*International Journal of Basic and Applied Sciences*, 15 (1) (2026) 161-176

International Journal of Basic and Applied Sciences

*Website: www.sciencepubco.com/index.php/IJBAS*
*https://doi.org/10.14419/xw7zzh51*
***Research paper***

# A Multi-Modal AI Framework for Intrusion Detection in Healthcare IoT

**Hani Al-Balasmeh [1] *, Fayzeh Abdulkareem Jaber [2], Sa'eed Serwan Abdulsattar [3]**

*[1] Dept. of Informatics Engineering, University of Technology Bahrain*
*[2] Dept of Computer Studies, University of Technology Bahrain*
*[3] Dept of Electrical and Electronics Engineering, University of Bahrain*
**Corresponding author E-mail: h.albalasmeh@utb.edu.bh*

*Received: December 31, 2025, Accepted: January 24, 2026, Published: January 26, 2026*

## Abstract

The rapid adoption of Healthcare Internet of Things (HIoT) technologies has increased the exposure of safety-critical medical infrastructures to sophisticated cyberattacks, while simultaneously imposing strict constraints on availability, privacy, and false alarm tolerance. Traditional network-centric intrusion detection systems (IDSs) struggle to reliably detect stealthy, low-rate attacks in healthcare environments, where encrypted traffic and predictable clinical communication patterns limit observability. This paper proposes a lightweight multimodal intrusion detection framework for HIoT networks that jointly analyzes network flow metadata and device telemetry using structured feature fusion and ensemble meta-learning. By integrating communication behavior with device-level operational regularity, the framework provides a more comprehensive representation of system activity without relying on payload inspection. An extensive evaluation was conducted using heterogeneous IoT and botnet datasets, including TON_IoT, IoT-23, and MedBIoT, under realistic conditions of class imbalance. Experi-mental results demonstrate that the proposed framework achieves 97.1% detection accuracy and 96.5% F1-score, outperforming state-of-the-art network-only IDS baselines. Most notably, the framework reduces the false positive rate to 2.4%, representing a relative reduction of over 38% compared to strong deep learning baselines. Attack-wise analysis shows the most significant performance gains for stealthy and low-rate attack classes, where network-centric IDSs exhibit substantial degradation. Ablation studies confirm that device telemetry im-proves recall for stealthy threats, while ensemble meta-learning stabilizes predictions and suppresses false alarms. Latency and scalability measurements further indicate that the framework remains suitable for real-time deployment at healthcare gateway nodes. Overall, the results provide strong empirical evidence that multi-modal intrusion detection materially improves detection reliability, clinical safety, and robust-ness in healthcare IoT environments, addressing key limitations of existing single-modality IDS approaches.

***Keywords***: *Healthcare Internet of Things (HIoT); Intrusion Detection System; Multi-Modal Security; Device Telemetry; Network Traffic Analysis; Ensemble Learning; False Positive Reduction; Cyber-Physical Systems Security.*

## 1. Introduction

The rapid integration of the Internet of Things (IoT) into modern healthcare environments has fundamentally transformed the delivery of medical services by enabling continuous patient monitoring, intelligent diagnostics, and seamless connectivity among medical devices. Healthcare Internet of Things (HIoT) networks now underpin mission-critical clinical functions, including remote vital-sign monitoring, infusion pump control, medical imaging systems, and coordinated emergency response workflows. While these technologies significantly enhance operational efficiency and the quality of care, they simultaneously expand the cyberattack surface of healthcare infrastructure due to their distributed architecture, device heterogeneity, and widespread reliance on legacy or resource-constrained hardware platforms [1], [2].

Unlike conventional enterprise networks, HIoT environments operate under stringent constraints on availability, reliability, and safety, where even brief service interruptions can have direct, potentially life-threatening consequences. Many medical IoT devices are deployed with limited computational resources and minimal built-in security mechanisms, rendering them particularly vulnerable to cyber threats such as botnet infiltration, denial-of-service (DoS) attacks, ransomware propagation, and unauthorized data exfiltration [3], [4]. Recent real-world incidents have demonstrated that cyber intrusions targeting hospital networks can disrupt clinical workflows, delay critical treatments, and compromise sensitive patient data, thereby posing severe risks to patient safety and institutional trust [5]. These realities underscore the urgent need for intrusion detection solutions that are specifically designed for the unique operational and security requirements of healthcare IoT ecosystems.

