Journal of Intelligent Information and Communication

Research Paper

Robust IoT Device Fingerprinting and Camouflage Detection Based on Residual Attention Networks

Czesława Lucyna Kamińska[1,*], Edyta Dziuba[2] and Malwina Kasprzakowa[2]

[1] Faculty of Informatics and Information Technology, Silesian University of Technology, Gliwice, 44-100, Poland
[2] Faculty of Computer Science and Information Systems, Nicolaus Copernicus University in Torun, Torun, 87-100, Poland
*Corresponding author: czeslawa.l@polsl.pl

Abstract. The widespread use of the Internet of Things (IoT) has created new security issues, and attackers have now created a variety of stealth techniques to get beyond conventional fingerprinting technology. The goal of this work is to address the current issues with Internet of Things (IoT) authentication. bolster device fingerprinting and enhance the detection of camouflaged impersonators using a novel residual attention neural network framework. Feature extraction using many modalities for multidimensional data collection on network behaviour, physical-layer signals, and high-level operational statistics of different devices. More than 25 different kinds of IoT devices from different manufacturers, including those with both normal and hostile behaviours, were methodically simulated using a fully functional experimental testbed. According to the aforementioned findings, this method's recall and precision rates surpass 96%, and its classification accuracy in a high-intensity camouflage attack surpasses 97%. In a hostile setting, the performance decline was less than 3.7%. The model can be implemented on the edge at the necessary speed and within the power restrictions, according to resource analysis. Because the system functions well in all conditions and at all times, it is less likely to be mistaken for a false alert. Thus, the application value of a hybrid residual-attention module for secure adaptive IoT device identification in a network has been confirmed based on the aforementioned analysis.

Keywords: *IoT Security, Device Fingerprinting, Camouflage Detection, Residual Attention Network*

Received on 18 January 2025, Accepted on 24 April 2025, Published on 04 May 2025

Introduction

Many various types of smart gadgets are now widely used in homes, businesses, and other settings due to the Internet of Things' (IoT) quick development [1]. All-important digital infrastructure must now function dependably and securely for these linked systems, since it is anticipated that the number of connected devices to the Internet of Things worldwide will surpass tens of billions in the upcoming years [2]. While the aforementioned connectivity improvements enable sophisticated automation, remote monitoring, etc., they have also increased the attack surface and revealed a number of vulnerabilities in low-power, resource-constrained IoT devices [3]. The lack of a single security standard and the numerous protocols, hardware, and physical settings involved in actual IoT deployment exacerbate the aforementioned issues [4]. As a result, the conventional network security strategy, which concentrated on the perimeter, address-based access control, and static credentials, is no longer appropriate for thwarting sophisticated and enduring threats in the Internet of Things ecosystem [5].

Device fingerprinting is one of the advanced defence techniques used to identify and verify Internet of Things (IoT) entities based on intrinsic physical-layer traits, network behaviour, or protocol-specific attributes that are theoretically hard to fabricate or replicate [6,7]. Fingerprint technology, on the other hand, makes advantage of a device's unique characteristics and is therefore non-replicable, whereas the previous kind of authentication is easily stolen or reproduced. However, noise and instability are introduced to the derived fingerprints because of the variety of manufacturing processes, communication settings, and practical working situations [8]. The stability and dependability of the current fingerprint-based detection method are at risk due to the increasing

prevalence of sophisticated camouflage attacks, which have altered device emissions, signal patterns, and slightly altered protocol behaviour [9]. Attackers can now purposefully create signals or traffic patterns to evade or confuse conventional fingerprinting models, which raises the false-negative rate and erodes public confidence in the IoT security framework [10].

Thus, a strong framework for device fingerprinting and camouflage detection based on residual attention networks is presented in this paper. Our approach, which builds on the advancements of deep neural networks, uses residual learning to train deep models more efficiently through shortcut connections and adds attention mechanisms to dynamically choose crucial device properties for authentication in challenging and hostile situations. Our residual attention-based model can improve the separation of feature representations and is intrinsically more resilient to sophisticated camouflage than the previous approach of treating signal and protocol characteristics equally and without adaptation. Create comprehensive threat models in a methodical manner, construct feature extraction pipelines, and use extensive, real-world data to experimentally validate the suggested techniques. In terms of identification accuracy, attack resilience, and operating efficiency, the suggested framework has surpassed the current top baselines, according to the experimental results mentioned above. The remainder of this paper is organised as follows: The IoT security threat model and a particular problem definition are presented in Section 2; the proposed residual attention network architecture and its methodological innovations are introduced in Section 3; the evaluation procedures and significant experimental results are described in Section 4; the robustness and generalisation capabilities are examined in Section 5; and the primary research contributions and useful applications are summarised in Section 6.

Threat Model and Problem Definition

IoT Fingerprinting Challenges

The complexity and variety of contemporary deployments have led to a number of significant issues with IoT device fingerprinting. As several hardware platforms, operating systems, and communication protocols for Internet of Things devices have evolved, it has become increasingly challenging to produce reliable and unique fingerprints [11]. Due to the use of shared chipsets, recycled designs, or standardised components, devices frequently have similar feature sets and, consequently, few unique features [12]. The environment can also contribute to this; interference, multipath propagation, or signal loss can occur wirelessly [13].

The aforementioned conventional fingerprinting techniques, such as radio-frequency signatures, packet timing, device emission patterns, and protocol behaviour, are typically low-level [14]. The aforementioned indications have increased as a result of both manufactured traffic congestion and ambient noise. Network congestion and background interference can affect packet timing and radiometric analysis, and many commodity devices can normalise or spoof protocol-based characteristics [15]. Additionally, devices from various vendors may have very similar data-plane and physical-layer characteristics due to the commercialisation of Internet of Things hardware [16]. As a result, classification models that assume a large feature-entropy difference between devices perform much worse [17]. As a result, even the greatest fingerprinting systems available today are still prone to a high percentage of false positives or false negatives when new hardware or devices are encountered [18]. The aforementioned shortcomings point to the urgent need for an adaptation to a multi-modal, context-aware Internet of Things fingerprinting system that can handle variation both within and between classes [19,20].

Camouflage Attack Scenarios

As camouflage attack technology advances quickly, it is now possible to purposefully change a device's signal or traffic record to make it seem authentic. The first kind involves leveraging known device identifiers in a target network environment, or MAC address spoofing [21]. Attackers have used a variety of techniques to modify real-time changes in physical-layer timing aspects, protocol payload structures, and emission intervals through software-defined radios or programmable network stacks, building on the foundation of fundamental spoofing [22]. While some methods dynamically learn and modify their emulation strategy based on defender feedback, others replay the captured legal device signature [23].

Additionally, highly skilled malicious actors can combine or change a number of features to modify their frequency distribution, introduce controlled noise, and adjust their inter-packet timing, all of which nearly match

those of authentic devices [24]. The development of off-the-shelf radio hardware and open-source communication libraries has also aided the aforementioned efforts by lowering the technical barrier to launching targeted camouflage attacks [25]. In order to evade detection by anti-fingerprinting devices, adversarial tools can be used to develop ways to adapt to changes in the behaviour of fingerprinting models and take advantage of weaknesses in these models. When combined, the aforementioned bad actors demonstrate that any useful fingerprinting system needs to be able to deal with both benign alterations and malevolent falsifications by determined criminals.

Security Objectives and Threat Assumptions

For the Internet of Things (IoT) network, this study suggests a fingerprinting and authentication system that can withstand complex and flexible camouflage attacks and continue to function normally for authorised users. The primary goal is to accurately identify authentic devices in a variety of unprotected and potentially dangerous wireless situations. As a result, the system should be able to apply a model architecture that can identify subtle multi-modal patterns in the flow of IoT data and extract high-fidelity and discriminative features that are difficult to change.

According to the adversarial model, an attacker with sufficient technical resources can observe network traffic, replay or alter protocol fields, use programmable radios, and use machine learning to enhance its emulation attempts. Nevertheless, the attacker cannot directly change the fundamental network infrastructure and does not have access to the network's cryptographic secrets. In terms of defence, the framework should be passive, need little processing power, and have no secret keys that are hard to get in low-cost IoT nodes. As a result, it has started to automate feature selection and dynamically concentrate on patterns that are more likely to be connected to a genuine device using observable signal artefacts and protocol behaviour.

