



Article

Weibull Mixture Models with Context-Specific Outliers for IoT Intrusion Detection

Hassen Sallay

Department of Networks and Communications, College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia; hmsallay@iau.edu.sa

Abstract

We address here anomaly detection in the context of smart homes and the industrial internet of things (IIoT) by proposing a statistical framework based on Weibull mixture models (WMM) with context specific outlier handling. We model the normal traffic according to IoT device traffic characteristics. The outliers are modeled by uniform distributions for smart homes and Weibull distributions for the IIoT, reflecting their distinct attack profiles. We validated the proposed framework on both real and synthetic datasets for the specific context of the IIoT. WMM outperforms benchmark methods representing the common outlier detection approaches, achieving consistent robust ROC-AUC of 0.98 and Precision-AUC of 0.97. For its deployment in smart home and IIoT contexts, the framework provides adaptable architectural design to improve its efficiency implementation and management.

Keywords: IoT security; outlier detection; statistical framework; weibull mixture model; IIoT; architecture

1. Introduction

Steve Ranger said “The IoT bridges the gap between the digital world and the physical world”. This bridging has transformed domestic and industrial environments. More than 55 billion IoT devices will be installed globally by 2035, with smart home devices constituting a significant portion [1]. This technology shift will enhance automation, energy efficiency, and connectivity, but also will open the door to vulnerabilities and security intrusions; as Mikko Hypponen said “If it’s smart, it’s vulnerable.”. Indeed, the attack surface for malicious actors to exploit continuously has expanded [2–5]. Moreover, manufacturers are prioritizing functionality over security and customers are unaware of the security risk and vulnerabilities. The mitigation of such security risks requires in-depth context awareness and defense strategies to be applied efficiently. Intrusion Detection Systems (IDSs) are a critical component of this layered defense, inspecting network and device traffic in real time to detect and alert potential security intrusions. Anomaly-based IDSs, in particular, monitor traffic behavior against established normal or attack profiles, identifying deviations or matches to generate security alerts.

From literature, ref. [6] surveys machine learning-based intrusion detection systems for IoT, covering research efforts from 2019 to 2024 across various domains such as medical, agricultural, industrial, and smart home networks. The study classifies machine learning-based IDSs and discusses their applications in diverse IoT environments. While powerful, machine learning and deep learning models require computational resources and may lack interpretability, posing challenges for resource-constrained IoT devices. Ref. [7] conducted a comprehensive survey of intrusion detection systems (IDSs) designed for IoT-based smart



Academic Editors: Olusola Tolulope
Odeyomi and Temitayo Olowu

Received: 6 May 2026

Revised: 23 July 2026

Accepted: 27 July 2026

Published: 4 August 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

environments, emphasizing the constraints of IoT devices, such as limited computing and storage capabilities. The survey reviews various IDSs and highlights the need for lightweight and efficient IDSs to mitigate IoT security threats, such as denial-of-service (DoS) attacks and unauthorized access. Among 109 studied references, we note a lack of references focusing on statistical frameworks using statistical modeling for smart homes and the IIoT. Ref. [8] provided a survey of anomaly detection methods and applications in IoT, analyzing 64 papers published between January 2019 and July 2021. The study categorizes anomaly detection algorithms, including statistical, machine learning, and hybrid approaches, and discusses the strengths and limitations of various detection methods. While this survey includes statistical methods, it does not specifically highlight parametric finite mixture models except in one reference that used a Gaussian mixture, which does not fit the bursty nature of the IoT traffic for several IoT devices. Furthermore, it lacks a focus on context-specific modeling for distinct IoT environments like smart homes and the IIoT.

Characterizing IoT devices in terms of traffic and attack types is required to develop appropriate models. Refs. [9,10] analyze smart home and smart city traffic to characterize usage patterns. They provide insights into IoT traffic characteristics. Devices exhibit periodic behavior with significant idle time, generate low traffic volumes, and use protocols like MQTT, CoAP, and HTTP(S), with security and privacy concerns.

Within smart homes and the IIoT context, a deep study of the devices, their traffic, and attack types is needed to design a statistically robust intrusion model. To do this, we have examined the similarities and differences in devices, network traffic, and security threats between the two environments, proposed a general taxonomy for IoT devices, and evaluated the suitability of statistical distributions for modeling normal traffic and attack-related outliers. This constitutes our first contribution of the paper. The second contribution is based on this analysis and a statistical framework for anomaly and intrusion detection in IoT smart homes and the IIoT was proposed. Mainly, for a flexible, scalable, generic and adaptable solution, we articulated the framework using tailored parametric finite mixture modeling to account for the diverse traffic behaviors such as periodic, event-driven, streaming, and control devices. More specifically, a Weibull mixture model (WMM) is effective for characterizing normal traffic in both contexts. Anomalies are modeled with a uniform distribution for outliers in smart homes and a Weibull distribution for the IIoT, aligning with the distinct threat landscapes of each environment. Third, we validate the proposed model using real and synthetic datasets to provide the usefulness of the framework and measure its performance compared to established outlier detection methods and strategies. Finally, we propose architectural insights for the implementation and deployment of the proposed framework.

The remainder of the paper is organized as follows: Section 2 presents the traffic and attack taxonomy and modeling in smart homes and the industrial IoT. Section 3 presents the mathematical formulation of the statistical framework. Section 4 validates the framework for the specific context Industrial IoT and discusses the benchmarking results against some common outlier detection models. Section 5 states architectural variants for the framework deployment, and Section 6 concludes the paper and highlights our future research directions.

2. Traffic and Attack Modeling in Smart Homes and the Industrial IoT

2.1. Similarities and Differences Between Smart Homes and the IIoT

The smart home IoT and IIoT share several characteristics. Both of them have heterogeneous devices with specific functionalities. They are using heterogeneous communication protocols; smart homes utilize protocols like HTTP(S), MQTT, CoAP, and Zigbee to manage device communication through gateways or routers, and IIoT systems use industrial proto-

cols such as Modbus, Profinet, Ethernet/IP, and MQTT, often integrated with gateways for data aggregation and cloud connectivity. This diversity exposes both environments to similar vulnerabilities. Moreover, the threat landscapes of the smart home IoT and the IIoT overlap significantly. They are facing denial of service (DoS), distributed denial of service (DDoS), unauthorized access, malware, and data breaches. Attack indicators are applicable across both contexts. Indicators such as increased traffic volume, unusual ports or protocols, device malfunctions, failed login attempts, and malware signatures, are common in both contexts.

However, they differ on attack criticality and impact. Smart home device failures cause inconvenience rather than catastrophic outcomes. In contrast, IIoT devices support critical industrial processes exposed to financial losses, physical damage, or safety hazards. On the other hand, human activities induce variable, unpredictable patterns in smart homes. User movements and the attacks are often random and unstructured compromising user privacy or disrupting services. In contrast, IIoT traffic is more structured and periodic due to industrial processes automation. Therefore, the attacks are frequently targeted and sophisticated, aiming to manipulate those processes or steal proprietary data.

2.2. Taxonomy of IoT Devices

We aim to guide our statistical modeling in the context of smart homes and the IIoT. For this sake, we have proposed a taxonomy of IoT devices (Table 1) based on their network traffic characteristics. We can distinguish that periodic reporting devices such as smart thermostats and temperature sensors generate consistent and low-volume traffic, which is ideal for modeling with distributions like Weibull. Conversely, event-driven devices like motion or vibration sensors are producing sporadic traffic requiring flexible models to capture heavy-tailed patterns. In addition, streaming devices such as cameras generate high-volume and a continuous data, while control systems exhibit mixed patterns necessitating adapted modeling approaches.

Table 1. Taxonomy of IoT Devices.

Category	Description	Examples (Smart Home)	Examples (IIoT)	Traffic Patterns
Periodic Reporting Devices	Send data at regular intervals	Smart Thermostat, Smart Meter	Temperature Sensor, Smart Meter	Small packets, periodic timing, low volume
Event-Driven Devices	Transmit data based on specific events	Motion Sensor, Smart Lock	Vibration Sensor, Actuator	Small packets, sporadic timing, low volume
Streaming Devices	Continuously stream data	IP Camera, Smart Speaker	Surveillance Camera	Bimodal packet sizes, continuous, high volume
Control Systems	Manage and control other devices/processes	Smart Hub, Voice Assistant	PLC, SCADA System	Variable packet sizes, mixed timing, moderate volume

2.3. Statistical Modeling of Normal Traffic and Attacks

Statistical modeling is pivotal for effective anomaly and intrusion detection in IoT environments. Mainly, we require distributions that accurately represent normal traffic and distinguish outliers [11]. Prior work on anomaly detection [12,13] gave us an insight to employ mixture models with outlier components. In the following sections, we examine Weibull mixture models for normal traffic with outlier following uniform distribution for smart homes and Weibull for the IIoT context.

2.3.1. Modeling Normal Traffic

The Weibull distribution is a very flexible modeling distribution given its shape and scale parameters. It can simulate an exponential-like regular intervals to heavy-tailed

bursty patterns. In both smart home IoT and the IIoT, a mixture of Weibull distributions seems to be an effective choice for modeling heterogeneous device behaviors. Indeed, smart home devices generate a mix of periodic and event-driven traffic according to user activities. A Weibull mixture model can capture these diverse patterns by combining components with different shape and scale parameters. So it is beneficial when representing the multimodal nature of devices like smart speakers (bursty during use) and smart plugs (periodic status updates). On the other hand, IIoT traffic is more structured and driven by automated industrial processes. For instance, suppose we are in the presence of temperature sensors and PLCs. The temperature sensors will transmit data at fixed intervals, whereas PLCs poll sensors regularly, with variability introduced by operational factors like equipment wear or network delays. We want to capture this periodicity of sensor data and the variability of control system interactions efficiently. For that, we believe in Weibull mixture modeling. By fitting multiple Weibull components, the model can represent the multimodal distributions observed in smart home devices (e.g., periodic vs. bursty) and the structured yet variable patterns in IIoT systems. For example, a smart meter reporting energy usage every 15 min may follow an exponential-like Weibull distribution, while a vibration sensor's sporadic alerts align with a heavy-tailed Weibull component. So, this flexibility enhances the model's ability to build a baseline for normal behavior that will be used for detecting anomalies.

