehri, M.S., et al. (2023). TNN-IDS: Transformer neural network-based intrusion detection system for MQTT-enabled IoT Networks. Computer Networks, 237: 110072. https://doi.org/10.1016/j.comnet.2023.110072

[18] Alaiz-Moreton, H., Aveleira-Mata, J., Ondicol-Garcia, J., Muñoz-Castañeda, A.L., et al. (2019). Multiclass classification procedure for detecting attacks on MQTT-IoT protocol. Complexity, 2019(1): 6516253. https://doi.org/10.1155/2019/6516253

[19] Hanif, A.A., Ilyas, M. (2024). Enhance the detection of DoS and brute force attacks within the MQTT environment through feature engineering and employing an ensemble technique. arXiv preprint arXiv:2408.00480. https://doi.org/10.48550/arXiv.2408.00480

[20] Mosaiyebzadeh, F., Rodriguez, L.G.A., Batista, D.M., Hirata, R. (2021). A network intrusion detection system using deep learning against MQTT attacks in IoT. In 2021 IEEE Latin-American Conference on Communications (LATINCOM), Santo Domingo, Dominican Republic, pp. 1-6. https://doi.org/10.1109/LATINCOM53176.2021.9647850

[21] Omotosho, A., Qendah, Y., Hammer, C. (2023). IDS-MA: Intrusion detection system for IoT MQTT attacks using centralized and federated learning. In 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC), Torino, Italy, pp. 678-688. https://doi.org/10.1109/COMPSAC57700.2023.00093

[22] Alasmari, R., Alhogail, A.A. (2024). Protecting smart-home IoT devices from MQTT attacks: An empirical study of ML-based IDS. IEEE Access, 12: 25993-26004. https://doi.org/10.1109/ACCESS.2024.3367113

[23] Siddharthan, H., Deepa, T., Chandhar, P. (2022). SENMQTT-SET: An intelligent intrusion detection in IoT-MQTT networks using ensemble multi cascade features. IEEE Access, 10: 33095-33110. https://doi.org/10.1109/ACCESS.2022.3161566

[24] Khan, M.A., Khan, M.A., Jan, S.U., Ahmad, J., et al. (2021). A deep learning-based intrusion detection system for MQTT enabled IoT. Sensors, 21(21): 7016. https://doi.org/10.3390/s21217016

[25] Boppana, T.K., Bagade, P. (2023). GAN-AE: An unsupervised intrusion detection system for MQTT

networks. Engineering Applications of Artificial Intelligence, 119: 105805. https://doi.org/10.1016/j.engappai.2022.105805

[26] Zeghida, H., Boulaiche, M., Chikh, R. (2023). Securing MQTT protocol for IoT environment using IDS based on ensemble learning. International Journal of Information Security, 22(4): 1075-1086. https://doi.org/10.1007/s10207-023-00681-3

[27] Rahmani, A.M., Yousefpoor, E., Yousefpoor, M.S., Mehmood, Z., et al. (2021). Machine learning (ML) in medicine: Review, applications, and challenges. Mathematics, 9(22): 2970. https://doi.org/10.3390/math9222970

[28] Suthishni, D.N.P., Kumar, K.S. (2022). A review on machine learning based security approaches in intrusion detection system. In 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 341-348. https://doi.org/10.23919/INDIACom54597.2022.9763261

[29] Mahdi, A.A. (2024). Machine learning applications of network security enhancement: Review. Computer Science & IT Research Journal, 5(10): 2283-2300. https://doi.org/10.51594/csitrj.v5i10.1635

[30] Liu, H., Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. Applied Sciences, 9(20): 4396. https://doi.org/10.3390/app9204396

[31] Saravanan, R., Sujatha, P. (2018). A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 945-949. https://doi.org/10.1109/iccons.2018.8663155

[32] Dubey, A., Ayubee, A.R., Singh, A.K., Singh, C., et al. (2023). A review: Most important and frequently used algorithms in supervised learning. In Proceedings of the KILBY 100 7th International Conference on Computing Sciences 2023 (ICCS 2023). https://doi.org/10.2139/ssrn.4482730

[33] Syed, I., Lokhande, V. (2024). An overview of the supervised machine learning. International Research Journal of Modernization in Engineering Technology and Science, 6(3): 6355-6360. https://doi.org/10.56726/irjmets51366

[34] Anthi, E., Williams, L., Słowińska, M., Theodorakopoulos, G., Burnap, P. (2019). A supervised intrusion detection system for smart home IoT devices. IEEE Internet of Things Journal, 6(5): 9042-9053. https://doi.org/10.1109/JIOT.2019.2926365

[35] Ashraf, W., Ahanger, A.S., Masoodi, F.S. (2024). Enhancing intrusion detection using supervised machine learning algorithms. In 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, pp. 1404-1408. https://doi.org/10.23919/INDIACom61295.2024.10498526