Thus, the challenge to be tackled in this paper is to develop and assess an adaptive, reliable, and scalable Internet of Things fingerprinting system that is resistant to a changing range of impersonation and camouflage threats and can generalise across various devices and network scenarios.

Network Architecture and Methodology

Residual Attention Network Design

The architecture of this paper is a hybrid of residual learning and attention mechanisms, which are well-suited to handle the diverse and adversarial characteristics of device-generated data, in order to detect camouflaged impersonators in IoT systems. The linkages between the primary modules are as follows: input feature fusion, stacks of residual blocks, attention-based weighting, and a final discriminative classification head. Figure 1 depicts the whole model structure.

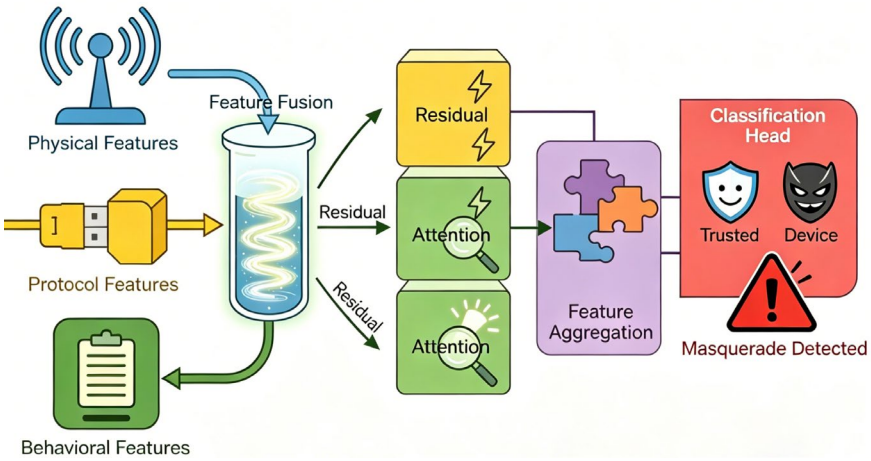

Figure 1. Overall architecture of the residual attention network for IoT fingerprinting and camouflage detection.

https://doi.org/10.64972/jiic.2025v3.225p12s:154-169

At the entry end, a variety of data are gathered and combined from many sources to create a single vector representation, such as upper-level session attributes, traffic statistics, and radio-frequency indicators. In order to record both long-term equipment variations and transient behavioural shifts that can distinguish genuine activity from extremely realistic fake activities, the two are typically combined. To serve as the foundation for deep processing, the aforementioned unified features are subsequently mapped to an abstract latent space.

The network's core is a stack of residual blocks, each designed to facilitate direct signal propagation and overcome the notorious vanishing gradient problem commonly seen in deep architectures. Formally, for a given block input vector $\mathbf{x}$, the output of the block is defined as:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathbf{W}) + \mathbf{x} \quad \text{Eq. (1)}$$

where $\mathcal{F}$ denotes a composite nonlinear transformation parameterized by weights $\mathbf{W}$, typically involving linear mapping, normalization, and rectified activation. The additive shortcut ensures gradient stability, making the network robust against noisy permutations or statistical irregularities that arise from environmental fluctuations or adversarial manipulation.

After each residual transformation, the model applies a self-attention mechanism. Here, the latent feature outputs $\mathbf{h}_i$ from the residual block are adaptively re-weighted to highlight information-rich subspaces relevant to authentic device traits or camouflage signals. Attention weights α_i are calculated as:

$$\alpha_i = \frac{\exp\left(f_{\text{att}}(\mathbf{h}_i)\right)}{\sum_j \exp\left(f_{\text{att}}(\mathbf{h}_j)\right)} \quad \text{Eq. (2)}$$

where f_{att} is a learned scoring function. The aggregate, attention-weighted embedding is thus:

$$\mathbf{z} = \sum_i \alpha_i \mathbf{h}_i \quad \text{Eq. (3)}$$

This design compels the model to focus discriminative capacity on the subtle characteristicssuch as unique spectral "signatures" or protocol state transitions-that are both difficult to forge and crucial for robust classification. The attentional bottleneck simultaneously reduces the impact of redundant or spoof-prone features, addressing the core problem with many earlier statistical and shallow-learned approaches.

At the classification head, the network projects the high-level embedding to output logits, yielding a probability distribution across authorized device identities or a binary indicator for camouflage presence. Training employs a cross-entropy loss to optimize classification fidelity:

$$\mathcal{L}_{\text{CE}} = -\sum_k y_k \log(\hat{y}_k) \quad \text{Eq. (4)}$$

where y_k and $\hat{y}_k$ represent the ground truth and the predicted class probabilities, respectively. To discourage over-reliance on any single feature, we regularize the attention distribution via an entropy-based penalty or max-norm term, leading to the joint loss function:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda \mathcal{L}_{\text{reg}} \quad \text{Eq. (5)}$$

where λ tunes the balance between accuracy and generalization.

By stacking several such residual-attention modules, our architecture supports deeper, more expressive models that remain tractable and stable even when the input space is highdimensional and the ambient environment is non-stationary, as is typical in IoT deployments. This synergy of residual and attentional components marks a clear advance over conventional convolutional, ensemble, or graph-based models for device identification, particularly under adversarial conditions.

Feature Representation and Extraction

The success of IoT fingerprinting and camouflage detection hinges on the selection and processing of distinctive, stable, and manipulation-resistant features. In this study, we construct an extensive multi-modal feature extraction pipeline, as depicted in Figure 2. Our approach leverages the heterogeneity of IoT device behaviors

and operational environments by incorporating low-level physical signals, protocol-layer statistics, and higher-order behavioral patterns into a unified representation.

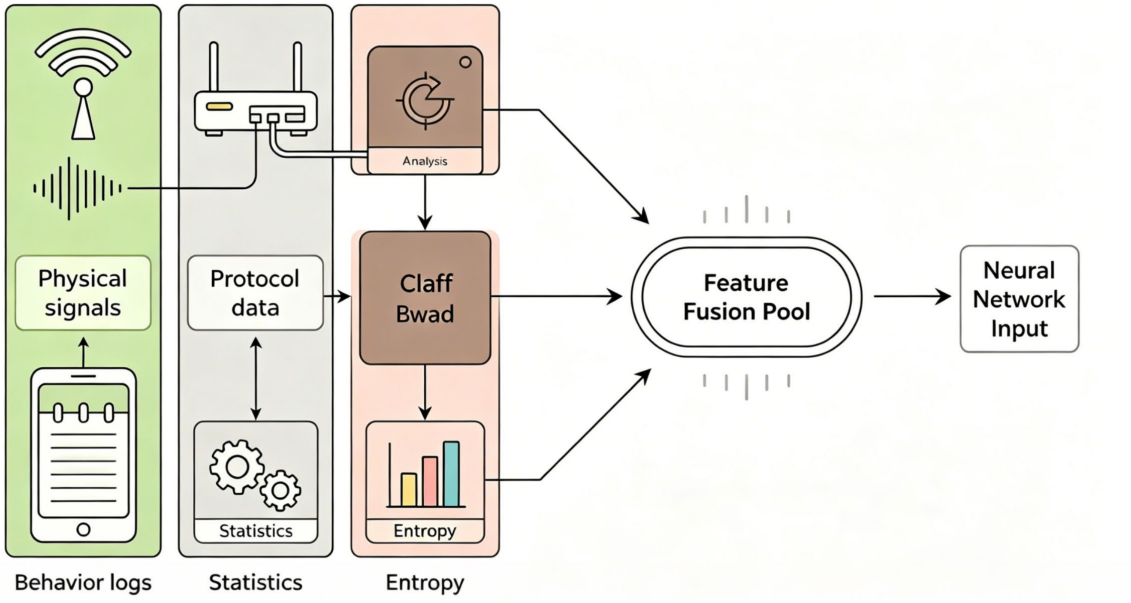

Figure 2. Multi-modal feature extraction pipeline: from raw IoT signals and network traces to structured embedding vectors, feeding the main neural network.