In the literature, ref. [14] demonstrates that Weibull distributions effectively model interarrival times on general Internet traffic. More specifically, ref. [15] analyzes real-world smart home traffic, focusing on volume, temporal patterns, and external endpoints. Although it does not explicitly reference Weibull distributions, the study identifies patterns which align with the characteristics of interarrival times that Weibull models can capture. The presence of “heartbeat” traffic, characterized by regular and irregular components, is supporting the potential use of Weibull mixtures to represent these patterns. Ref. [16] characterizing traffic properties, such as interarrival times, indicates that statistical models like Weibull are worth exploring. In industrial the IIoT context, ref. [17] identifies the benefits of the Weibull distribution among heavy-tailed models, such as Pareto and lognormal, in industrial settings, especially where traffic may include both regular updates and bursty communications. Although the paper does not explicitly discuss mixture models, regarding the concept of multiple failure modes in reliability analysis, ref. [18] provides a theoretical basis for extending Weibull mixtures to traffic modeling due to the analogy between device failures and traffic events. Furthermore [19] demonstrated the efficacy of Weibull mixtures in survival analysis. Indeed, we can treat interarrival times in IoT traffic as time-to-event data, with right-censoring applicable in scenarios involving incomplete observation periods. This theoretical alignment enhances the case for Weibull mixtures in IoT applications such as smart homes and the IIoT.

2.3.2. Modeling Attack Outliers

Smart Home: Uniform Outlier Distribution

As we know, the uniform distribution assumes all values within a specified range are equally likely. It is capturing the randomness of attacks without requiring assumptions about their structure. In smart homes, many cyberattacks generate random and unstructured traffic that deviates significantly from normal patterns. This attack traffic nature makes the uniform distribution an effective model for outliers. For instance, in the DDOS flooding attack, devices like hubs or cameras are flooded with packets at random intervals. The inter-arrival times of these packets are uniformly distributed across a wide range, reflecting the high-volume, unstructured nature of the attack. Port scanning is another example. As attackers test different ports, the generated traffic is fitting a uniform

distribution due to their exploratory nature. As a third example, suppose the attacker has compromised a smart speaker; this later will send erratic data producing interarrival times that can be detected as uniform outliers.

While structured attacks, such as malware sending periodic malicious commands, may occur, the prevalence of random attacks in smart homes justifies prioritizing a uniform outlier model. Extensions to include Weibull or heavy-tailed components for more complex scenarios are also possible.

The IIoT: Weibull Outlier Distribution

In the IIoT, security threats such as targeted malware, unauthorized control commands and data exfiltration, often involve sophisticated attack processes. They may mimic normal traffic patterns to evade detection which necessitates a more nuanced outlier model. For instance, a malware which alters PLC logic to sabotage equipment, may generate periodic but abnormal traffic, such as manipulated sensor readings or control commands. A second example of an attacker producing a traffic with structured timing by sending fraudulent commands to actuators. Consider also covert data theft from SCADA systems by involving periodic, low-volume traffic designed to blend with normal operations. All of these attack examples require a Weibull outlier model to identify these subtle deviations.

While random attacks like DDoS also occur in the IIoT and could be modeled with a uniform distribution, the prevalence of targeted threats justifies prioritizing a Weibull outlier model. This approach enhances detection precision, particularly for devices with structured traffic, such as PLCs or SCADA systems, where anomalies may closely resemble normal patterns.

We conclude that the smart home IoT and the IIoT have shared and distinct features that inform the design of effective IDS. While Weibull mixture models for normal traffic can capture diverse patterns in both contexts, the outlier distribution should be tailored to the specific threat landscape of each one. The uniform is appropriate for detecting random attacks in smart homes, while the Weibull is more suitable to the structured threats in the IIoT.

3. Mathematical Formulation of the Model

3.1. Weibull Mixture Model with Uniform Outlier Component

We assume that normal IoT traffic follows a mixture of K Weibull distributions, while outliers (intrusions) are modeled by a uniform distribution. Let $X = (x_1, \dots, x_N)^T$ represent a set of N network flows, where each flow $x_i = (x_{i1}, \dots, x_{iD})$ is a D -dimensional feature vector characterizing communication and data exchange (e.g., packet size, inter-arrival time).

The probability density function (PDF) of the model is a convex linear combination of K Weibull components and one uniform component, expressed as

$$p(x_i | \Theta_K) = \sum_{k=1}^K \pi_k p(x_i | \theta_k) + \pi_{K+1} U(x_i), \quad (1)$$

where π_k (for $k = 1, \dots, K$) are the mixing proportions for the Weibull components, with $\pi_k > 0$ and $\sum_{k=1}^{K+1} \pi_k = 1$. $\pi_{K+1} = 1 - \sum_{k=1}^K \pi_k$ is the mixing proportion for the uniform component. $\Theta_K = (\theta_1, \dots, \theta_K)$ is the set of parameters for the Weibull components, where each $\theta_k = (\alpha_k, \beta_k)$ includes the scale (α_k) and shape (β_k) parameters. $p(x_i | \theta_k)$ is the Weibull PDF for component k , defined as

$$p(x_i | \theta_k) = \frac{\beta_k}{\alpha_k} \left(\frac{x_i}{\alpha_k} \right)^{\beta_k-1} e^{-\left(\frac{x_i}{\alpha_k} \right)^{\beta_k}}, \quad (2)$$

$-U(x_i)$ is the averaged density over the data range (heuristic uniform component), defined as:

$$U(x_i) = \frac{1}{N} \sum_{i=1}^N p(x_i | \hat{\Theta}), \quad (3)$$

where $\hat{\Theta}$ represents the estimated parameters of the model. The uniform component is averaged over all observations to capture outliers.

Expanding Equation (1), the full model is

$$p(x_i | \Theta_K) = \sum_{k=1}^K \pi_k \left[\frac{\beta_k}{\alpha_k} \left(\frac{x_i}{\alpha_k} \right)^{\beta_k-1} e^{-\left(\frac{x_i}{\alpha_k} \right)^{\beta_k}} \right] + \pi_{K+1} U(x_i). \quad (4)$$

3.1.1. Parameter Estimation

We estimate parameters through maximum likelihood estimation (MLE) via the Expectation-Maximization (EM) algorithm. The complete log-likelihood, incorporating latent membership vectors $z_i = (z_{i1}, \dots, z_{iK+1})$, where $z_{ik} = 1$ if x_i belongs to component k and $z_{ik} = 0$ otherwise, is

$$\log p(X, z | \Theta) = \sum_{i=1}^N \sum_{k=1}^{K+1} z_{ik} \log[\pi_k p(x_i | \theta_k)]. \quad (5)$$

The EM algorithm iterates an expectation and maximization steps. The Expectation (E) step computes posterior probabilities at iteration q :

$$p(z_{ik} | x_i, \theta_k^{(q)}) = \frac{\pi_k^{(q-1)} p(x_i | \theta_k^{(q-1)})}{\sum_{k=1}^K \pi_k^{(q-1)} p(x_i | \theta_k^{(q-1)}) + \pi_{K+1}^{(q-1)} U(x_i)}, \quad (6)$$

for $k = 1, 2, \dots, K$, and for the uniform component

$$p(z_{iK+1} | x_i, \theta_{K+1}^{(q)}) = \frac{\pi_{K+1}^{(q-1)} U(x_i)}{\sum_{k=1}^K \pi_k^{(q-1)} p(x_i | \theta_k^{(q-1)}) + \pi_{K+1}^{(q-1)} U(x_i)}. \quad (7)$$

The Maximization (M) step updates parameters to maximize the complete log-likelihood:

$$\pi_k^{(q)} = \frac{1}{N} \sum_{i=1}^N p(z_{ik} | x_i, \theta_k^{(q-1)}), \quad (8)$$

$$\alpha_k^{(q)} = \left(\frac{\sum_{i=1}^N p(z_{ik} | x_i, \theta_k^{(q-1)}) x_i^{\beta_k^{(q-1)}}}{\sum_{i=1}^N p(z_{ik} | x_i, \theta_k^{(q-1)})} \right)^{\frac{1}{\beta_k^{(q-1)}}}. \quad (9)$$

The shape parameter β_k is updated using the Newton–Raphson method:

$$\beta_k^{(q)} \approx \beta_k^{(q-1)} - \frac{\partial Q(\Theta | \Theta^{(q-1)})}{\partial \beta_k^{(q-1)}} \left(\frac{\partial^2 Q(\Theta | \Theta^{(q-1)})}{\partial \beta_k^{(q-1)2}} \right)^{-1}, \quad (10)$$

where $Q(\Theta | \Theta^{(q-1)})$ is the expected complete log-likelihood.

3.1.2. Model Selection

The number of components K is determined using the Mixture Minimum Description Length (MMDL) criterion:

$$\text{MMDL}(K) = \frac{n_{cp}}{2} \sum_{k=1}^K \log(\pi_k) + \frac{1}{2} N_{fp} \log(N) - \log(p(X | \theta)), \quad (11)$$

where n_{cp} is the number of parameters per component, and N_{fp} is the total number of free parameters.