Traditional intrusion detection systems (IDS), including signature-based and rule-based approaches, have proven inadequate for protecting IoT environments. Signature-based IDSs are inherently ineffective against zero-day and polymorphic attacks. At the same time, rule-based systems require continuous manual updates and expert intervention, which is impractical in large-scale, dynamic healthcare deployments

[6]. Consequently, artificial intelligence (AI) and machine learning (ML) techniques have emerged as promising alternatives, offering adaptive learning capabilities and improved detection of previously unseen attack patterns [7], [8].

Despite their potential, existing AI-based intrusion detection solutions for IoT and healthcare networks suffer from critical limitations. Most proposed approaches rely predominantly on network-level traffic features—such as packet statistics, flow durations, or protocol distributions—while largely neglecting device-level behavioral and operational characteristics [9]. In healthcare settings, however, malicious activities often manifest through subtle deviations in device behavior, including irregular communication periodicity, abnormal operational states, or unexpected resource utilization, without immediately producing conspicuous network anomalies. As a result, single-modality detection strategies are prone to reduced detection accuracy and elevated false positive rates, which can overwhelm security teams and disrupt sensitive clinical operations [10].

Moreover, healthcare IoT environments exhibit extreme heterogeneity, encompassing wearable sensors, bedside monitoring devices, imaging systems, implantable devices, and mobile medical equipment from diverse vendors and manufacturers. This diversity leads to highly variable traffic patterns and operational behaviors across devices and departments, significantly challenging the generalization capability of conventional intrusion detection models [11]. In such safety-critical environments, high false alarm rates are particularly problematic, as unnecessary device isolation or network restrictions may interfere with patient care and compromise clinical safety.

To overcome these challenges, this paper adopts a multi-modal intrusion detection paradigm that jointly analyzes network-level communication behavior and device-level telemetry information. By integrating these complementary perspectives, the proposed approach constructs a more comprehensive and context-aware representation of system activity, enabling more accurate discrimination between benign clinical behavior and malicious activity. The framework is designed to enhance robustness against traffic variability, device heterogeneity, and stealthy attack strategies, while maintaining low false-positive rates, which are essential for healthcare environments. Importantly, the proposed detection architecture is lightweight and suitable for deployment at healthcare gateway nodes, ensuring real-time operation without imposing excessive computational or privacy overhead.

The remainder of this paper is organized as follows. Section II reviews related work on AI-based intrusion detection for IoT and healthcare networks. Section III presents the system and threat models. Section IV introduces the proposed multi-modal intrusion detection framework and feature fusion strategy. Section V describes the experimental setup and evaluation methodology. Section VI provides an in-depth analysis of experimental results and comparative security performance, and Section VII concludes the paper with key findings and directions for future research.

## 2. Literature Review

Recent advances in intrusion detection for Healthcare Internet of Things (HIoT) networks reflect growing recognition that conventional security mechanisms are insufficient to protect safety-critical medical infrastructure. Unlike generic IoT deployments, healthcare environments impose strict constraints on availability, latency, and false-alarm tolerance while operating with heterogeneous, often resource-limited medical devices. Existing research in this domain can be broadly categorized into four streams: (i) traditional intrusion detection systems, (ii) AI-based network-centric detection, (iii) device telemetry and behavioral analysis, and (iv) emerging multi-modal intrusion detection approaches. This section reviews recent contributions in each stream and highlights the limitations that motivate this work.

### 2.1. Traditional intrusion detection in healthcare

Signature-based and rule-driven intrusion detection systems, such as Snort and Suricata, have historically been deployed in hospital networks due to their low implementation overhead and deterministic operation [1]. However, multiple studies have demonstrated that these systems are poorly suited to healthcare IoT environments, where encrypted traffic, proprietary medical protocols, and dynamic clinical workflows significantly reduce detection effectiveness [2], [3].

More critically, signature-based IDSs are inherently incapable of detecting zero-day and polymorphic attacks, which pose a substantial threat to medical devices characterized by long operational lifecycles and infrequent firmware updates [4]. As a result, traditional IDS solutions are increasingly regarded as inadequate for securing modern HIoT infrastructures, particularly in large hospitals and distributed care settings.