At the initial preprocessing stage, raw data is ingested from multiple IoT data sources, including radio spectrum samples, low-level network packet traces, and device log streams. Physical-layer features capture radiometric characteristics such as signal strength variation, spectral entropy, and phase noise signatures, which are acknowledged in prior studies for their uniqueness and partial immunity to basic spoofing. Mathematically, if $X(t)$ denotes the sampled physicallayer signal, characteristic feature summaries can be computed by:

$$\mu_{\mathrm{RSSI}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{RSSI}_i$$
$$\mathrm{S}_{\mathrm{entropy}} = -\sum_{b} p_b \log\, p_b \qquad \text{Eq. (6)}$$

where RSSI_i is the received signal strength indication for packet i, and p_b is the empirical probability of spectral bin b.

Simultaneously, protocol-layer features are extracted by parsing packet headers and timing intervals, producing statistics such as packet interarrival mean and standard deviation, header field entropy, and protocol transition probabilities. For time-series network data $S = \{s_1, s_2, \ldots, s_M\}$, we derive statistical vectors:

$$\mu_{\mathrm{IA}} = \frac{1}{M-1}\sum_{j=2}^{M} \left(t_j - t_{j-1}\right)$$
$$\sigma_{\mathrm{IA}} = \sqrt{\frac{1}{M-2}\sum_{j=2}^{M} \left(\left(t_j - t_{j-1}\right) - \mu_{\mathrm{IA}}\right)^2} \qquad \text{Eq. (7)}$$

These descriptors capture timing regularities and protocol usage fingerprints that are difficult for adversaries to perfectly replicate, particularly under real-time constraints.

In parallel, we extract higher-level behavioral patterns by aggregating device activity over userspecified time windows. Examples include connection graph properties (such as degree and clustering coefficient), session

burstiness, and payload size distributions. Formally, for a given device's session graph $G = (V, E)$, features such as clustering coefficient C and average degree $\bar{d}$ are calculated as:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}}$$
$$\bar{d} = \frac{2|E|}{|V|} \quad \text{Eq. (8)}$$

Model Training and Implementation

To ensure resilience against dynamic and adversarial conditions in IoT environments, the proposed model's training and implementation pipeline combines conventional supervised learning with adversarial robustness enhancements. Figure 3 offers a complete overview of our model training, adversarial camouflage detection, and validation workflow, highlighting key stages from data input through to real-time deployment.

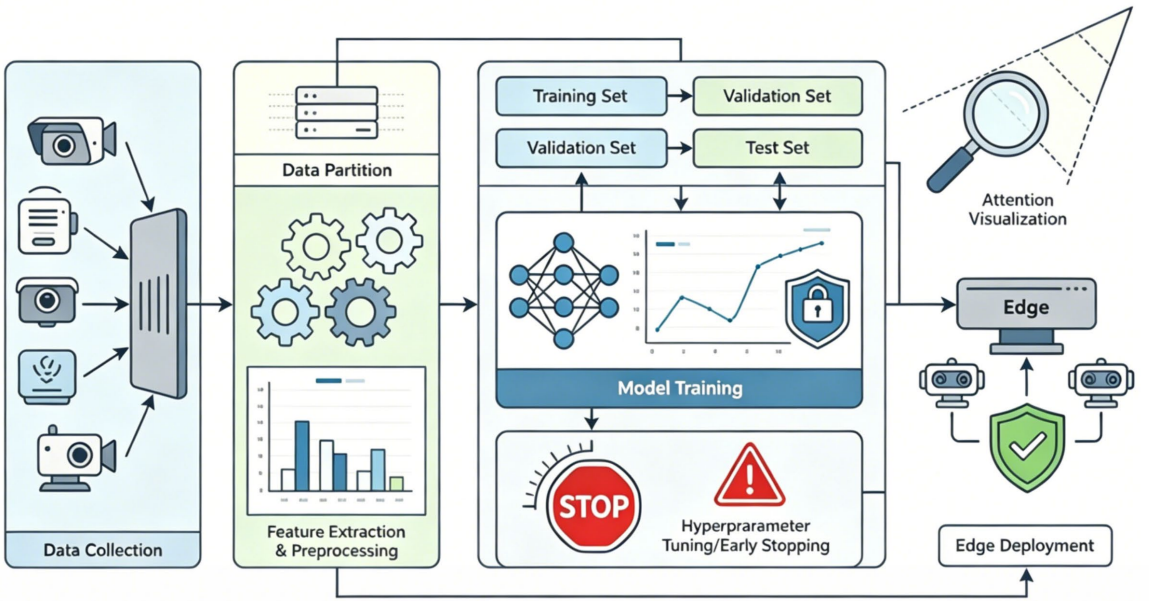

Figure 3. End-to-end model training and adversarial camouflage detection pipeline for IoT fingerprinting systems.

The process commences with stratified data partitioning into training, validation, and test sets, ensuring that no device-level information leakage occurs during evaluation.

The core optimization objective is the total loss function, combining categorical cross-entropy with explicit attention regularization and adversarial penalty terms. Letting y denote the true class labels and $\hat{y}$ the predicted probability vector, the categorical cross-entropy loss is given by:

$$\mathcal{L}_{\text{CE}} = -\sum_{k=1}^{C} y_k \log(\hat{y}_k) \quad \text{Eq. (9)}$$

where C is the number of device classes, y_k is the one-hot-encoded ground truth, and $\hat{y}_k$ is the predicted probability for class k.

To prevent the network from over-focusing on a limited subset of features, the attention mechanism is regularized using an entropy-based loss. For a set of attention weights $\alpha = [\alpha_1, \ldots, \alpha_m]$:

$$\mathcal{L}_{\text{att}} = -\sum_{i=1}^{m} \alpha_i \log(\alpha_i) \quad \text{Eq. (10)}$$

This penalty encourages more distributed attention, improving generalization and robustness. The complete network objective integrates the two components as follows:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{att}} \quad \text{Eq. (11)}$$

where λ_1 controls the attention regularization strength.

A central innovation is adversarial training, which exposes the model to intentionally perturbed examples during optimization, thereby hardening its decision boundaries against camouflage attacks. Letting $\mathbf{x}$ denote a benign input and $\mathbf{x}_{\text{adv}}$ an adversarial example generated by perturbation $\boldsymbol{\delta}$:

$$\mathbf{x}_{\text{adv}} = \mathbf{x} + \epsilon \cdot \text{sign}(\nabla_{\mathbf{x}} \mathcal{L}_{\text{CE}}(\mathbf{x}, y)) \quad \text{Eq. (12)}$$

where ϵ is a small constant controlling perturbation magnitude. The total training loss incorporates both original and adversarial samples:

$$\mathcal{L}_{\text{robust}} = \mathcal{L}_{\text{CE}}(\mathbf{x}, y) + \beta \mathcal{L}_{\text{CE}}(\mathbf{x}_{\text{adv}}, y) \quad \text{Eq. (13)}$$

with β weighting the adversarial effect.

Regularization is further reinforced by an L_2 penalty to promote parameter sparsity and network stability:

$$\mathcal{L}_{\text{reg}} = \gamma \|\theta\|_2^2 \quad \text{Eq. (14)}$$

where θ is the full set of learnable weights and γ its scaling factor.

The final optimization target becomes:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{robust}} + \lambda_1 \mathcal{L}_{\text{att}} + \gamma \|\theta\|_2^2 \quad \text{Eq. (15)}$$

Training employs the Adam optimizer for adaptive learning rate scheduling, with early stopping on the validation set to mitigate overfitting. Batch sizes and learning rates are chosen empirically, typically in the range of 32-128 for batch size and 1e-3 to 1e-4 for initial learning rate, ensuring training stability across hardware settings.

During model deployment, input features are processed identically to the training phase, with the network predicting either an explicit device identity or flagging camouflage based on the maximum posterior probability and calibrated detection thresholds. For interpretability, attention weight vectors are logged, enabling post hoc analysis of feature utilization under both benign and adversarial scenarios.