3.2. Weibull Mixture Model with Weibull Outlier Component

An alternative model replaces the uniform outlier component with a Weibull distribution to capture structured attack patterns. The PDF is

$$p(x_i | \Theta_K) = \sum_{k=1}^K \pi_k \left[\frac{\beta_k}{\alpha_k} \left(\frac{x_i}{\alpha_k} \right)^{\beta_k-1} e^{-\left(\frac{x_i}{\alpha_k} \right)^{\beta_k}} \right] + \pi_{K+1} \left[\frac{\beta_{K+1}}{\alpha_{K+1}} \left(\frac{x_i}{\alpha_{K+1}} \right)^{\beta_{K+1}-1} e^{-\left(\frac{x_i}{\alpha_{K+1}} \right)^{\beta_{K+1}}} \right], \quad (12)$$

where β_{K+1} and α_{K+1} are the shape and scale parameters of the outlier Weibull component, and other terms are as defined in Equation (1).

Parameter Estimation

The log-likelihood for the Weibull outlier model is

$$\log p(X, z | \Theta) = \sum_{i=1}^N \sum_{k=1}^{K+1} z_{ik} \log[\pi_k p(x_i | \theta_k)], \quad (13)$$

where $p(x_i | \theta_{K+1})$ is the Weibull PDF for the outlier component, as in Equation (2). The EM algorithm proceeds similarly, with posterior probabilities

$$p(z_{ik} | x_i, \theta_k^{(q)}) = \frac{\pi_k^{(q-1)} p(x_i | \theta_k^{(q-1)})}{\sum_{k=1}^{K+1} \pi_k^{(q-1)} p(x_i | \theta_k^{(q-1)})}, \quad (14)$$

for $k = 1, 2, \dots, K + 1$. Parameter updates follow Equations (8)–(10), extended to include the outlier Weibull component's parameters β_{K+1} and α_{K+1} .

The MMDL criterion (Equation (11)) is used to select K , with adjusted n_{cp} and N_{fp} to account for the additional Weibull parameters.

4. Model Validation for the IIoT

In this part, we validate the proposed models against real and synthetic datasets. We focus mainly on the IIoT. We led a principled comprehensive performance study including multiple baseline models for benchmarking, multiple runs of experiments, sensitivity analysis, and an ablation study for the model components. Furthermore, we tested perturbation with different kinds of noises to study the robustness of the proposed model and its adaptability to real deployment scenarios. In the following, we present both datasets used to validate the proposed model, the baseline benchmarking models used in the performance study, the implementation of the framework, and the model validation results.

4.1. Benchmarking Datasets

4.1.1. Real Dataset

For model validation and benchmarking, we are faced with a lack of datasets that follow the Weibull distributions. In order to choose a suitable real dataset that complies with the proposed model, we studied the goodness of fit of common widely used and

publicly available IoT, IIoT, and smart home IoT datasets [20,21] using the fit R package. Unfortunately, they did not fit the Weibull mixture model. This is not surprising since most datasets were generated experimentally with a limited and preconditioned scope, whereas the datasets from the industry and telecommunications companies are not made public.

Recently, a new dataset called DataSense specific for the IIoT was published [22]. DataSense is a dataset with realistic attack scenarios for IIoT systems. The dataset integrates multi-source data from interconnected devices and network equipment and retains the most discriminative features to balance detection accuracy and computational efficiency, making it particularly suitable for real-time IIoT security applications. The dataset includes several categorical and numeric features with their specific statistical measures. We assessed the fitness of numeric features to Gaussian, Weibull, Log-Normal, Gamma, and Exponential distributions using the Kolmogorov–Smirnov (KS) statistic. In contrast to the previous cited datasets, the DataSense dataset shows Weibull behavior in some of its features and also heavy tailed skewed distributions such as the lognormal distribution that can be approached by Weibull since there are some similarities in the shape that they describe.

Figure 1a shows which distributions most frequently had the best fit for individual features providing a high-level view of the distributional characteristics of the dataset. We note the dominance of Gaussian and lognormal distributions with the presence of Weibull behavior of some features. Figure 1b shows the KS statistic box plot visualizing the range and central tendency of KS statistics across distributions to compare the fit quality and consistency. Note that a lower KS statistic indicates a better fit. Figure 1c shows the correlation heatmap revealing correlations between the KS statistics of different distributions. We note that there is a strong positive correlation (approximately 0.72) between the KS statistics for the lognormal and Weibull distributions. This suggests that features fitting one of these distributions well often fit the other relatively well too, implying similarities in the data shapes they describe. Therefore, we used this dataset in our benchmarking as a real benchmarking dataset for the IIoT context even though most of the features did not fit to Weibull distribution.

4.1.2. Synthetic Dataset for the IIoT

To validate thoroughly the proposed Weibull Mixture Model for anomaly detection in industrial IoT environments, we built a synthetic dataset. The synthetic dataset includes common categorical features and numeric features for benign traffic including packet count, flow duration, inter-arrival time, packet size statistics, to name a few. The majority of numeric features follow a Weibull distribution, but it includes also some lognormal/Gaussian features to approximate the real-world scenarios where feature distributions are more heterogeneous. Figure 2a shows the KS statistic with p -values of 12 selected numeric features. We note that several features have p -values above the 0.05 threshold, meaning the Weibull hypothesis is not rejected for those features. However, some features fall below the significance threshold, indicating that not every feature is truly Weibull-distributed. By such design of the synthetic dataset, we want to experiment with the WMM against a dataset with enough features which are sufficiently positive, skewed, and Weibull-like rather than perfectly Weibull. On the other hand, for the anomaly construction in the context of the IIoT, we include feature-swap anomalies, followed by extreme, noise, and uniform anomalies (see Figure 2c). The main idea is that the benchmark should include multiple perturbation mechanisms in which some alter marginal distributions, some inject extreme values, some add random noise, and some disturb feature consistency. The two-sample KS validation further confirms that synthetic anomalies are statistically distinguishable from benign samples across features (see Figure 2b). The KS statistics are consistently high, indicating that the synthetic anomalies are not merely random label assignments but represent measurable

distributional shifts. This strengthens the validity of the synthetic benchmark, although it also means that the synthetic task may be cleaner and more separable than real-world intrusion detection.

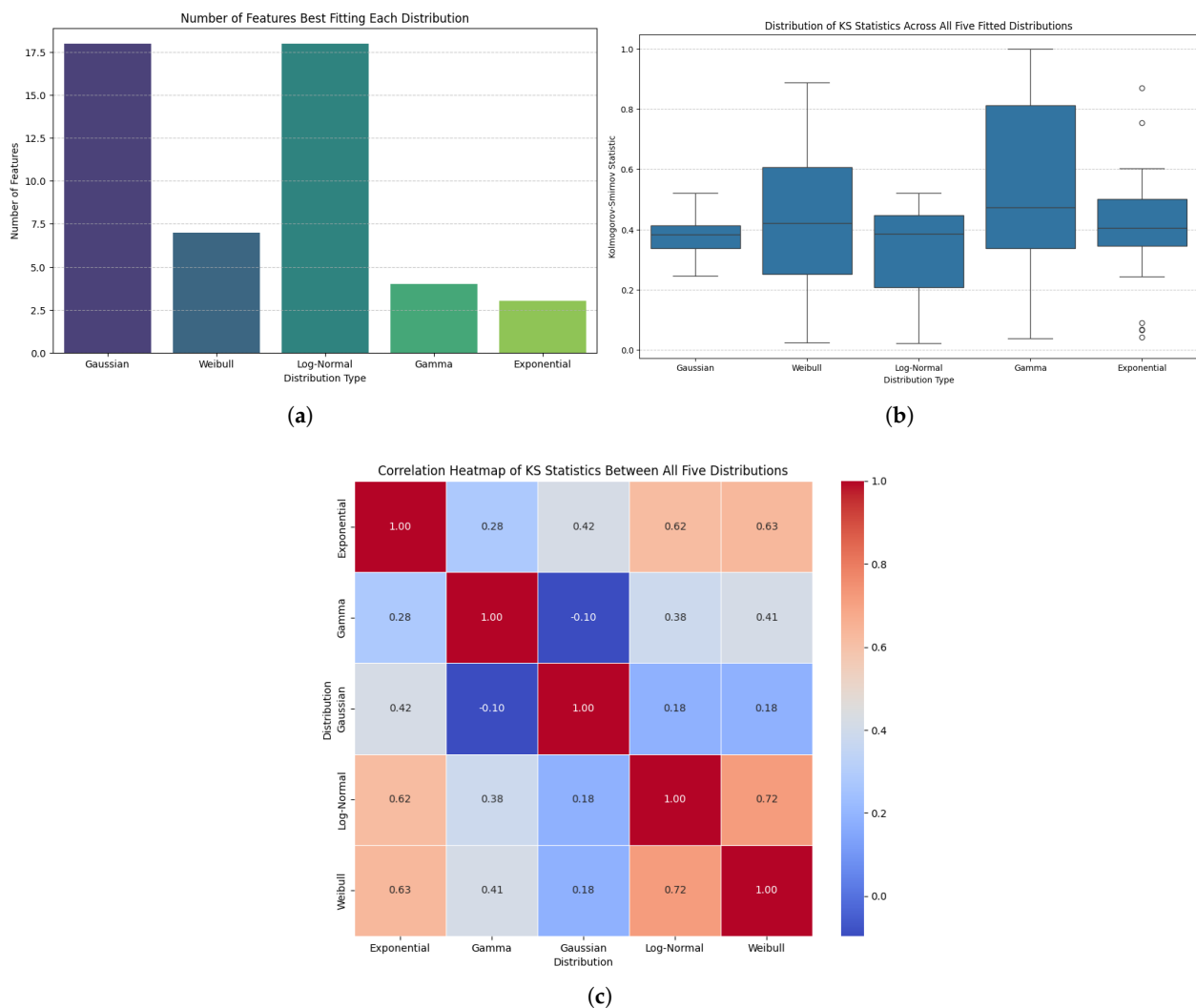


Figure 1. (a) Number of features best fitting distribution based on Kolmogorov–Smirnov test. (b) Distribution of KS statistics across all distributions. (c) Correlation heatmap of KS statistics.