### 2.2. AI-based network-centric intrusion detection

To overcome the rigidity of rule-based systems, recent research has widely adopted machine learning and deep learning techniques for intrusion detection. Supervised learning approaches, including Random Forests, Gradient Boosting, and Support Vector Machines, have demonstrated strong performance on benchmark intrusion datasets [5], [6]. Deep learning models—most notably convolutional neural networks (CNNs) and recurrent neural networks (RNNs)—have further improved detection accuracy by modeling spatial and temporal traffic patterns [7].

Since 2022, several studies have applied deep learning to intrusion detection in IoT and healthcare-oriented contexts. Transformer-based models and attention mechanisms have been proposed to analyze encrypted IoT traffic and capture long-range dependencies [8]. At the same time, hybrid CNN–LSTM architectures have been evaluated in healthcare network scenarios [9]. Despite these advances, the majority of such approaches remain network-centric and rely exclusively on traffic features. Consequently, they often suffer from elevated false positive rates and limited robustness when attackers mimic legitimate clinical traffic patterns. Moreover, the computational complexity of deep models restricts their deployment at resource-constrained healthcare gateways.

### 2.3. Device telemetry and behavioral analysis

Recognizing that many IoT attacks manifest through deviations in device operation rather than overt traffic anomalies, recent work has explored telemetry-driven intrusion detection techniques. These approaches analyze device-level indicators such as transmission periodicity, resource utilization, communication stability, and operational state transitions to identify compromised devices [10], [11]. In healthcare environments, where medical devices typically exhibit highly predictable behavior, behavioral profiling has proven effective for detecting botnet participation and insider threats [12].

However, telemetry-only detection suffers from notable limitations. Device heterogeneity, vendor-specific telemetry formats, and inconsistent data availability reduce model generalizability across healthcare deployments. Furthermore, attacks that primarily exploit network-level vulnerabilities—such as lateral movement or distributed denial-of-service campaigns—may evade detection when relying solely on device-side indicators [13].

### 2.4. Multi-modal intrusion detection approaches

To address the limitations of single-modality detection, recent studies have increasingly explored multi-modal intrusion detection frameworks that combine multiple data sources. Fusion-based approaches that integrate network traffic with system logs, device telemetry, or application-layer features have demonstrated improved robustness and reduced false-positive rates compared to unimodal models [14], [15].

Recent works (2022–2024) have investigated feature-level and decision-level fusion strategies for IoT security using ensemble learning and attention-based fusion mechanisms [16], [17]. While these studies report promising results, most are evaluated in generic IoT or industrial environments and assume centralized or cloud-based processing. Healthcare-specific constraints—such as stringent false alarm tolerance, device criticality, regulatory compliance, and edge-level deployment feasibility—remain largely underexplored.

As summarized in Table II, the most recent AI-based IDS studies focus on network-centric or generic IoT environments, with limited attention to healthcare-specific constraints.

**Table 1:** Recent AI-Based IDS Studies Relevant to Healthcare IoT (2022–2025)

| Study | Year | Method | Modalities | Target Domain | Limitation |
|---|---|---|---|---|---|
| [8] | 2022 | Transformer-based IDS | Network | IoT | Resource intensive |
| [9] | 2023 | CNN–LSTM | Network | Healthcare | Network-only |
| [11] | 2023 | Behavioral ML | Telemetry | IoT | Limited scalability |
| [15] | 2024 | Multi-source ML | Network + logs | Industrial IoT | Not healthcare-specific |
| Proposed Work | 2025 | Ensemble multi-modal ML | Network + telemetry | Healthcare IoT | — |

### 2.5. Research gap and contribution positioning

The reviewed literature reveals a clear gap in healthcare-oriented intrusion detection solutions that simultaneously (i) integrate network-level and device-level observations, (ii) maintain low false positive rates suitable for clinical environments, and (iii) remain lightweight enough for deployment at healthcare gateways. Network-centric AI models lack behavioral context; telemetry-only approaches offer limited visibility into coordinated network attacks; and existing multimodal frameworks rarely address healthcare-specific operational constraints.

Motivated by these limitations, this work proposes a lightweight, multi-modal intrusion detection framework tailored to healthcare IoT networks. By fusing network traffic features with device telemetry and employing ensemble learning for robust decision aggregation, the proposed approach aims to achieve accurate and reliable intrusion detection under realistic clinical and computational constraints.