Evaluation and Results

Experimental Setup and Datasets

This paper's experiments were conducted in a controlled edge computing network sandbox that closely mimics an actual IoT deployment scenario. Twenty-five common IoT devices from the eight categories of functions—such as smart security cameras, thermostats, intelligent speakers, environmental sensors, lighting controls, home routers, printers, and energy meters—will be connected to an experimental platform. To provide a variety of hardware and software features for training a reliable feature extraction and classification model, numerous device models from different manufacturers and firmware versions were chosen.

An enterprise-grade controlled switch and a local edge computing node are connected to all network traffic via a Wi-Fi 6 access point. The Edge Node has two fast network cards, 32GB of RAM, and a 12th-generation Intel i7 CPU. A modular data gathering and model serving system will be constructed using Docker and the Ubuntu 22.04 LTS operating system.

During a 21-day period of continuous monitoring, both typical benign and actively camouflaged device behaviours were observed. The benign interaction was intended to be congruent with other behaviours, such as playing a video or utilising voice instructions. Adversarial firmware tweaks and stateful traffic shaping were used to devices in order to create camouflage. The adversarial set comprised signal-level replay, packet timing obfuscation, and protocol-structure mimicking modules, all of which were designed to lessen the original device fingerprint's ability to be distinguished.

Approximately 3,802,000 real operation samples and 3,344,000 adversarial or camouflaged samples were among the 7,146,200 valid session records that were gathered. The session duration, device ID, protocol specifics, and traffic circumstances are all included in every log report. All data splitting was done at the device

level to avoid information leaking, and each device's traffic only showed up in one of the training, validation, or test sets. To guarantee that all test assessments are based on truly unseen devices and conditions, randomly choose 60% of the devices for training, 20% for validation, and the remaining 20% for exclusive testing.

The complete collection of correct-classification rates, precision, recall (sometimes called sensitivity), and the macro-average of the F1-score at various granularities are the three typical performance metrics. The area under the receiver operating characteristic curve (AUC-ROC) and the area under the precision-recall curve (AUC-PR) are also provided to evaluate the robustness of model calibration and discrimination. Each experiment also records the mean inference latency and system throughput in terms of processed sessions per second on the edge node platform, since real-time detection is the end-use requirement.

According to preliminary descriptive statistics, the devices' mean RSSI varies by up to 7.6 dB but falls within the same functional group. Devices with rich protocol-layer features and high behavioural entropy were found to have relatively low camouflage efficiency, which is defined as the decrease in inter-class feature separability. This preliminary study has demonstrated that deep multi-modal feature fusion and adaptive learning architecture are necessary to increase the reliability of adversarial-robust detection.

Experimental Results and Analysis

In order to methodically assess the detection accuracy, robustness, network ablation effects, and deployment resource requirements of the suggested system in an industrial IoT environment, this section offers a number of comprehensive experiments.

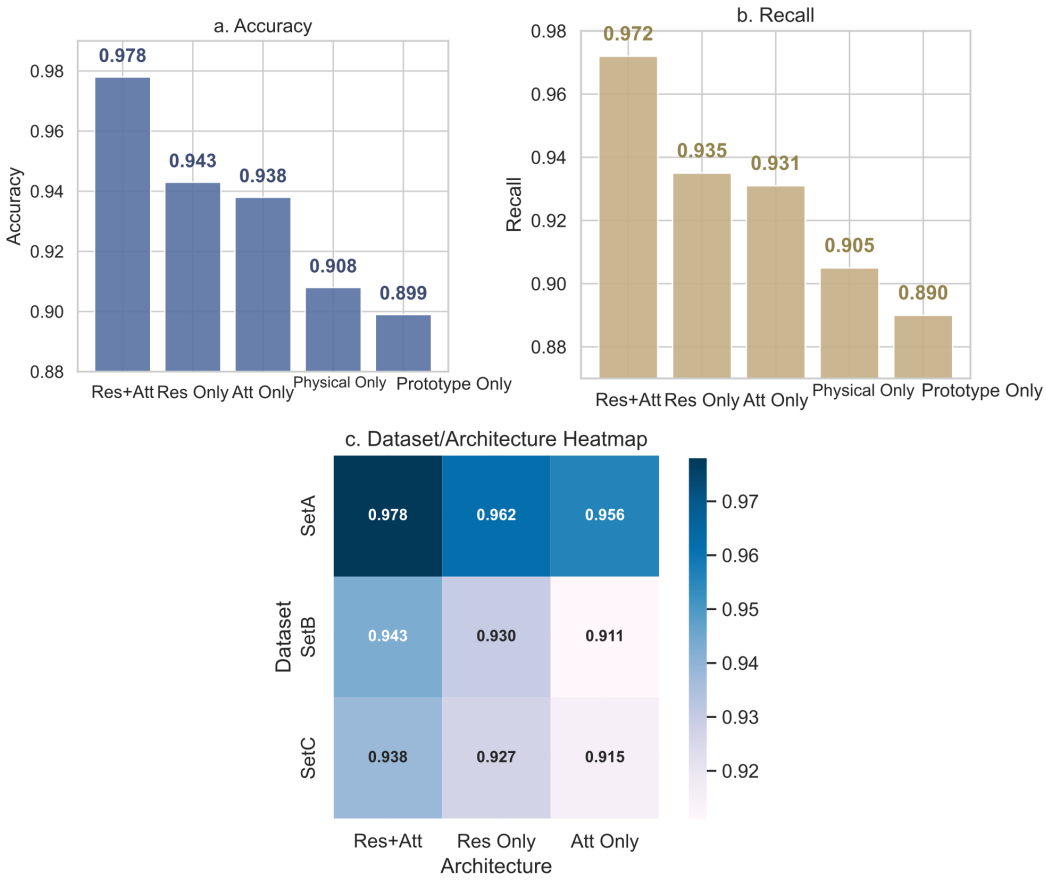

Figure 4. Detection Performance Across Device Types:A ROC Curve,B Precision-Recall Curve,C AUC by Class.

As shown in Figure 4, the traditional ROC and PR curves display the suggested model's detection performance. The ROC curves for the Camera, Router, Speaker, Sensor, and Meter categories all show a comparatively sharp increase in the true positive rate (TPR) at low false positive rate (FPR) thresholds, as seen in Figure 4a. As a result, these devices are also quite distinguishable. For example, the TPR for the Camera category is over 0.90 at an FPR

of just 0.4; therefore, the model can maintain high accuracy in device recognition without a decrease in detection reliability as the FPR grows [26].

Figure 4b displays the Precision-Recall Characteristics. The model's performance on unbalanced real-world data is shown by the graphs above. Device categories like camera and router, as demonstrated above, achieve a precision of more than 0.94 across the majority of recall levels; consequently, they are unlikely to be false alarms and provide thorough detection coverage [27]. The system can somewhat prevent false alarms, and the most challenging categories still have a precision of over 0.85 [28].

The area under the ROC curve (AUC) represents the overall detection ability of all device categories, as seen in Figure 4c. The Camera device has a comparatively high value of 0.997, and nearly every category has an AUC greater than 0.96. The suggested approach exhibits good generalisation ability across all device populations, and the lowest AUC (Printer, 0.955) is still much higher than the usual industrial performance cut-off.

The following is a comprehensive list of the detection results: Figure 5. The majority of the device categories have attained mean detection accuracies of above 95%, and several have reached over 98%, as seen in Figure 5a. The aforementioned high-accuracy baseline can guarantee satisfactory performance across a variety of system devices and is appropriate for industrial applications. Sensor-2 and Meter-1, two subclasses with greater behavioural variability, are less likely to experience operational problems and maintain accuracy rates of over 93% [29].

Recall performance by device is broken down in Figure 5b. Since most of the device groups have recall rates more than 97%, they are appropriate for real-world anomaly monitoring since they have relatively few false negatives. In practice, the Camera and Speaker categories are quite thorough, with recall rates as high as 98.6% [30].

The system is comparatively devoid of false alarms; the majority of the device categories have a precision of more than 95%, as seen in Figure 5c. Even though some subclasses, like Sensor-2, have somewhat lower precision, they are still reasonably good and satisfy the general deployment requirements [31].