4.2. Benchmark Baseline Models

The purpose of benchmarking the Weibull Mixture Model (WMM) is not to exhaustively compare against every possible anomaly detector but to validate whether WMM is competitive against representative models from the main families of anomaly detection such as histogram and marginal modeling, local-neighborhood deviation, boundary learning, and reconstruction error techniques. The main idea is that the benchmark should include methods that challenge the WMM from different perspectives. Therefore, we selected the common baseline models that cover these classes: HBOS (Histogram-Based Outlier Score) is a marginal density model that tests whether simple independent feature-density modeling is enough or not. It is based on the assumption that anomalies have low probability density. LOF (Local Outlier Factor) is a local neighborhood method that detects anomalies that lie in sparse local regions. It is based on the assumption that anomalies lie in sparse local neighborhoods. One-Class SVM (Support Vector Machine) is a boundary method that supports the boundary around the normal data Kernel and tests whether a flexible nonlinear decision boundary outperforms the WMM or not. It is based on the

assumption that anomalies fall outside the normal decision boundary [23]. PCA (Principal Component Analysis) and the autoencoder are based on the assumption that anomalies are difficult to reconstruct. PCA is a linear subspace method that tests whether anomalies are deviations from a low-dimensional normal manifold. The autoencoder challenges the WMM using non-linear reconstruction-based detection. It is a deep learning method that tests whether learned nonlinear reconstruction improves detection.

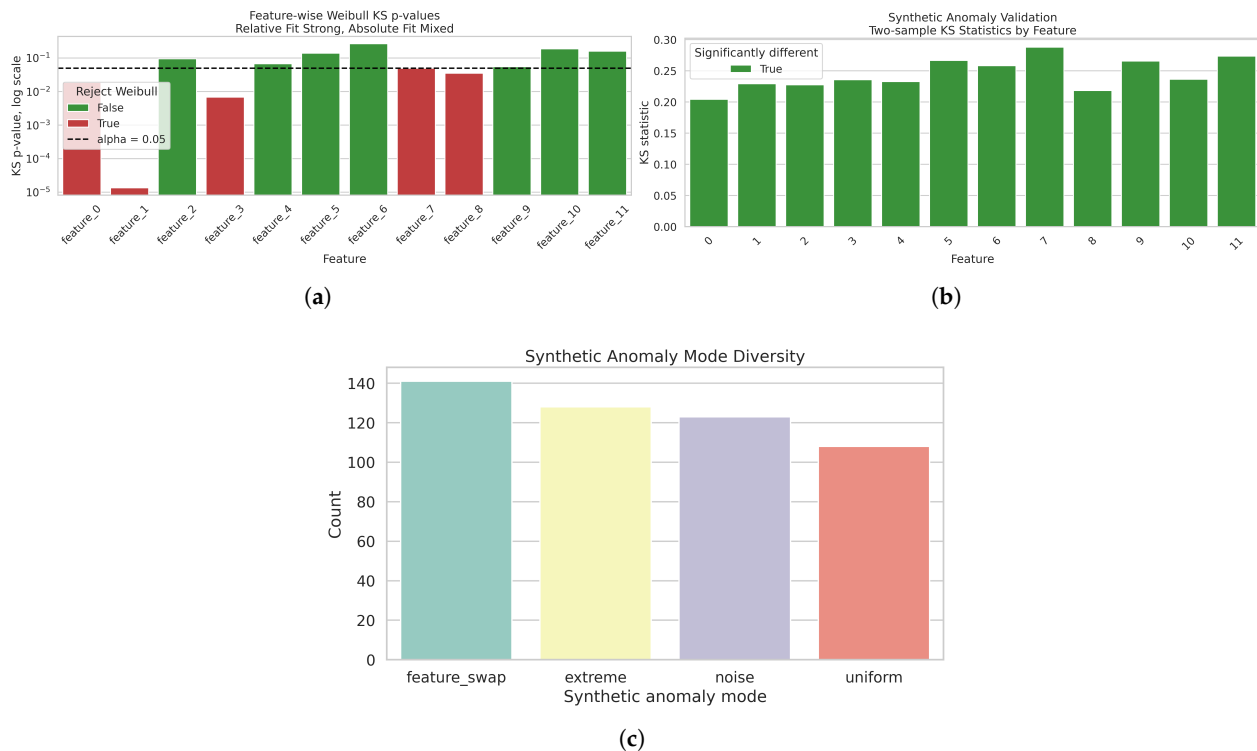


Figure 2. (a) Feature-wise Kolmogorov–Smirnov test results for distribution fit with $p = 0.05$. (b) Two samples of KS per feature. (c) Anomaly modes.

Many additional ML or DL anomaly detectors such as deep, variational, and LSTM autoencoder are variants of the same core ideas. Moreover, given the fact that the WMM operates on extracted IoT/IloT traffic features which are tabular features, we include an autoencoder, which is the common deep anomaly-detection baseline for tabular data not raw packets, images, or long temporal sequences where deep models are most useful. In addition to that, we are more interested in explainable models that are clear to interpret and easy to deploy and which are very important to identify and understand the intrusion and anomaly root causes for the IloT setting.

4.3. Implementation and Performance Metrics

The model was implemented using numerical optimization (L-BFGS-B) for estimating the maximum likelihood (MLE), with K-means initialization for the EM algorithm. For that, we used Python packages 3.11 such as numpy for the vectorized operations for the EM algorithm's E-step, M-step, and log-likelihood calculations, as well as computing MMDL's log-likelihood and penalty terms. SciPy supports dataset generation with Weibull and uniform distributions, provides PDF calculations and beta optimization for the EM algorithm, and contributes to MMDL's likelihood term. Moreover, the evaluation includes repeated trials (30 runs), confidence intervals (95%), statistical tests, ablation studies, robustness analysis, and anomaly-ratio sensitivity. This provides sufficient evidence to validate the proposed WMM against the main competing anomaly-detection principles. Scikit-learn 1.9.0 facilitates K-means initialization for both dataset generation and the EM algorithm's

E-step, and provides baseline models and metrics for validation. Finally Pandas organizes validation results into DataFrames, while Matplotlib 3.11.1 and Seaborn 0.13.2 visualize performance metrics with plots. Among the main metrics are

- Precision: measures the fraction of raised alarms that are truly attacks;
- Recall: measures the fraction of attacks detected;
- Specificity: measures the fraction of normal traffic correctly accepted;
- F1: measures the balance between precision and recall;
- PR-AUC: measures the precision–recall quality, especially important under imbalanced datasets;
- ROC-AUC: measures the global ranking ability between normal and attack samples;
- MCC: measures the balanced correlation between prediction and truth.

4.4. Results for the Real Dataset

Figure 3a presents a heatmap plot showing the main results for models benchmarking with several baseline models and performance metrics. We observe that, on the real IIoT dataset, the WMM achieves the highest recall and F1-score, indicating strong sensitivity to intrusive behavior. Indeed, the Weibull Mixture Model demonstrates an excellent ability to detect a large proportion of attack samples, with an attack recall of approximately 96.1%. This means that the WMM is very effective at flagging abnormal traffic and rarely misses malicious events. From an IIoT cybersecurity perspective, this is valuable because false negatives are often more dangerous than false positives since an undetected intrusion may lead to compromise, data leakage, or service disruption.



Figure 3. (a) Heatmap plot for main metrics. (b) Specificity across models. (c) MCC of benchmark models.

However, the real dataset results also expose a major limitation of the WMM, which is poor benign traffic recognition. Figure 3b reported specificity of around 23.9%, which indicates that many benign samples are incorrectly classified as attacks. Consequently, the model produces a high false alarm rate. This directly affects operational usability because an IDS that raises too many false alerts can overwhelm analysts and reduce trust in the system. The moderate precision of around 45% confirms this issue: fewer than half of the WMM alerts correspond to actual attacks.

This is an expected result for the real dataset due to its nature. Indeed, this behavior can be partly explained by the fact that the dataset contains few features following Weibull distributions, while most of the features follow non-Weibull distributions such as lognormal laws. Indeed, lognormal tails often allow extreme normal values. A fitted Weibull may treat those values as extremely unlikely, causing false positives. More specifically, since the WMM imposes Weibull marginals on all features, normal samples associated with the lognormal marginals will receive underestimated likelihoods, especially in the heavy-tail region. Consequently, the WMM does not catch the true variability of benign traffic, causing normal but rare patterns to be incorrectly treated as anomalies since these benign observations are assigned high anomaly scores, increasing false positives and degrading specificity, Precision, ROC-AUC, PR-AUC, and MCC. Indeed, the WMM has an ROC-AUC of 0.6 which is only slightly above random ranking. LOF achieves 0.873. This means LOF ranks attacks above normal samples much more consistently than the WMM. Moreover, the WMM achieves a weak PR-AUC (0.448) Compared with LOF and the autoencoder. In imbalanced IDS scenarios, PR-AUC is often more informative than ROC-AUC. Therefore, the low PR-AUC is a significant weakness. In addition, the lower MCC of the WMM confirms that its predictions are unbalanced: it detects most attacks but misclassifies many normal samples, and even Figure 3c shows that the WMM has the best MCC. In contrast, the heatmap shows that the other models largely outperform the WMM model for the other metric except the recall, F1 score, and MCC, where they perform badly.

These results suggest that the WMM is effective as a high-recall detector, but false-alarm control depends strongly on the validity of its marginal distribution assumptions. Consequently, the WMM is particularly suitable for contexts where false negatives are more dangerous than false positives, such as the IIoT context. In practice, the WMM can play a complementary role with the other models and can be suitable as a first-stage screening detector benefiting from a second-stage verifier such as LOF, One-Class SVM, or an autoencoder. Indeed, it can play the role of catching nearly all suspicious traffic, minimizing missed attacks, and forward suspicious samples to a second-stage verifier, mainly LOF and the autoencoder, which have high precision and specificity, so they can be used after the WMM to filter alerts. Otherwise, the WMM alone would generate too many false alarms.