[36] Akintoye, K.A., Adu, M.K., Akinwamide, S. (2024). Network intrusion detection and classification system: A supervised machine learning approach. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 12(7): 657-670. https://doi.org/10.22214/ijraset.2024.63548

[37] Huang, Z., Li, Z.M., Zhang, J.M. (2023). Enhancing network security through machine learning: A study on intrusion detection system using supervised algorithms. Applied and Computational Engineering, 19: 50-66. https://doi.org/10.54254/2755-2721/19/20231008

[38] Farooq, M. (2022). Supervised learning techniques for intrusion detection system based on multi-layer classification approach. International Journal of Advanced Computer Science and Applications (IJACSA), 13(3): 311-315. https://doi.org/10.14569/ijacsa.2022.0130338

[39] Yazici, İ., Shayea, I., Din, J. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. Engineering Science and Technology an International Journal, 44: 101455. https://doi.org/10.1016/j.jestch.2023.101455

[40] Naeem, S., Ali, A., Anam, S., Ahmed, M.M. (2023). An unsupervised machine learning algorithms: Comprehensive review. International Journal of Computing and Digital Systems, 13(1): 911-921. https://doi.org/10.12785/ijcds/130172

[41] Oja, E. (2004). Finding clusters and components by unsupervised learning. In Structural, Syntactic, and Statistical Pattern Recognition, pp. 1-15. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-27868-9_1

[42] Bhadauria, S., Mohanty, T. (2021). Hybrid intrusion detection system using an unsupervised method for anomaly-based detection. In 2021 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Hyderabad, India, pp. 1-6. https://doi.org/10.1109/ANTS52808.2021.9936919

[43] Wang, Y., Sun, G., Cao, X., Yang, J. (2022). An intrusion detection system for the Internet of Things based on the ensemble of unsupervised techniques. Wireless Communications and Mobile Computing, 2022(1): 8614903. https://doi.org/10.1155/2022/8614903

[44] Jha, M., Acharya, R. (2016). An immune inspired unsupervised intrusion detection system for detection of novel attacks. In 2016 IEEE Conference on Intelligence and Security Informatics (ISI), Tucson, AZ, USA, pp. 292-297. https://doi.org/10.1109/ISI.2016.7745493

[45] Alom, M.Z., Taha, T.M. (2017). Network intrusion detection for cyber security using unsupervised deep learning approaches. In 2017 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, pp. 63-69. https://doi.org/10.1109/NAECON.2017.8268746

[46] Idrissi, I., Azizi, M., Moussaoui, O. (2022). An unsupervised generative adversarial network based-host intrusion detection system for Internet of Things devices. Indonesian Journal of Electrical Engineering and Computer Science, 25(2): 1140-1150. https://doi.org/10.11591/ijeecs.v25.i2.pp1140-1150

[47] Zhou, Z.H. (2012). Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC, New York. https://doi.org/10.1201/b12207

[48] Dietterich, T.G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine

Learning, 40: 139-157. https://doi.org/10.1023/A:1007607513941

[49] Opitz, D., Maclin, R. (1999). Popular ensemble methods: An empirical study. Journal of Artificial Intelligence Research, 11: 169-198. https://doi.org/10.1613/jair.614

[50] Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? The Journal of Machine Learning Research, 15(90): 3133-3181. http://jmlr.org/papers/v15/delgado14a.html.

[51] Breiman, L. (2001). Random forests. Machine Learning, 45: 5-32. https://doi.org/10.1023/A:1010933404324

[52] Schapire, R.E. (1990). The strength of weak learnability. Machine Learning, 5(2): 197-227. https://doi.org/10.1007/BF00116037

[53] Wolpert, D.H. (1992). Stacked generalization. Neural Networks, 5(2): 241-259. https://doi.org/10.1016/s0893-6080(05)80023-1

[54] Krawczyk, B., Minku, L.L., Gama, J., Stefanowski, J., Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. Information Fusion, 37: 132-156. https://doi.org/10.1016/j.inffus.2017.02.004

[55] Vujović, Ž. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6): 599-606. https://doi.org/10.14569/IJACSA.2021.0120670

[56] Hossin, M., Sulaiman, M.N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2): 1-11. https://doi.org/10.5121/ijdkp.2015.5201

[57] Hindy, H., Bayne, E., Bures, M., Atkinson, R., Tachtatzis, C., Bellekens, X. (2020). Machine learning based IoT intrusion detection system: An MQTT case study (MQTT-IoT-IDS2020 dataset). In Selected Papers from the 12th International Networking Conference, pp. 73-84. https://doi.org/10.1007/978-3-030-64758-2_6

[58] Stiawan, D., Wahyudi, D., Septian, T.W., Idris, M.Y., Budiarto, R. (2023). The development of an Internet of Things (IoT) network traffic dataset with simulated attack data. Journal of Internet Technology, 24(2): 345-356. https://doi.org/10.53106/160792642023032402013