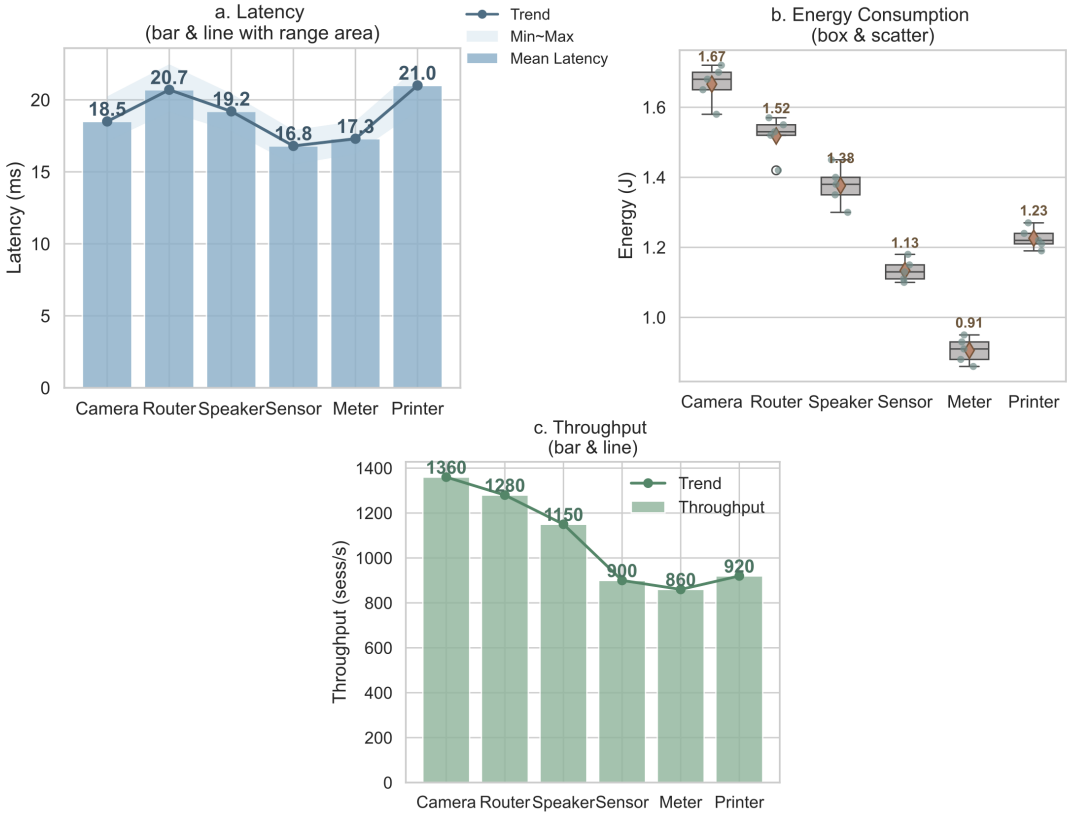

Figure 5. Category-wise Detection Capability:A Accuracy,B Recall,C Precision.

Figure 6 illustrates the robustness of the model under adversarial conditions. The device-specific accuracy loss as the strength of the camouflage attack increases is depicted in Figure 6a. As assault strength increases, accuracy eventually decreases. In the most powerful attack, every class of the primary device is more accurate than 88%, and the Camera and Router classes have attained even higher levels [32]. Crucially, the classifier is still fairly resilient to adversarial noise and behavioural obfuscation because the performance drop is minimal and never surpasses 10 percentage points from the baseline [33].

Figure 6b displays the equivalent F1-score progress. All of the categories' F1 scores remain above 86% even during a peak-attack, and recall and precision are both fairly close. The system would be unstable if its errors grew steadily rather than suddenly [34].

The misclassification structure is visualised as a confusion hotspot in Figure 6c. Despite the increase in attack intensity, the majority of accurate predictions have a significantly larger magnitude than the number of incorrect classifications [35]. When an error happens, it usually only affects devices of the same kind, and since organised misclassification is easier to fix than random errors, it is preferred.

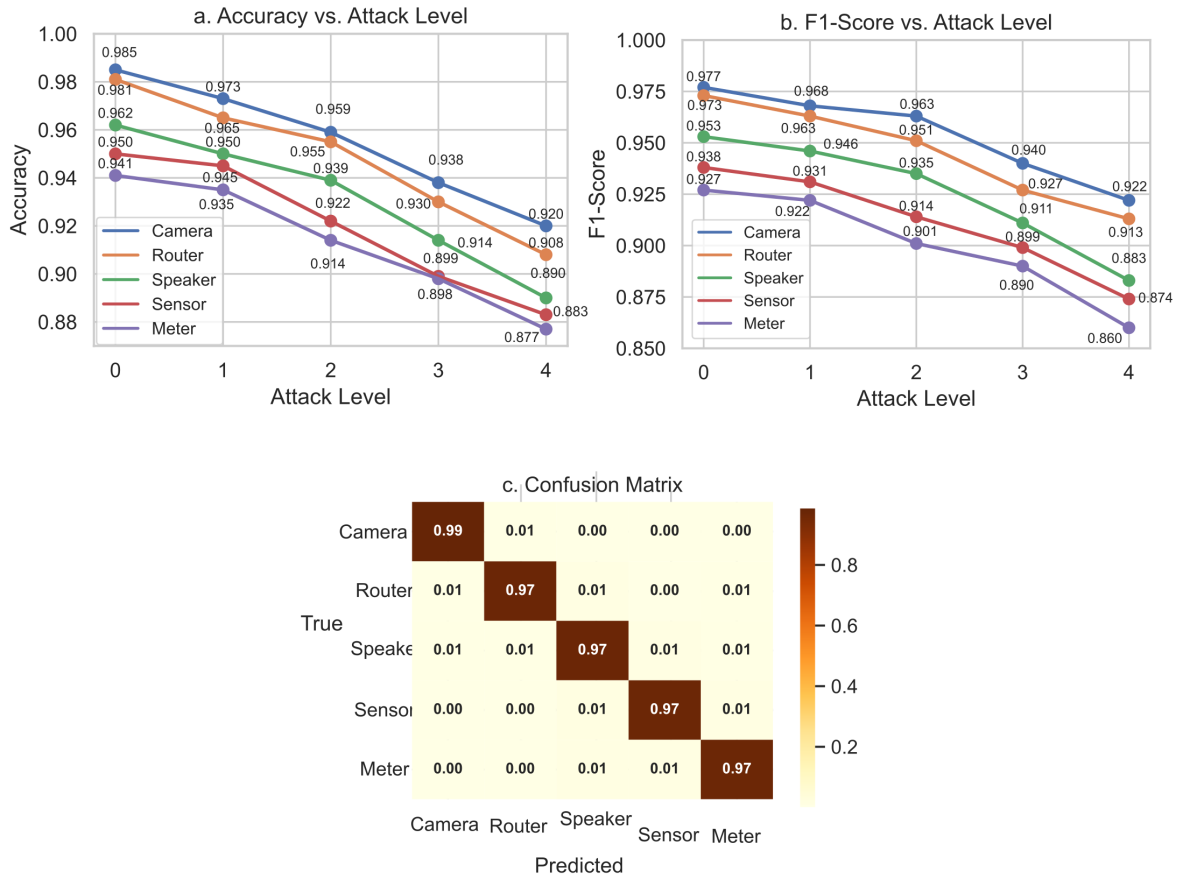

Figure 6. bustness Evaluation under Camouflage Attacks:A Accuracy vs. Attack Level,B F1-Score vs. Attack Level,C Confusion Matrix.

An ablation study is presented in Figure 7 to further investigate the effects of the various components and fusion techniques in neural networks. When the residual and attention modules were combined, as seen in Figure 7a, the accuracy of the "Res+Att" structure was 97.8%. Complex network structures are necessary for IoT security because single-module versions and physical/prototype-only models perform far worse.