4.5. Results on the IIoT Synthetic Dataset

We present in this section a broad benchmarking study of the WMM, compared mainly against anomaly baseline models. The study evaluates the model from several angles: model selection, ranking quality, classification performance, statistical significance, robustness, runtime, ablation, and anomaly-ratio sensitivity.

4.5.1. Model Selection and Ablation Study for the WMM

Figure 4a shows that both BIC and MMDL select the same mixture complexity across repeated runs. Both criteria prefer $K = 4$, indicating that the benign traffic distribution is not well represented by a single Weibull component. Instead, it appears to contain several sub-regimes or latent traffic modes. In IIoT traffic, this is plausible because benign

flows can vary by device type, protocol behavior, time interval, communication pattern, or workload intensity. This result supports the use of a mixture model rather than a single parametric Weibull distribution. The agreement between BIC and MMDL is also reassuring because these criteria penalize over-complexity differently but still converge to the same dominant choice. This suggests that the selected model complexity is not arbitrary and that the four-component WMM captures meaningful heterogeneity in the benign data.

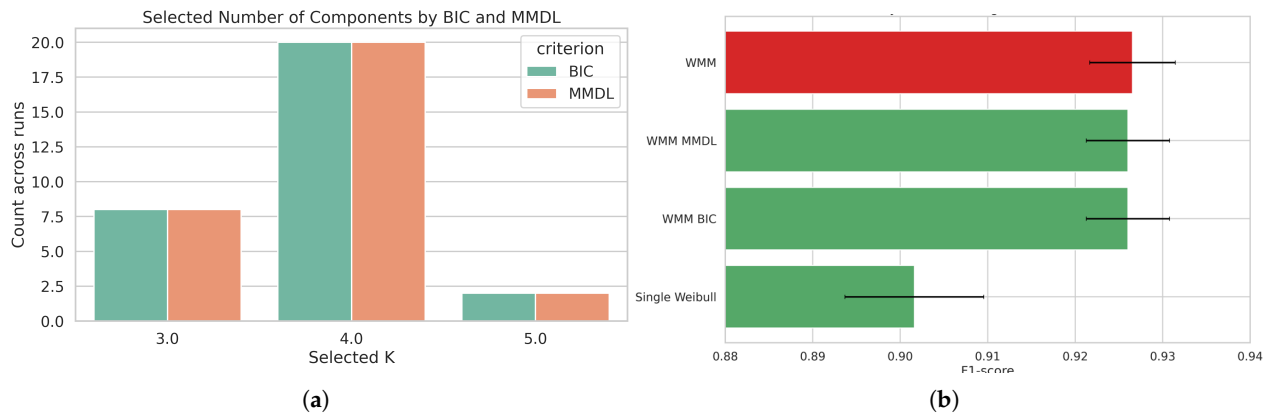


Figure 4. (a) Model selection for BIC and MMDL. (b) Ablation for metric F1.

Figure 4b shows the ablation plot comparing four scenarios: full WMM, WMM with MMDL selection technique, WMM with BIC selection technique, and single Weibull model for the metric F1. We note that the full mixture-based WMM and its BIC/MMDL variants perform similarly, with F1-scores around 0.92–0.93. The single Weibull model performs clearly worse, around 0.90 or slightly below. This demonstrates that the mixture structure provides real performance benefits. A single Weibull distribution is too rigid to capture benign traffic heterogeneity. The mixture model captures several benign modes, which reduces false alarms and improves anomaly discrimination. The similarity between BIC and MMDL variants also confirms that model-selection strategy is stable. Both criteria identify useful complexity, and neither appears to overfit severely.

4.5.2. Main Results

Figure 5 shows the main benchmarking results within a heatmap plot and a bar chart. As we can observe, the WMM achieves the best performance in central metrics for anomaly detection mainly in ROC-AUC, PR-AUC, F1-score, recall, and MCC.

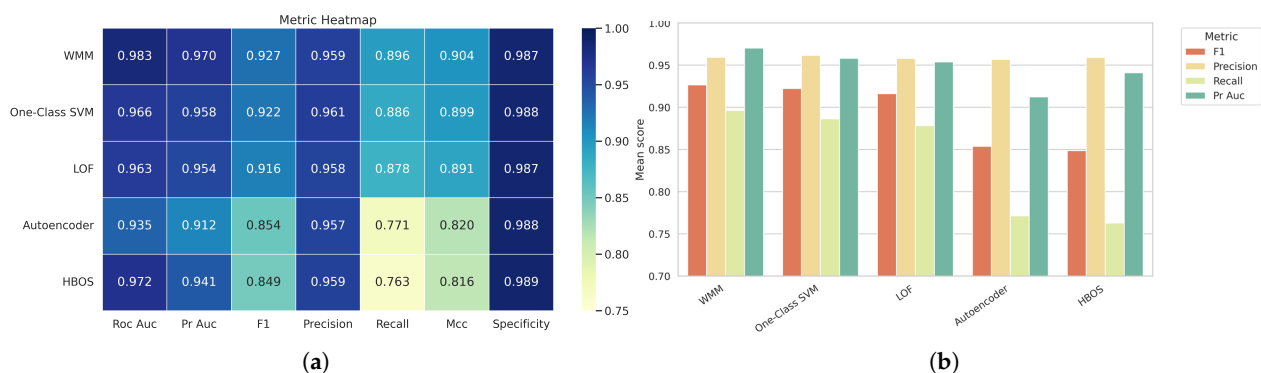


Figure 5. (a) Heatmap plot for all metrics. (b) Bar chart for selected metrics.

The One-Class SVM slightly exceeds the WMM in precision and specificity, but its recall, F1-score, PR-AUC, and MCC are lower. The WMM, by contrast, detects more anomalies while maintaining almost the same precision and specificity.

Concerning LOF, it also performs well, but it remains below the WMM and One-Class SVM across most metrics. HBOS has very high specificity but weaker recall and F1-score making it missing more anomalies. Now for the autoencoder, the figure shows that it has the weakest overall performance among the main models, especially in recall, F1-score, PR-AUC, and MCC.

4.5.3. Tradeoffs Between the Models

Figure 6 shows scatter plots clarifying the trade-offs between the models.

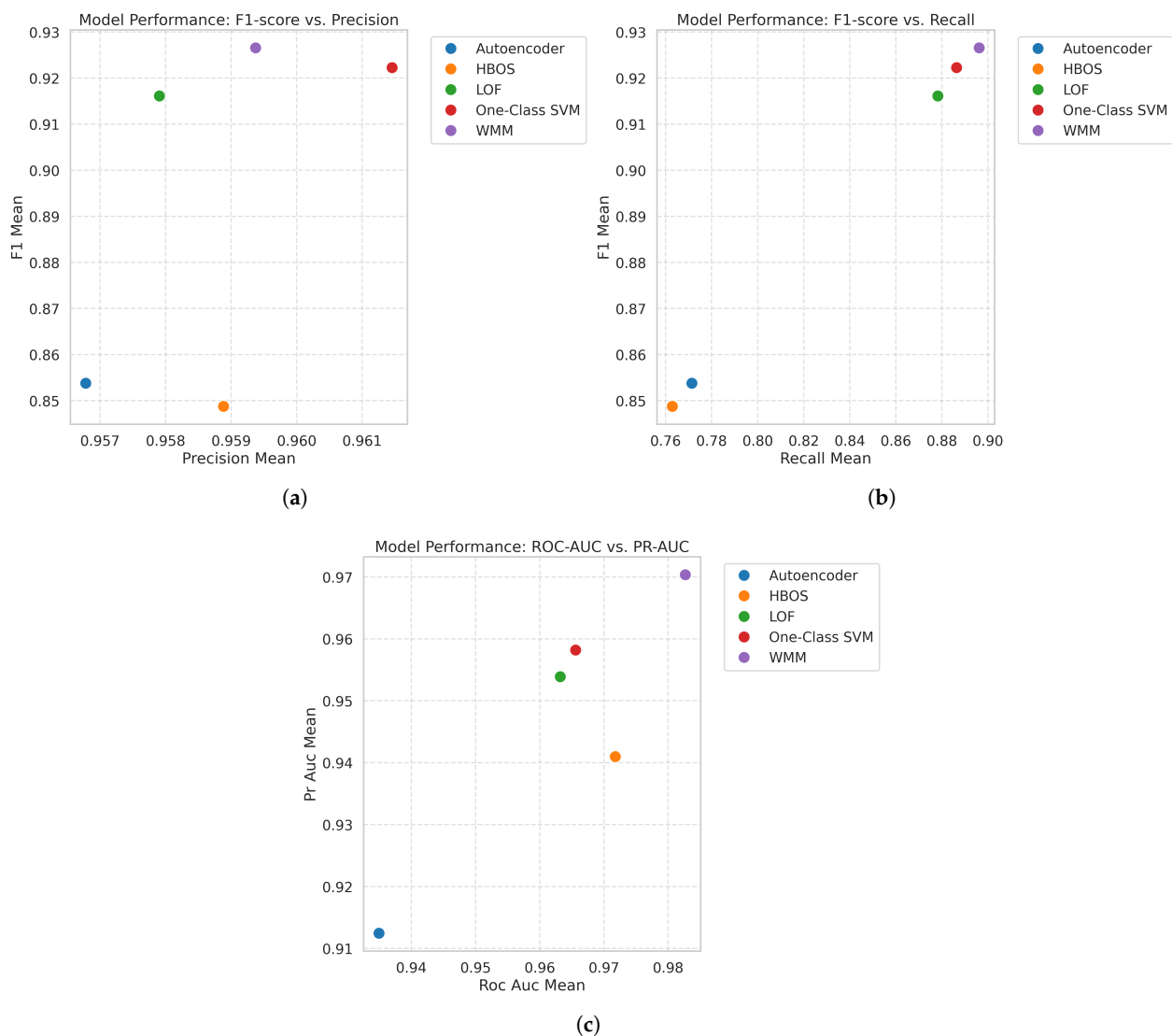


Figure 6. (a) F1-score vs. precision. (b) F1-score vs. recall. (c) ROC-AUC vs. PR-AUC.