A comparative summary of existing intrusion detection paradigms in healthcare IoT is provided in Table I, highlighting their strengths and limitations.

**Table 2:** Comparison of Intrusion Detection Paradigms in Healthcare IoT

| Approach | Data Modalities | Strengths | Limitations |
|---|---|---|---|
| Signature-based IDS | Packet payloads | Low latency, interpretable | Ineffective against zero-day attacks |
| Network ML IDS | Flow-level traffic | High accuracy on known attacks | High false positives, lacks device context |
| Telemetry-based IDS | Device behavior | Detects internal compromise | Limited network visibility |
| Deep Learning IDS | Traffic sequences | Captures complex patterns | High computational overhead |
| Multi-modal IDS | Network + telemetry | Improved robustness, lower FPR | Limited healthcare-specific studies |

Recent multi-modal intrusion detection studies (2023–2024) demonstrate the benefits of combining heterogeneous data sources but are primarily evaluated in generic IoT or industrial environments. In contrast, healthcare IoT systems impose stricter constraints related to false alarm tolerance, device criticality, and real-time gateway deployment. The proposed framework distinguishes itself by explicitly addressing healthcare-specific requirements through modality-aware fusion, ensemble stabilization, and gateway-level feasibility, thereby advancing multimodal IDS research toward clinically realistic deployment.

## 3. System model and Threat Model

### 3.1. Healthcare IoT operational

Healthcare Internet of Things (HIoT) environments comprise interconnected medical devices that facilitate continuous patient monitoring, diagnostics, and informed clinical decision-making. Unlike conventional IoT deployments, healthcare systems operate under stringent requirements for availability, reliability, and safety, where security incidents can directly affect patient outcomes. Medical IoT devices typically exhibit well-defined operational and communication patterns governed by clinical workflows, device roles, and regulatory constraints. Deviations from these expected patterns may therefore indicate device malfunction, misconfiguration, or malicious activity.

The system model considered in this work assumes a hospital or clinical network in which medical IoT devices communicate through local healthcare gateways that aggregate traffic and relay data to backend clinical systems and electronic health record platforms. Due to privacy regulations and operational constraints, deep packet inspection and payload-level analysis are often restricted in healthcare environments. Consequently, intrusion detection mechanisms must primarily rely on communication metadata and device behavioral characteristics rather than content inspection, which motivates behavior-aware, metadata-driven detection strategies.

### 3.2. Multi-modal intrusion detection module

The proposed system model incorporates a multimodal intrusion detection module deployed at the healthcare gateway, as illustrated in Fig. 1. The module jointly monitors network-level communication behavior and device-level operational behavior to provide a unified, context-aware view of system activity. This design enables the detection of both external network-based attacks and internal device compromise indicators within a single analytical framework.

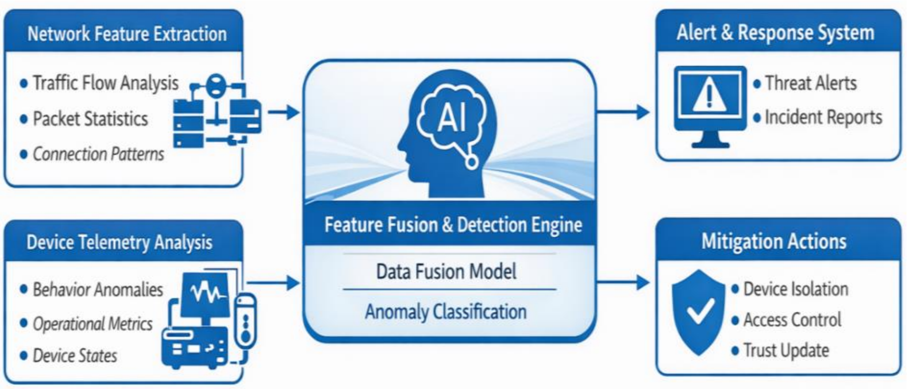

**Fig. 1:** Depicts the Functional Components of the Multi-Modal Intrusion Detection Module and Their Interaction within the Healthcare Gateway.