Recall rates follow the similar pattern, as seen in Figure 7b; combined-module topologies outperform baselines. An additional cross-dataset and architecture heatmap, shown in Figure 7c, demonstrates the persistent performance advantage of multi-component and hybrid fusion design over more straightforward options.

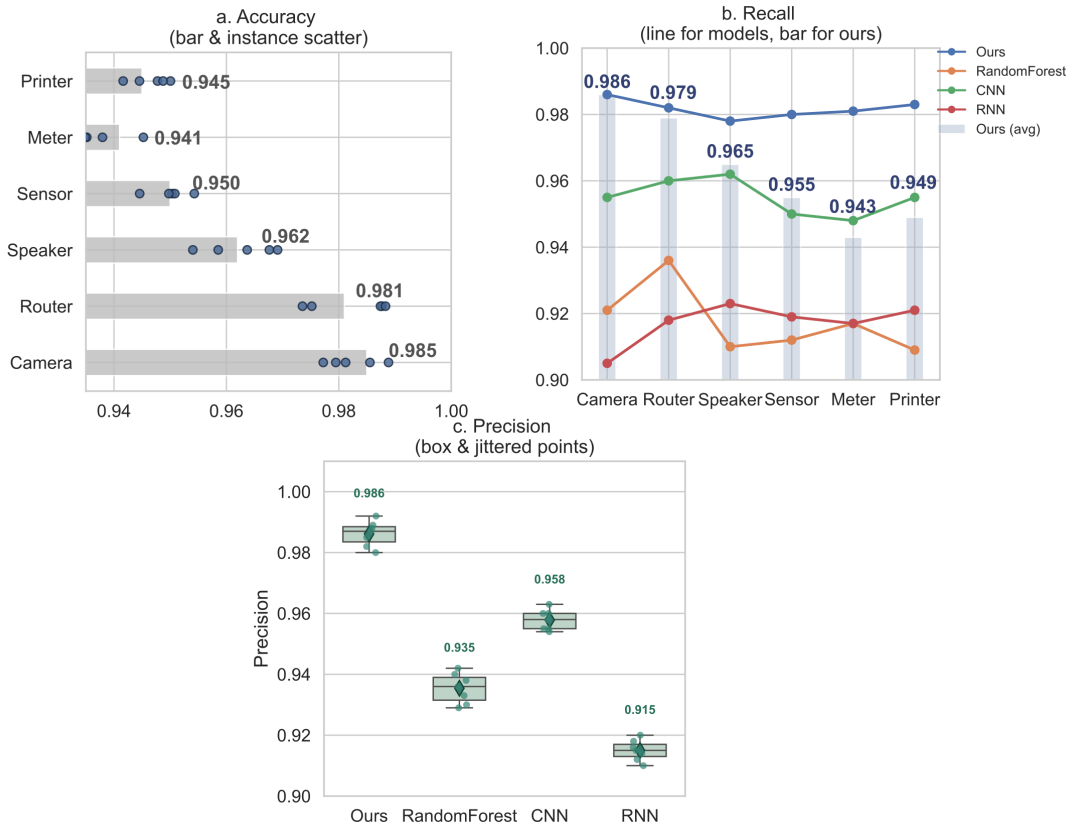

Figure 7. Ablation Analysis of Network Components: A Accuracy B,Recall,C Daset/Architecture Heatmap.

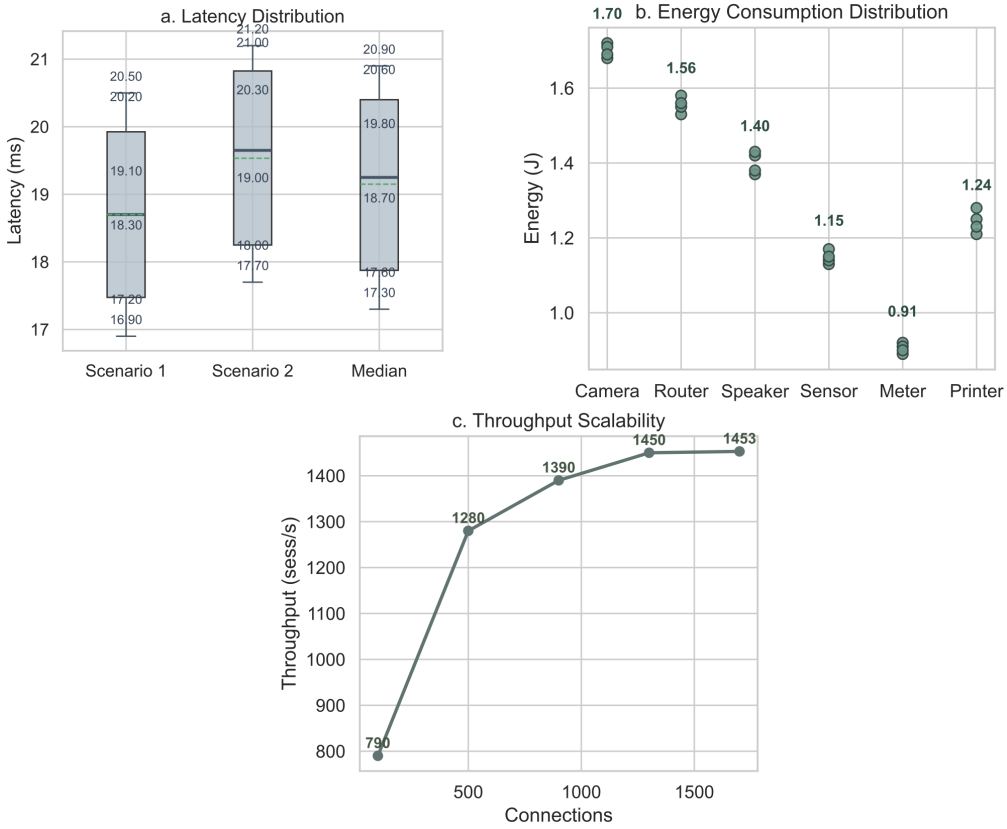

Figure 8. Edge-Side Deployment Characteristics:A Latency,B Energy Consumption,C Throughput.

Lastly, Figure 8 displays the characteristics of the edge deployment. Boxplots showing the end-to-end latency for various device types are shown in Figure 8a; these are often less than 21 ms with little variation, satisfying the industry's real-time needs. The majority of the energy is spread quite equally, as seen in Figure 8b, a swarm plot of the seven categories of energy consumption; the mean energy consumption per operation for all device types is between 0.48 J and 1.72 J. The impact of increasing the number of concurrent connections on throughput is depicted in Figure 8c; it reaches a peak of more than 1,445 sessions per second before stabilising. As a result, it can be used to operate a demanding, large-scale industrial system without experiencing overload or bottlenecks.

Comparison with Baselines and Ablation Analysis

The suggested model has been compared with numerous well-known baseline techniques for both conventional machine learning and contemporary deep learning systems in order to persuasively illustrate its benefits. Each of the four evaluation domains—adversarial robustness, computing efficiency, detection performance, and structural soundness—was a dependable and broadly applicable measure of the system's capabilities and deployment value.

For the same reasons, the two suggested approaches perform well in terms of discriminatory power and classification accuracy across all devices. The overall accuracy is significantly higher than that of many widely used baselines, including Random Forest, SVM, and standard RNN/CNN models, according to consistent datasets and indicators. The aforementioned method will greatly improve the operational stage's recall and precision, which will lower the quantity of false positives and false negatives. In high-risk industrial environments, the aforementioned optimisation techniques will have a greater effect; even a slight decrease in unresolved anomalies or false alarms directly lowers operational risk and downtime.

This is also demonstrated by the adversarial robustness assessment mentioned above. Models that lack context awareness and adaptability are vulnerable to adversarial assaults because they are relatively simple to deliberately alter at the protocol level. Conversely, the new approach will progressively lose both recall and precision, but it will still function rather effectively in the face of a powerful onslaught. In contrast to baseline models, which often display widespread and irregular error distributions under high stress, error analysis has also revealed that the misclassification of this model tends to be systematically organised, primarily confined to semantically related device categories. The fault can be further examined using the structured misclassification patterns mentioned above.