We can observe that the WMM performs the best on most metrics. In particular, its PR-AUC of approximately 0.970 is especially important because PR-AUC is more informative than ROC-AUC when anomalies are rare. It indicates that WMM separates benign and anomalous samples effectively.

Even if One-Class SVM has slightly higher precision than the WMM, the latter achieves better recall and better ranking quality. LOF is behind the WMM and One-Class SVM, while HBOS and the autoencoder have lower F1-scores because their recall is substantially lower even they have high precision and specificity which is very high for all models, around 0.987–0.989. On other hand, the WMM has the highest MCC, which is around 0.904.

This high MCC confirms that the WMM is balanced, not merely optimized for one class of confusion matrix outcomes.

Referring to the real-dataset results, where the WMM showed high recall but low specificity, on this synthetic benchmark, the WMM is both sensitive and specific. This is due to the fact that the data-generating distribution aligns with Weibull-like assumptions.

4.5.4. Anomaly-Ratio Sensitivity

We aim here to study the impact of the anomaly prevalence change on the different performance metrics. Figure 7 shows this anomaly-ratio sensitivity into two plots. We can remark that ROC-AUC remains very stable, staying around 0.98 across anomaly ratios. So we can say that the ranking quality of the WMM score is independent of anomaly prevalence. Concerning the PR-AUC, we observe a steady increase as the anomaly becomes more frequent. It is not surprising since it is known that precision-based metrics are strongly affected by class imbalance. This is why low prevalence penalizes Precision and F1-score even when recall and ROC-AUC remain strong. We deduce that we should take care when we deploy the model to avoid excessive alerts in such situations of rare outliers.

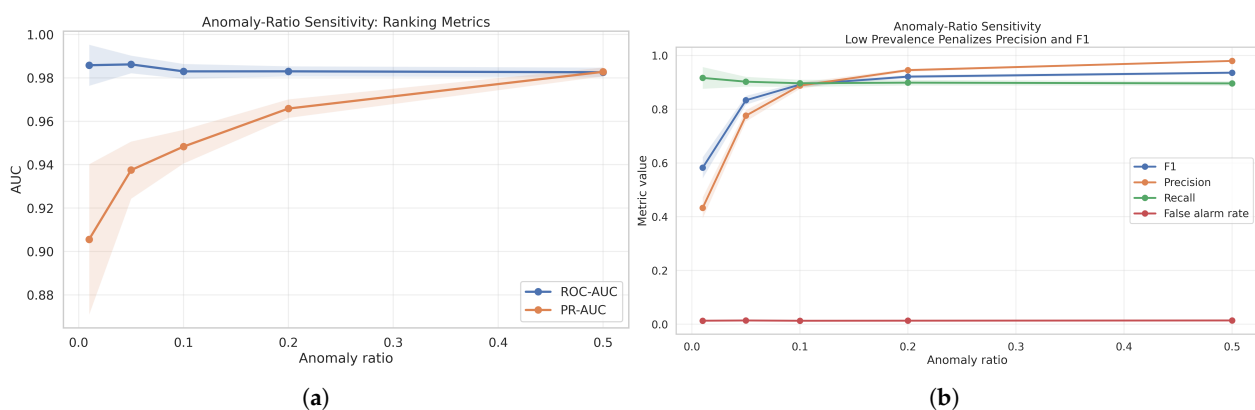


Figure 7. (a) Sensitivity of ROC-AUC. (b) Sensitivity of different metrics.

4.5.5. Robustness Analysis

Under three perturbation types, which are feature corruption, Gaussian noise, and scale drift, we aim to test our model robustness. Figure 8 shows the experimental results. The model is robust under feature corruption and mild perturbation levels, but its performance degrades under stronger Gaussian noise and scale drift.

Figure 8a,b show the F1-score robustness bar chart across perturbation levels. The model confirms strong robustness for F1-score (slight shift from 0.927 to 0.91) under feature corruption. However, it drops sharply under Gaussian noise, and scale shift, mainly for severe perturbation levels, and struggles to maintain a balance between Precision and recall. Figure 8c explains this drop and shows that it is driven by a large loss of precision, and therefore the model produced more false positives, as shown by Figure 8e for the false alert rate. Moreover, this fact is confirmed by the recall plot (Figure 8d) showing that the recall remains relatively high for most perturbation types. Consequently, we can deduce that the model continues to detect many actual positive cases, even under perturbation but also generates more false alarms.

On other side, Figure 8f,g show the PR-AUC and ROC-AUC results. The metrics remain high across all perturbation types, but they decrease under Gaussian noise and scale drift, becoming less reliable across different decision thresholds, especially in maintaining high precision while preserving recall.

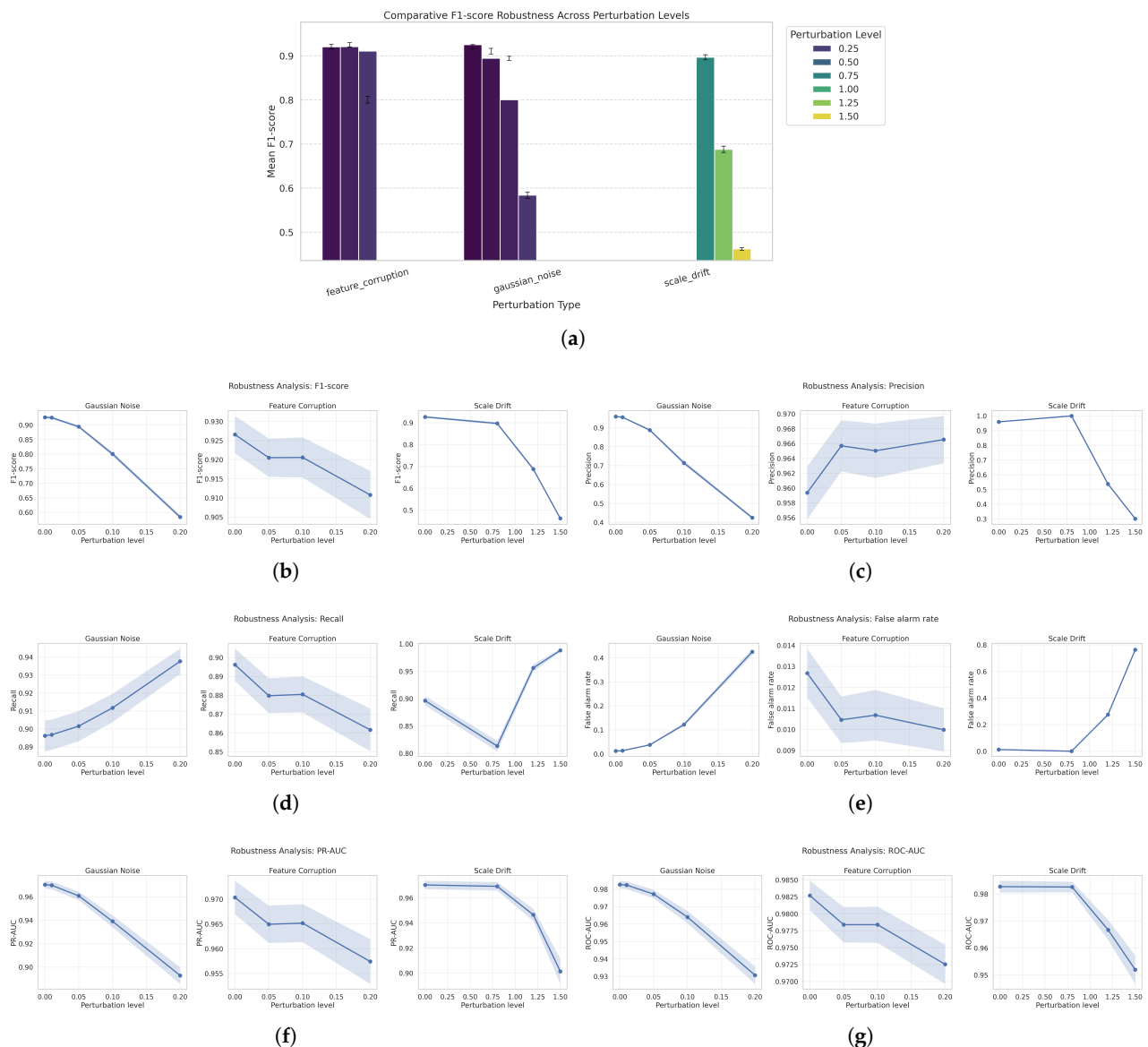


Figure 8. Robustness of WMM against perturbations for performance metrics: (a) F1-score robustness bar chart, (b) F1-score, (c) precision, (d) recall, (e) false alarm rate, (f) PR-AUC, and (g) ROC-AUC.

Overall, the robustness study shows that the framework is highly robust to feature corruption but more vulnerable to Gaussian noise and scale drift. This finding implies that, when deployed, the WMM should be paired with periodic retraining. However, the good thing is that high PR-AUC and ROC-AUC values indicate that the model still retains useful discriminative power across perturbation conditions.

4.5.6. Statistical Significance of WMM and Runtime vs. Efficiency

Figure 9a shows the mean performance difference between the WMM and competing models across metrics with a significance threshold at $p = 0.05$. Most WMM advantages are positive, and many are statistically significant. The autoencoder and HBOS perform badly and the WMM's competitors are mainly One-Class SVM and LOF. The differences against these two competitors are smaller but still often positive. Therefore, the WMM advantage is incremental rather than absolute. Scientifically, the WMM should be described as the best-performing model in this study, but with One-Class SVM as a strong and competitive baseline.