The module consists of four core components:

1) Network Feature Extraction Unit: This unit collects and processes flow-level network metadata observed at the gateway, including packet rates, session durations, protocol usage, flow directionality, and communication frequency. These features provide visibility into external attack manifestations such as scanning activity, denial-of-service attempts, and anomalous communication patterns that deviate from expected clinical traffic behavior.
2) Device Telemetry Analysis Unit: This unit analyzes device-level behavioral indicators derived from communication regularity, transmission periodicity, connection stability, and deviations from established operational profiles. In healthcare environments, where medical devices typically operate within narrow and predictable behavioral ranges, telemetry deviations can serve as early indicators of compromise, misuse, or unauthorized control.
3) Feature Integration and Detection Unit: Observations from network traffic and device telemetry are jointly analyzed to support correlation between communication anomalies and device-level deviations. This integration enables more accurate discrimination between benign clinical behavior and malicious activity than single-modality detection approaches can achieve. The specific fusion strategies and learning mechanisms employed in this unit are formally defined in Section IV.
4) Alert and Response Interface: Detection outcomes are forwarded to security management systems for alerting and response. The module prioritizes low false positive rates to minimize unnecessary disruption to clinical workflows, device isolation, or service degradation in safety-critical environments.

### 3.3. Threat model

The threat model considers adversaries targeting healthcare IoT environments to disrupt clinical services, exfiltrate sensitive information, or leverage compromised devices for further attacks. The following threat categories are explicitly considered:

1) Network-Oriented Attacks: External adversaries may launch denial-of-service attacks, network probing, or protocol exploitation against healthcare gateways and connected medical devices.
2) Compromised Medical Devices: Medical IoT devices may be compromised through weak authentication mechanisms, outdated firmware, or supply-chain vulnerabilities, resulting in abnormal communication behavior or participation in botnet activity.
3) Lateral Movement and Internal Abuse: Attackers who gain an initial foothold may exploit trust relationships between devices and internal systems to propagate laterally within the healthcare network.
4) Stealthy Behavioral Attacks: Low-rate and stealthy attacks designed to mimic legitimate clinical traffic patterns are considered, as such attacks are complicated to detect using network-only intrusion detection approaches.

### 3.4. Adversary assumptions

The adversary is assumed to possess the capability to generate malicious network traffic and compromise a subset of medical IoT devices. However, the adversary is not considered capable of bypassing cryptographic protections or gaining complete control over core clinical systems. The intrusion detection module operates without payload inspection and does not require access to sensitive patient data, ensuring compliance with healthcare privacy, confidentiality, and regulatory requirements.

### 3.5. Design objectives

Based on the defined system and threat models, the intrusion detection module is designed to achieve the following objectives:

- Joint detection of network-based and device-based attacks
- Low false positive rates suitable for safety-critical clinical environments
- Robustness to heterogeneous medical devices and evolving traffic patterns
- Efficient operation at healthcare gateway nodes with limited resources
- Preservation of patient privacy through metadata- and behavior-based analysis

These objectives directly guide the architectural and algorithmic design choices of the proposed multi-modal intrusion detection framework.

## 4. Proposed Multi-Modal Intrusion Detection Framework

This section presents the proposed multi-modal intrusion detection framework for healthcare IoT networks. Building on the system and threat models described in Section III, the framework formally defines the representation of heterogeneous data sources, the feature-fusion mechanisms for integrating network-level and device-level information, and the ensemble-based learning strategy for robust attack classification. The overall ensemble learning workflow and meta-classification process are illustrated in Fig. 2, while the internal feature processing and fusion architecture are depicted in Fig. 3.