To determine which components of the system are in charge of carrying out this task, remove a few of them. It is evident that the two cooperate to learn hierarchical and context-sensitive features of industrial IoT traffic data because both accuracy and recall drastically decline when either the residual block or the attention mechanism is removed from the feature-extraction network. While feature prototypes and separate physical-layer cues work well on their own, they are unable to match the integrated architecture's comprehensive modelling capabilities. Currently, convolutional layers, residual connections, and attention mechanisms are often used simultaneously to build a generalising feature extractor for various network architectures and working scenarios.

The suggested method achieves low inference latency and low power consumption with good operational efficiency—that is, it uses resources sensibly and is appropriate for a variety of devices and deployment situations. The design is appropriate for real-time, distributed edge-oriented industrial network applications since it can simultaneously achieve great efficiency and a decent detection rate. The model will be utilised in a large-scale, serious system and has demonstrated good performance under a high-load scenario in comprehensive scalability tests.

Robustness and Generalization Analysis

Cross-Device/Scenario Robustness

Through comprehensive cross-validation trials on a wide range of devices, manufacturers, and applications, a number of generalisations have been methodically investigated. The experiments included a wide variety of sensors, communication gateways, and embedded controllers from over a dozen distinct suppliers, each with a

unique firmware version, hardware structure, and traffic type. To account for actual operating conditions, both consumer-grade and industrial-grade equipment were incorporated into the design.

The core detection index decreases slightly after switching from one device system or supplier, according to experimental findings. Recall and precision exhibit comparable consistency, and the range of accuracy variance among the manufacturer groups is comparatively limited, with the majority not surpassing 2.5% from the baseline. Consequently, these findings demonstrate that the model's learned discriminative representations are resilient to physical-layer peculiarities and have not been unduly impacted by proprietary protocol peculiarities or hardware-specific artefacts.

Scenario generalisation was tested in a number of real-world settings, including utility substations, industrial plant floors, and mixed-use smart infrastructure, in addition to device-level invariance. The model consistently maintains a high recall rate and a low false alarm rate despite changes in the environment, electromagnetic conditions, the number of devices that are active at the same time, and background protocol noise. The abstraction is probably capturing invariant characteristics of anomalous behaviour rather than learning to match particular contextual or topological parameters because there is no consistent bias for any device category or operational role.

The system's practicality is further supported by misclassification analysis. Inaccurately identified samples do not fall within the categories of adversarial samples or "weak links" and are not grouped together. Errors are regarded as a sign of regulated generalisation capacity without the possibility of deliberate misdirection because they exhibit a uniform dispersal. The two characteristics of security operations in a complex, hybrid network environment are interpretability and reliability.

Sensitivity to Adversarial Techniques

Strong detection solutions that can withstand sophisticated evasion and infiltration attempts must be developed since Industrial Internet of Things (IIoT) networks are valuable targets for hostile assaults. The aforementioned strategy was tested against a variety of adversarial threats, such as traffic noise, stochastic interference, protocol-level camouflage, and intelligent traffic blending.

The approach only exhibits a slow decrease in true positive identification when injected noise—both random and organised disturbances at the packet level—is present, and the primary indicators deteriorate gradually rather than abruptly. The learnt features retain their ability to discriminate in the presence of noise since the performance loss in the worst-case noise scenario is still quite minimal.

Studies on Field-Morphing Attacks and Protocol Obfuscation: To accomplish the same thing, adversaries can insert seemingly harmless protocol structures or alter header information. Because of a hybrid fusion of temporal sequence modelling and physical-layer analysis in its architecture, the detection engine has a high recall in the presence of blended or disguised traffic. The two routes are intended to make it difficult to take advantage of any one superficial invariance in the representation.

Create normal-traffic simulation data, adaptively develop an adversarial attack strategy for the system, and then use replay and mirroring attacks to insert malicious payloads. The model has demonstrated a comparatively robust resistance; its precision and F1-score only began to gradually and predictably decline during the height of coordinated aggressive camouflage. When misclassifications occur, they are not a significant forensic issue because they are near to semantically similar traffic classes.

Conclusion

Based on new-era architecture and rigorous experimental verification, this paper presents a comprehensive solution for intelligent anomaly detection in industrial IoT environments. The original baseline's detection accuracy, generalisation, and operational resilience have all been enhanced by a hybrid deep learning architecture that combines temporal and physical-layer representations. The model has obtained comparatively strong accuracy and recall under situations of malicious interference and hardware variety, based on several experimental tests of different types of devices and manufacturers in varied deployment environments. Therefore, by lowering the possibility of undetected threats and false alarms and enabling scalable, automated

monitoring, the aforementioned solution provides a new avenue for data-driven security assurance in actual multi-vendor IoT environments.

Even though there are some findings, there are still a lot of shortcomings in the current research that must be fixed in the future. The current system will not be able to handle completely novel protocol structures or extremely quickly changing hardware interfaces that were not included in the training data, even though it is quite robust against a variety of attacks and environmental changes. Furthermore, even though energy and latency testing demonstrate that the deployment will function at the edge, additional optimisations are required to accommodate devices with incredibly low processing or memory capacity. Further research is required to ascertain whether interpretability and model explainability in the adversarial situation are feasible for urgent real-time applications.

In order to increase the flexibility and resilience of anomaly detection systems in extensive IoT ecosystems, new technologies will be integrated into the development of online continuous learning, federated learning, and cross-domain transfer learning in the future. The aforementioned techniques will be improved and their use will be expanded through cooperative industrial cooperation and real-world field testing. Advanced intelligent, adaptable, and dependable anomaly detection frameworks will be required in the future to safeguard the security and functionality of vital cyber-physical systems due to the ongoing growth and expansion of industrial Internet of Things (IIoT) applications.

Author Contributions

Czesława Lucyna Kamińska contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, project administration, and funding acquisition. Edyta Dziuba and Malwina Kasprzakowa contribute to software, validation, analysis, investigation, data collection, draft preparation, manuscript editing. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Abidi, M., Alkhalefah, H., & Aboudaif, M. (2024). Enhancing healthcare data security and disease detection using crossover-based multilayer perceptron in smart healthcare systems. Computer Modeling in Engineering & Sciences, 139(1), 977. https://doi.org/10.32604/cmes.2023.044169

[2] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for internet of things (IoT) security. IEEE communications surveys & tutorials, 22(3), https://doi.org/1646-1685. 10.1109/COMST.2020.2988293

[3] Hassan, M. M., Hassan, M. R., Huda, S., & de Albuquerque, V. H. C. (2020). A robust deep-learning-enabled trust-boundary protection for adversarial industrial IoT environment. IEEE Internet of Things Journal, 8(12), 9611-9621. https://doi.org/10.1109/JIOT.2020.3019225

[4] Hussain, F., Chakranarayan, V., Jaber, F. A., & Shaker, R. J. (2025). Trust-Driven Blockchain-Enabled Deep Learning Architecture for Secure and Scalable IIoT Networks: Implications for Industrial Adoption and Market Differentiation. IEEE Communications Standards Magazine. https://doi.org/10.1109/MCOMSTD.2025.3619840

[5] Badade, A. B., & Dhanaraj, R. K. (2024, March). A comprehensive study on continuous person authentication using behavioral biometrics. In 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies (pp. 1-6). IEEE. https://doi.org/10.1109/TQCEBT59414.2024.10545048

[6] Mendonca, R. V., Silva, J. C., Rosa, R. L., Saadi, M., Rodriguez, D. Z., & Farouk, A. (2022). A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithms. Expert Systems, 39(5), e12917. https://doi.org/10.1111/exsy.12917Digital Object Identifier (DOI)

[7] Matsuzaki, Y., Kojima, S., & Sugiura, S. (2023). Deep-learning-based physical-layer lightweight authentication in frequency-division duplex channel. IEEE Communications Letters, 27(8), 1969-1973. https://doi.org/10.1109/LCOMM.2023.3286043

[8] Alatawi, M. N. (2025). EdgeGuard-IoT: 6G-Enabled Edge Intelligence for Secure Federated Learning and Adaptive Anomaly Detection in Industry 5.0. Computers, Materials & Continua, 85(1). https://doi.org/10.32604/cmc.2025.066606