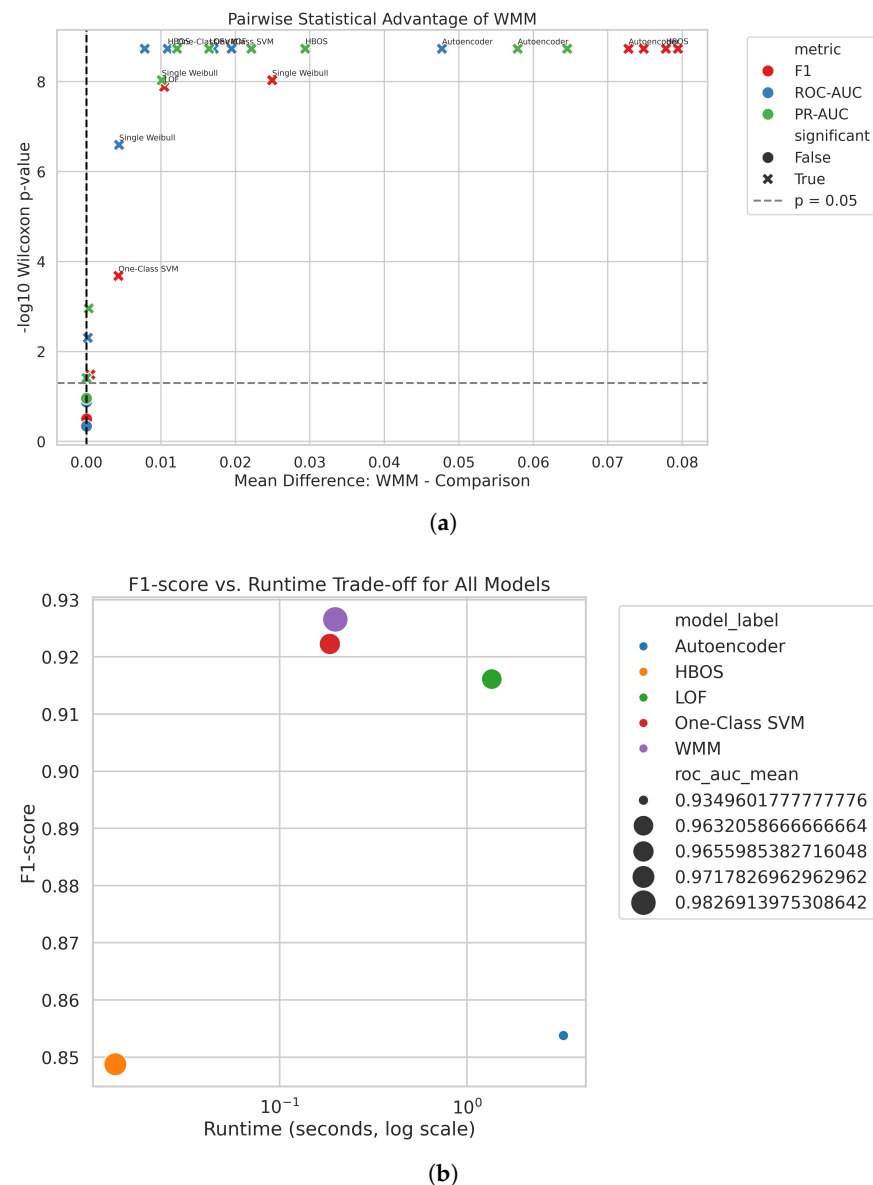


Figure 9. (a) Pairwise significance with baseline models. (b) Runtime–F1 trade-off.

Figure 9b presents the runtime–F1 trade-off plot. It shows that the WMM provides a favorable balance between accuracy and computational cost. HBOS is the fastest model but has lower F1-score. LOF is slower and does not outperform the WMM. One-Class SVM is competitive in F1-score but slightly below the WMM. The autoencoder has a lower F1-score and requires more model complexity.

The WMM therefore occupies a desirable position with a high detection quality and moderate runtime. This supports deployment at network gateways or edge nodes, especially when compared with heavier reconstruction-based approaches. The model is also more interpretable than an autoencoder because anomaly scores can be traced back to feature likelihoods and mixture-component responsibilities.

4.6. Overall Benchmarking Result Discussion

The contrast between the real and synthetic results is highly informative. On the synthetic dataset, the WMM performs as a strong and balanced detector because the majority of features follow the distribution family assumed by the model. On the real dataset, the WMM still detects attacks effectively, but it came at the cost of many false

positives. On other hand, across both real and synthetic datasets, the WMM shows strong recall, meaning it is effective at identifying abnormal traffic.

We conclude that the combined benchmarking results provide a balanced view of the WMM. On the synthetic dataset, the WMM is achieving excellent performance on most metrics. On the real dataset, the WMM remains highly effective at detecting attacks but with very high false alarms. The WMM is a promising and interpretable IDS model with a practical advantage over opaque methods such as the autoencoder; anomaly scores can be decomposed through feature-wise likelihoods, mixture-component responsibilities, and Weibull shape/scale parameters. However, it has a main limitation revealed by the real dataset, which is the false alarm control. We conclude that the WMM is mathematically appropriate and empirically strong when IIoT features are positive, skewed, and Weibull-like, but it is sensitive to distributional mismatch and benign-traffic complexity. Its performance depends not only on the number of Weibull features but also on feature separability, tail behavior, and correlations.

5. Model Deployment

This section starts by discussing, in a macro view, the proposed statistical framework deployment architecture in term of the protocols, components, and management issues for both the smart home and IIoT contexts. It ends by providing a deployment architecture of the proposed WMM framework.

5.1. Architectural Components

Smart IoT devices are usually connected to a central hub or gateway for communication and internet connectivity. They communicate through Wi-Fi, Zigbee, Z-Wave, and Bluetooth generating distinct traffic patterns. So, as a first option, we can embed the WMM framework directly into the smart home hub, enabling real-time traffic analysis. If we have to consider the lack of computational power for complex statistical models, necessitating optimization or lightweight implementations, we can opt to send traffic data to a cloud service to bypass the computational constraints. This option raises privacy issues and needs to be mitigated by encryption and anonymization techniques. Alternatively, as a third architectural option, a hybrid approach (local lightweight monitoring with cloud-based intensive analysis) offers a balanced solution, minimizing latency while addressing resource constraints.

In the IIoT, devices use specialized protocols such as Modbus, Profinet, and Ethernet/IP, often integrated with gateways for data aggregation. Considering that any disruptions in the IIoT can lead to significant consequences, the architecture should prioritizes real-time detection and low latency without sacrificing scalability and precision. We can opt for a distributed architecture with three layers: (1) an edge component performing initial data collection and preprocessing; (2) a gateway component that aggregates data from the edge device and forwards it to centralized systems; and (3) a component that analyzes traffic for anomalies by running the model on private SOC or private cloud. A fourth component should be dedicated to WMM tuning and management.

Overall, for both smart home and IIoT contexts, the architectural components could be the following:

- A data collection and preprocessing component using network sniffers, packet capture tools, or device APIs to gather traffic data and prepare it for analysis.
- A Model component to train and run the Weibull Mixture Model in order to detect anomalies based on preprocessed data.
- An alerting component that notifies users (smart home) or administrators (IIoT) via mobile apps, email, or dashboards.

- A management console equipped with a centralized interface overseeing model updates, parameter tuning, and performance monitoring.

Furthermore, deploying the model introduces management issues with several challenges mainly related to (1) model update and retraining since new devices or changing behaviors necessitate periodic retraining and therefore model parameter updates, (2) tuning the model balance between false positives and negatives using user feedback, and finally (3) scalability by balancing load to ensure performance and privacy and security by considering encryption and anonymization to protect sensitive data, especially in the cloud deployments.

Based on these architectural discussion insights, we propose in the next section our deployment architecture for the WMM. First we resume the proposed framework and after that we present the proposed WMM deployment architecture.

5.2. Deployment Architecture of the WMM

Figure 10a shows the overall architecture of the proposed context-aware Weibull Mixture Model-based intrusion detection framework. IoT traffic will be first collected and converted into statistical features. The extracted features are assigned to device or operation specific contexts, where dedicated benign Weibull Mixture Models and context-specific outlier components are used to model normal and abnormal behavior. The combined context-specific models generate anomaly scores, which are evaluated against validation-derived thresholds to produce the final benign or anomaly decision. The operational feedback layer supports monitoring, drift detection, and periodic model updates.