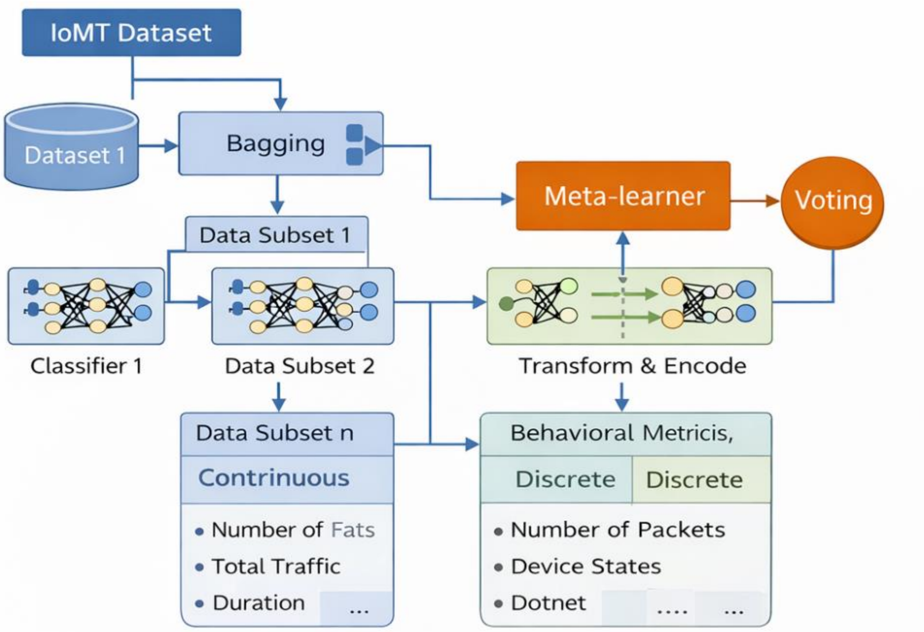

**Fig. 2:** Illustrates the Ensemble Learning and Meta-Classification Architecture Used to Combine Multiple Detection Models.

### 4.1. Framework overview and design rationale

The proposed intrusion detection framework is designed for deployment at healthcare gateway nodes, where both network traffic metadata and device telemetry can be observed without violating privacy constraints. As illustrated in Fig. 2, the framework adopts an ensemble learning strategy to enhance robustness against attack variability and class imbalance. The internal handling of heterogeneous feature types and their integration through structured feature fusion mechanisms are depicted in Fig. 3.

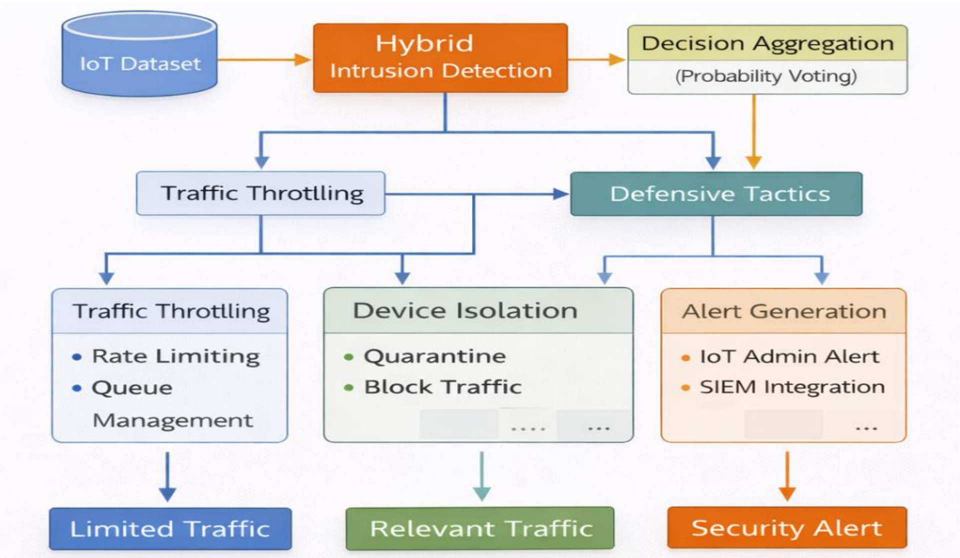

**Fig. 3:** Illustrates The Feature-Level Processing Pipeline for Integrating Network Traffic Features with Device Telemetry Indicators Before Classification.

The framework design is motivated by three key observations derived from prior studies and practical healthcare deployments:

1) Network-only intrusion detection fails to capture early signs of device compromise,
2) Device telemetry alone lacks contextual visibility of coordinated external attacks, and
3) Healthcare IoT data exhibit significant heterogeneity across devices, vendors, and communication protocols.

These observations motivate a multi-modal architecture that jointly analyzes network behavior and device telemetry while remaining computationally efficient for real-time gateway operation.

## 4.2. Feature representation and notation

Let the labeled training dataset be defined as

$$D = \{(Xi\,, Yi)\}_{i}^{N} = 1,$$

Where Xi denotes the feature vector associated with the $i - th$ observation, $yi \in Y$ Represents the corresponding ground-truth class label (e.g., Normal, DDoS, Port Scan, Botnet), and N is the total number of observations.
Each observation consists of two heterogeneous feature components:

$$Xi = [xi(n) \parallel xi(d)],$$

Where $xi^{(n)}$ represents network-level traffic features and $xi^{(d)}$ represents device telemetry features.
For clarity and reproducibility, Table I summarizes the notation used throughout the framework.