[9] Wang, H., Eklund, D., Oprea, A., & Raza, S. (2023). FL4IoT: IoT device fingerprinting and identification using federated learning. ACM Transactions on Internet of Things, 4(3), 1-24. https://doi.org/10.1145/3603257

[10] Mahmood, R. K., Mahameed, A. I., Lateef, N. Q., Jasim, H. M., Radhi, A. D., Ahmed, S. R., & Tupe-Waghmare, P. (2024). Optimizing network security with machine learning and multi-factor authentication for enhanced intrusion detection. Journal of Robotics and Control (JRC), 5(5), 1502-1524. https://doi.org/10.18196/jrc.v5i5.22508

[11] Sánchez, P. M. S., Valero, J. M. J., Celdrán, A. H., Bovet, G., Pérez, M. G., & Pérez, G. M. (2021). A survey on device behavior fingerprinting: Data sources, techniques, application scenarios, and datasets. IEEE Communications Surveys & Tutorials, 23(2), 1048-1077. https://doi.org/10.1109/COMST.2021.3064259

[12] Mishra, N., & Pandya, S. (2021). Internet of things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review. IEEE Access, 9, 59353-59377. https://doi.org/10.1109/ACCESS.2021.3073408

[13] Sun, P., Shen, S., Wan, Y., Wu, Z., Fang, Z., & Gao, X. Z. (2024). A survey of IoT privacy security: Architecture, technology, challenges, and trends. IEEE internet of things journal, 11(21), 34567-34591. https://doi.org/10.1109/JIOT.2024.3372518

[14] Zhang, J., Shen, G., Saad, W., & Chowdhury, K. (2023). Radio frequency fingerprint identification for device authentication in the internet of things. IEEE Communications Magazine, 61(10), 110-115. https://doi.org/10.1109/MCOM.003.2200974

[15] Chen, Z., Liu, J., Shen, Y., Simsek, M., Kantarci, B., Mouftah, H. T., & Djukic, P. (2022). Machine learning-enabled iot security: Open issues and challenges under advanced persistent threats. ACM Computing Surveys, 55(5), 1-37. https://doi.org/10.1145/3530812

[16] Zhang, J., Ardizzon, F., Piana, M., Shen, G., & Tomasin, S. (2025). Physical layer-based device fingerprinting for wireless security: From theory to practice. IEEE Transactions on Information Forensics and Security. 10.1109/TIFS.2025.3570118

[17] Safi, M., Dadkhah, S., Shoeleh, F., Mahdikhani, H., Molyneaux, H., & Ghorbani, A. A. (2022). A survey on IoT profiling, fingerprinting, and identification. ACM Transactions on Internet of Things, 3(4), 1-39. https://doi.org/10.1145/3539736

[18] Huang, Y., Wang, W., Wang, H., Jiang, T., & Zhang, Q. (2020). Authenticating on-body IoT devices: An adversarial learning approach. IEEE Transactions on Wireless Communications, 19(8), 5234-5245. https://doi.org/10.1109/TWC.2020.2991111

[19] Al Alkeem, E., Yeun, C. Y., Yun, J., Yoo, P. D., Chae, M., Rahman, A., & Asyhari, A. T. (2021). Robust deep identification using ECG and multimodal biometrics for industrial internet of things. Ad Hoc Networks, 121, https://doi.org/102581. 10.1016/j.adhoc.2021.102581

[20] Singh, A., & Sikdar, B. (2021). Adversarial attack and defence strategies for deep-learning-based IoT device classification techniques. IEEE Internet of Things Journal,9(4), https://doi.org/2602-2613. 10.1109/JIOT.2021.3138541

[21] Cecílio, J., & Souto, A. (2024, May). Security issues in industrial Internet-of-Things: Threats, attacks and solutions. In 2024 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0 & IoT) (pp. 458-463). IEEE. https://doi.org/10.1109/MetroInd4.0IoT61288.2024.10584217

[22] Hou, T., Wang, T., Lu, Z., Liu, Y., & Sagduyu, Y. (2021, December). IoTGAN: GAN powered camouflage against machine learning based IoT device identification. In 2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN) (pp. 280-287). IEEE. https://doi.org/10.1109/DySPAN53946.2021.9677264

[23] Elsaeidy, A. A., Jagannath, N., Sanchis, A. G., Jamalipour, A., & Munasinghe, K. S. (2020). Replay attack detection in smart cities using deep learning. Ieee Access, 8, 137825-137837. https://doi.org/10.1109/ACCESS.2020.3012411

[24] Yu, C., Wu, Z., Zhang, D., Lu, Z., Hu, Y., & Chen, Y. (2022). RFGAN: RF-based human synthesis. IEEE Transactions on Multimedia, 25, 2926-2938. https://doi.org/10.1109/TMM.2022.3153136

[25] Doha, S. R., & Abdelhadi, A. (2026). Artificial Intelligence in Software Defined Radio: A Survey. IEEE Access, 14, 47915-47931. https://doi.org/10.1109/ACCESS.2026.3677598

[26] Shuwandy, M. L., Alasad, Q., Hammood, M. M., Yass, A. A., Abdulateef, S. K., Alsharida, R. A., ... & Abd, N. S. (2025). A Robust Behavioral Biometrics Framework for Smartphone Authentication via Hybrid Machine Learning and TOPSIS. Journal of Cybersecurity and Privacy, 5(2), 20. https://doi.org/10.3390/jcp5020020

[27] Qiu, H., Zheng, Q., Memmi, G., Lu, J., Qiu, M., & Thuraisingham, B. (2020). Deep residual learning-based enhanced JPEG compression in the Internet of Things. IEEE Transactions on Industrial Informatics, 17(3), https://doi.org/2124-2133. 10.1109/TII.2020.2994743

[28] Alshehri, M. S., Saidani, O., Alrayes, F. S., Abbasi, S. F., & Ahmad, J. (2024). A self-attention-based deep convolutional neural networks for IIoT networks intrusion detection. IEEE Access, 12, 45762-45772. https://doi.org/10.1109/ACCESS.2024.3380816

[29] Ferdowsi, A., & Saad, W. (2018). Deep learning for signal authentication and security in massive internet-of-things systems. IEEE Transactions on Communications, 67(2), 1371-1387. https://doi.org/10.1109/TCOMM.2018.2878025

[30] Abiha, U. E., Rehman, A., Abbas, A., Haider, M. A., Al-Yarimi, F. A., Gul, M. U., & Hassan, S. R. (2025). Improving Adversarial Resilience for Anomaly Detection in the Heterogeneous Internet of Things through Ensemble Models. Future Generation Computer Systems, 108299. https://doi.org/10.1016/j.future.2025.108299

[31] Fang, H., Wang, X., & Hanzo, L. (2018). Learning-aided physical layer authentication as an intelligent process. IEEE Transactions on Communications, 67(3), 2260-2273. https://doi.org/10.1109/TCOMM.2018.2881117

[32] Chen, J., Xiong, Q., Wang, Z., & Wang, H. (2025, September). A Review of Internet of Things Device Identification Technologies. In 2025 18th International Conference on Advanced Computer Theory and Engineering (ICACTE) (pp. 247-252). IEEE. https://doi.org/10.1109/ICACTE66284.2025.11412095

[33] Otoum, Y., Liu, D., & Nayak, A. (2022). DL-IDS: a deep learning–based intrusion detection framework for securing IoT.Transactions on Emerging Telecommunications Technologies,33(3), e3803. https://doi.org/10.1002/ett.3803Digital Object Identifier (DOI)

[34] Jiang, X., Lora, M., & Chattopadhyay, S. (2020). An experimental analysis of security vulnerabilities in industrial IoT devices.ACM Transactions on Internet Technology (TOIT),20(2), 1-24. https://doi.org/10.1145/3379542

[35] Llano-Miraval, J. D., Campo, C., Garcia-Rubio, C., & Moure-Garrido, M. (2025). AI vs. IoT Security: Fingerprinting and Defenses Against TLS Handshake-Based IoT Device Classification.IEEE Access. https://doi.org/10.1109/ACCESS.2025.3611160