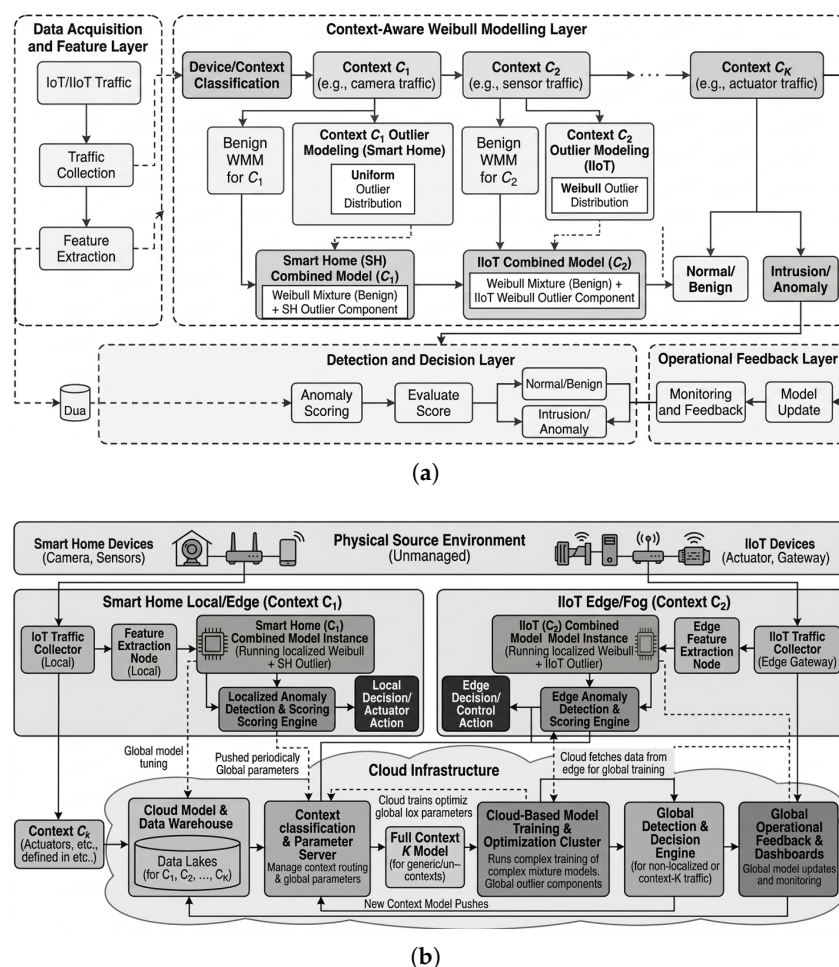


Figure 10. (a) WMM framework architecture. (b) Deployment architecture.

Figure 10b presents the deployment architecture of the proposed context-aware Weibull Mixture Model-based intrusion detection system. The architecture follows a hybrid edge—cloud design in which heterogeneous smart home and IIoT devices generate traffic in an unmanaged physical source environment, while local edge or fog nodes perform traffic collection, feature extraction, anomaly scoring, and immediate decision-making. Separate context-specific model instances are deployed for smart home traffic, IIoT traffic, and additional eventual contexts. Each local model combines a benign Weibull mixture component with a context-specific outlier component, enabling lightweight and interpretable anomaly detection close to the traffic source. The cloud layer stores historical data and model artifacts, manages context routing and global parameters, trains and optimizes context-specific and full-context models, and provides global monitoring dashboards. The bidirectional feedback path allows edge observations, analyst feedback, and drift indicators to be used for model retraining and recalibration, and periodic deployment of updated parameters to the edge.

6. Conclusions

The understanding of device behaviors, traffic patterns, and threat landscapes helps us to propose a taxonomy of IoT devices for smart home IoT and industrial IoT. We then built a soundly formulated statistical framework that uses Weibull Mixture Models for normal traffic and context-specific outlier distributions for anomaly detection. We validated the WMM for the IIoT on a real and synthetic datasets. The results are encouraging and demonstrate the framework's efficacy and its usefulness. However, the WMM needs to consider the variability of user-driven traffic in smart homes and attacks sophistication in the IIoT by continuously retraining the model to adapt to the changing patterns. A preliminary architectural and managerial deployment of the WMM was proposed to tackle that challenge and will be detailed additionally and experimented with as future work. Moreover, we will focus on several other directions to enhance the framework's applicability. We intend to extend the framework to incorporate other heavy-tailed distributions such as lognormal and experiment with a hybrid mixture with the WMM to adequately capture real IoT environments. In particular, we will investigate future extensions considering hybrid marginals, such as Weibull–lognormal–Gaussian mixtures. Currently, the WMM ignores dependencies between features, such as packet rate and byte count, duration and volume, or protocol type and connection behavior which reduces density-estimation quality. As another direction, we will investigate dependence modeling using copulas, grouped features, or enriched multivariate mixture components to improve its quality. We will also explore time-series analysis or sequential modeling techniques to improve the detection of temporal anomalies, such as gradual data exfiltration or evolving attack patterns. Finally, exploring hybrid outlier models that combine uniform and Weibull distributions in both smart home and IIoT contexts could provide comprehensive coverage against a broader spectrum of threats.

Funding: The author would like to thank SAUDI ARAMCO Cybersecurity Chair, Imam Abdulrahman Bin Faisal University, for funding this project.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the author on request.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. IoT Analytics. State of IoT 2025: Number of Connected IoT Devices Growing 14% to 21.1 Billion Globally. IoT Analytics. 2024. Available online: <https://iot-analytics.com/number-connected-iot-devices/> (accessed on 1 May 2026).
2. Atlam, H.F.; Hemdan, E.E.-D.; Alenezi, A.; Alassafi, M.O.; Wills, G.B. Internet of Things Forensics: A Review. *Internet Things* **2020**, *11*, 100120. [\[CrossRef\]](#)
3. Alrawi, O.; Lever, C.; Antonakakis, M.; Monrose, F. SoK: Security evaluation of home-based IoT deployments. In *Proceedings of the 2019 IEEE Symposium on Security and Privacy*; IEEE: New York, NY, USA, 2019.
4. Ronen, E.; Shamir, A.; O'Flynn, C.; Shamir, A.; Weingarten, A. IoT goes nuclear: Creating a ZigBee chain reaction. *IEEE Secur. Priv.* **2018**, *16*, 54–62. [\[CrossRef\]](#)
5. Celik, Z.B.; Babun, L.; Sikder, A.K.; Aksu, H.; Tan, G.; McDaniel, P.; Uluagac, A.S. Sensitive information tracking in commodity IoT. In *Proceedings of the 27th USENIX Security Symposium*, Baltimore, MD, USA, 15–17 August 2018.
6. Manivannan, D. Recent endeavors in machine learning-powered intrusion detection systems for the Internet of Things. *J. Netw. Comput. Appl.* **2024**, *229*, 103925. [\[CrossRef\]](#)
7. Elrawy, M.F.; Awad, A.I.; Hamed, H.F.A. Intrusion detection systems for IoT-based smart environments: A survey. *J. Cloud Comput.* **2018**, *7*, 21. [\[CrossRef\]](#)
8. Chatterjee, A.; Ahmed, B.S. IoT anomaly detection methods and applications: A survey. *Internet Things* **2022**, *20*, 100568. [\[CrossRef\]](#)
9. Sivanathan, A.; Sherratt, D.; Gharakheili, H.H.; Radford, A.; Wijenayake, C.; Vishwanath, A.; Sivaraman, V. Characterizing and classifying IoT traffic in smart cities and campuses. In *Proceedings of the 2017 IEEE Conference on Computer Communications Workshops*; IEEE: New York, NY, USA, 2017.
10. Kumar, S.; Cunche, M.; Kaafar, M.A.; Shafiq, Z.; Uzmi, Z.A. Characterizing IoT traffic in smart home and campus environments. In *Proceedings of the 2020 Proceedings IEEE INFOCOM*; IEEE: New York, NY, USA, 2020; pp. 2706–2715.
11. Fan, W.; Bouguila, N.; Sallay, H. Anomaly intrusion detection using incremental learning of an infinite mixture model with feature selection. In *Rough Sets and Knowledge Technology (RSKT 2013)*; Lingras, P., Wolski, M., Cornelis, C., Mitra, S., Wasilewski, P., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8171, pp. 364–375.
12. Leung, K.; Leckie, C. Unsupervised anomaly detection in network intrusion detection using clusters. In *Proceedings of the Twenty-Eighth Australasian Conference on Computer Science*, Newcastle, Australia, 1 January 2005; Volume 38, pp. 333–342.
13. Ngueajio, M.K.; Washington, G.; Rawat, D.B.; Ngueabou, Y. Intrusion Detection Systems Using Support Vector Machines on the KDDCUP99 and NSL-KDD Datasets: A Comprehensive Survey. In *Intelligent Systems and Applications. IntelliSys 2022*; Arai, K., Ed.; Lecture Notes in Networks and Systems; Springer: Berlin/Heidelberg, Germany, 2023; Volume 543.
14. Arfeen, A.; Pawlikowski, K.; McNickle, D.; Willig, A. The role of the Weibull distribution in modelling traffic in Internet access and backbone core networks. *J. Netw. Comput. Appl.* **2019**, *141*, 1–22. [\[CrossRef\]](#)
15. Mazhar, M.H.; Shafiq, Z. Characterizing Smart Home IoT Traffic in the Wild. *arXiv* **2020**, arXiv:2001.08288.
16. Nguyen-An, H.; Silverston, T.; Yamazaki, T.; Miyoshi, T. IoT Traffic: Modeling and Measurement Experiments. *IoT* **2021**, *2*, 140–162. [\[CrossRef\]](#)
17. Kirichek, R.; Kulik, V. Industrial Internet of Things Traffic Research and Generation. *Proc. Telecommun. Univ.* **2019**, *5*, 27–36. [\[CrossRef\]](#)
18. Halima, B.; Chaker, A. Weibull Mixture Model for Reliability Analysis. *Int. Rev. Electr. Eng.* **2014**, *9*, 986–990. [\[CrossRef\]](#)
19. Bennis, A.; Mouysset, S.; Serrurier, M. Estimation of conditional mixture Weibull distribution with right-censored data using neural network for time-to-event analysis. *arXiv* **2020**, arXiv:2002.09358.
20. Canadian Institute for Cybersecurity. *IoT Dataset 2023*; University of New Brunswick: Fredericton, NB, Canada, 2023.
21. Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. *IEEE Access* **2022**, *10*, 40281–40306. [\[CrossRef\]](#)
22. Firouzi, A.; Dadkhah, S.; Maret, S.A.; Ghorbani, A.A. DataSense: A Real-Time Sensor-Based Benchmark Dataset for Attack Analysis in IIoT with Multi-Objective Feature Selection. *Electronics* **2025**, *14*, 4095. [\[CrossRef\]](#)
23. Khreich, W.; Khosravifar, B.; Hamou-Lhadj, A.; Talhi, C. An anomaly detection system based on variable N-gram features and one-class SVM. *Inf. Softw. Technol.* **2012**, *54*, 1178–1191.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.