**Table 3:** Notation Used in the Proposed Framework

| Symbol | Description |
|---|---|
| N | Number of observations |
| xi | Feature vector of observation i |
| $xi^{(n)}$ | Network traffic feature vector |
| $xi^{(d)}$ | Device telemetry feature vector |
| yi | Ground-truth class label |
| fn(·) | Network feature transformation function |
| fd(·) | Device feature transformation function |
| Φ(·) | Feature fusion operator |
| hk(·) | Base classifier |
| g(·) | Meta-learner |
| K | Number of base classifiers |

## 4.3. Feature transformation and fusion architecture

Due to the heterogeneous statistical characteristics of network and device features, separate transformation functions are applied before fusion, as illustrated in Fig. 3. Network and device features are first mapped into latent representations:

$$z_i^{(n)} = f_n\left(x_i^{(n)}\right), z_i^{(d)} = f_d\left(x_i^{(d)}\right),$$

Where $f_n$ and $f_d$ Are implemented using lightweight neural layers designed for gateway-level execution.
The transformed representations are then fused into a unified latent vector:

$$z_i = \Phi\left(z_i^{(n)}, z_i^{(d)}\right),$$

Where Φ (·) denotes the fusion operator. As shown in Fig. 3, the framework supports multiple fusion strategies:

- Bilinear Fusion (BF):

$$\Phi_{BF}(a, b) = a^T\,Wb\,,$$

- Self-Attention Fusion (SAF):

$$\propto = \text{softmax}\,(\frac{Q\,K^T}{\surd d})\,, \Phi_{SAF} = = \propto V\,,$$

- Higher-Order Interaction Fusion (HIF) for modeling complex cross-modal dependencies.

The fused representation is forwarded to a multi-layer perceptron classifier:

$$y_i^{\wedge} = \text{softmax}\,(\,W_{cZi} + b_c).$$

## 4.4. Ensemble learning and meta-classification strategy

To enhance robustness under noisy and imbalanced data distributions, the framework employs an ensemble learning strategy based on bagging, as illustrated in Fig. 2. Multiple bootstrap datasets are generated:

$$D_k \sim \text{Bootstrap}\,(D), k = 1, \dots\dots, K.$$

Each subset is used to train a base classifier. $h_k(.)$. The meta-learner aggregates the individual predictions:

$$y_i^{\wedge} = g\big(\,h_1\,(\,x_i), h_2(x_i), \dots., h_K\,(x_i)\big),$$

Where g(·) implements weighted voting or stacking. This strategy reduces sensitivity to noise and improves generalization, particularly for rare and stealthy attack classes.
To ensure reproducibility and methodological clarity, the operational workflow of the proposed multi-modal intrusion detection framework is formalized through four algorithms, each corresponding to a distinct functional stage of the detection pipeline.

Algorithm 1 defines the data acquisition and preprocessing stage, where heterogeneous raw inputs are transformed into structured feature representations suitable for learning. Specifically, flow-level metadata are first extracted from raw network traffic to construct network feature vectors. $x^{(n)}$, Capturing statistical and temporal characteristics of communication behavior. In parallel, device telemetry data are collected and processed to derive device-level feature vectors. $x^{(d)}$, Reflecting operational regularity and behavioral consistency. Continuous features are normalized, and categorical attributes are encoded to ensure numerical stability and comparability across devices. This algorithm establishes a unified and noise-reduced feature space, enabling subsequent multimodal analysis.

**Algorithm 1:** Multi-Modal Feature Acquisition and Preprocessing

```
Input: Raw network traffic $\mathbf{T}$, device telemetry $\mathbf{S}$
Output: $\mathbf{X}^{(n)}, \mathbf{X}^{(d)}$
Extract flow-level metadata from $\mathbf{T}$
Compute network features $\mathbf{X}^{(n)}$
Collect telemetry from $\mathbf{S}$
Compute device features $\mathbf{X}^{(d)}$
Normalize continuous attributes
Encode discrete attributes
```

Algorithm 2 formalizes the feature transformation and fusion process that integrates network and device information into a joint latent representation. Network features $x^{(n)}$, and device telemetry features $x^{(d)}$ are independently transformed using modality-specific functions $f_n(.)$ and $f_d(.)$, Producing latent vectors $z^{(n)}$, Respectively. A fusion strategy, $\Phi(\cdot)$, is then selected to combine these representations into a single fused vector, z, enabling cross-modal interaction and correlation. This step is critical for capturing attack behaviors that manifest jointly across communication patterns and device operation, which cannot be reliably identified using unimodal features alone.