[59] Hindy, H., Tachtatzis, C., Atkinson, R., Bayne, E., Bellekens, X. (2020). MQTT-IoT-IDS2020: MQTT Internet of Things intrusion detection dataset. IEEE Dataport. https://doi.org/10.21227/bhxy-ep04

[60] Vaccari, I., Chiola, G., Aiello, M., Mongelli, M., Cambiaso, E. (2020). MQTTset, a new dataset for machine learning techniques on MQTT. Sensors, 20(22): 6578. https://doi.org/10.3390/s20226578

[61] Cherrington, M., Thabtah, F., Lu, J., Xu, Q. (2019). Feature selection: Filter methods performance challenges. In 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, pp. 1-4. https://doi.org/10.1109/ICCISci.2019.8716478

[62] Rout, S., Mallick, R., Sahu, S.K. (2023). Exploring the significance of feature analysis in AI/ML modeling. In 2023 OITS International Conference on Information Technology (OCIT), Raipur, India, pp. 580-585. https://doi.org/10.1109/OCIT59427.2023.10431396

[63] Dally, J.W. (2008). Statistical analysis of experimental data. In Springer Handbook of Experimental Solid Mechanics, pp. 259-280. https://doi.org/10.1007/978-0-387-30877-7_11

[64] Efron, B., Tibshirani, R. (1991). Statistical data analysis in the computer age. Science, 253(5018): 390-395. https://doi.org/10.1126/science.253.5018.390

[65] Vlachos, M. (2010). Dimensionality reduction. In Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_216

[66] Mulla, F., Gupta, A. (2022). A review paper on dimensionality reduction techniques. Journal of Pharmaceutical Negative Results, 13(SO3): 1263-1272. https://doi.org/10.47750/pnr.2022.13.s03.198

[67] Velliangiri, S., Alagumuthukrishnan, S., Joseph, S.I.T. (2019). A review of dimensionality reduction techniques for efficient computation. Procedia Computer Science, 165: 104-111. https://doi.org/10.1016/j.procs.2020.01.079

[68] Sarveniazi, A. (2014). An actual survey of dimensionality reduction. American Journal of Computational Mathematics, 4(2): 55-72. https://doi.org/10.4236/ajcm.2014.42006

[69] Kelmendi, A. (2021). Feature importance for black-box models. Czech Technical University in Prague. https://dspace.cvut.cz/bitstream/handle/10467/92887/F8-BP-2021-Kelmendi-Ard-thesis.pdf?sequence=-1.

[70] Merrick, L. (2019). Randomized ablation feature importance. arXiv preprint arXiv:1910.00174. https://doi.org/10.48550/arXiv.1910.00174

[71] Langelier, G., Sahraoui, H., Poulin, P. (2005). Visualization-based analysis of quality for large-scale software systems. In Proceedings of the 20th IEEE/ACM International Conference on Automated Software Engineering, pp. 214-223. https://doi.org/10.1145/1101908.1101941

[72] Tatu, A., Albuquerque, G., Eisemann, M., Schneidewind, J., et al. (2009). Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In 2009 IEEE Symposium on Visual Analytics Science and Technology, Atlantic City, NJ, USA, pp. 59-66. https://doi.org/10.1109/vast.2009.5332628

[73] Ma, K.L. (2006). Cyber security through visualization. In Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation-Volume 60, Tokyo, Japan, pp. 3-7. https://dl.acm.org/doi/pdf/10.5555/1151903.1151904.

[74] Vos, G., van Eijk, L., Sarnyai, Z., Azghadi, M.R. (2024). Stabilizing machine learning for reproducible and explainable results: A novel validation approach to subject-specific insights. arXiv preprint arXiv:2412.16199. https://doi.org/10.48550/arxiv.2412.16199

[75] Sarhan, M., Layeghy, S., Portmann, M. (2021). Feature analysis for machine learning-based IoT intrusion detection. arXiv preprint arXiv:2108.12732. https://doi.org/10.48550/arXiv.2108.12732

[76] Wang, M., Zheng, K., Yang, Y., Wang, X. (2020). An explainable machine learning framework for intrusion detection systems. IEEE Access, 8: 73127-73141. https://doi.org/10.1109/access.2020.2988359

[77] Khediri, A., Slimi, H., Yahiaoui, A., Derdour, M., Bendjenna, H., Ghenai, C.E. (2024). Enhancing machine learning model interpretability in intrusion detection systems through SHAP explanations and LLM-generated

descriptions. In 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS), EL OUED, Algeria, pp. 1-6. https://doi.org/10.1109/PAIS62114.2024.10541168

[78] Lundberg, S.M., Lee, S. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874. https://doi.org/10.48550/arxiv.1705.07874

[79] Hariharan, S., Rejimol Robinson, R.R., Prasad, R.R., Thomas, C., Balakrishnan, N. (2023). XAI for intrusion detection system: Comparing explanations based on global and local scope. Journal of Computer Virology and Hacking Techniques, 19(2): 217-239. https://doi.org/10.1007/s11416-022-00441-2