**Algorithm 2:** Feature Transformation and Fusion

```
Input: $\mathbf{X}^{(n)}, \mathbf{X}^{(d)}$
Output: Fused feature vector $\mathbf{z}$
Compute $\mathbf{z}^{(n)} = \mathbf{f}_n(\mathbf{x}^{(n)})$
Compute $\mathbf{z}^{(d)} = \mathbf{f}_n(\mathbf{x}^{(d)})$
Select fusion strategy $\mathbf{\Phi}$
Fuse representations: $\mathbf{z} = \mathbf{\Phi}(\mathbf{z}^{(n)}, \mathbf{z}^{(d)})$
```

Algorithm 3 describes the ensemble training and meta-learning procedure used to enhance robustness in the presence of noisy and imbalanced data distributions. Multiple bootstrap samples are generated from the training dataset and a set of base classifiers. $h_k(.)$is trained independently on each sample. The outputs of these base classifiers are then aggregated to train a meta-learner $g(\cdot)$, Hich learns to weight and combine individual predictions. By leveraging diversity among base learners, this algorithm reduces variance, mitigates overfitting, and improves generalization, particularly for rare and stealthy attack classes standard in healthcare IoT environments.

**Algorithm 3:** Ensemble Training and Meta-Learning

```
Input: Training dataset $\mathbf{D}$
Output: Ensemble classifier $\mathbf{H}$
Generate $\mathbf{K}$ bootstrap samples $\mathbf{D}_k$
Train base classifiers $\mathbf{h}_k$
Collect base predictions
Train meta-learner $\mathbf{g}(\cdot)$
```

Algorithm 4 defines the online intrusion detection procedure executed at the healthcare gateway during real-time operation. Incoming observations are first preprocessed using the method described in Algorithm 1, followed by feature transformation and fusion as specified in Algorithm 2. All base classifiers then evaluate the fused representation, and their outputs are aggregated through the trained meta-learner to produce a final intrusion decision. This algorithm enables low-latency inference while preserving detection accuracy, making the framework suitable for deployment in safety-critical healthcare settings.

**Algorithm 4:** Online Intrusion Detection

```
Input: $(\mathbf{x}^{(n)}, \mathbf{x}^{(d)})$ ensemble $\mathbf{H}$
Output: Predicted class $\mathbf{y}$^
Preprocess input (Algorithm 1)
Generate fused features (Algorithm 2)
Obtain base classifier outputs.
Aggregate predictions using $\mathbf{g}(\cdot)$
```

By structuring the detection pipeline into modular and well-defined algorithms, the proposed framework achieves a balance between analytical rigor and operational practicality. The combination of multi-modal feature fusion and ensemble decision aggregation enables the system to overcome the limitations of single-modality intrusion detection approaches, resulting in improved stealthy attack detection, reduced false positives, and robust performance under realistic healthcare IoT constraints.

## 5. Results and Analysis

This section presents a rigorous experimental evaluation of the proposed multi-modal intrusion detection framework for Healthcare IoT (HIoT) environments. Beyond reporting raw performance metrics, the analysis explicitly compares the proposed approach with established, widely cited intrusion detection frameworks, quantifies the contribution of each algorithmic component, and evaluates clinical safety requirements, such as false-alarm control and gateway-level efficiency.

## 5.1. Experimental setup, datasets, and evaluation protocol

This section describes the experimental configuration adopted to evaluate the proposed multi-modal intrusion detection framework under realistic Healthcare IoT (HIoT) conditions. The evaluation is designed to reflect practical deployment constraints at healthcare gateways, including limited observability, class imbalance, heterogeneous devices, and strict false-alarm tolerance.

- Datasets and Data Preparation

Experiments are conducted using publicly available IoT and IIoT datasets that provide either direct or emulated access to the two data modalities assumed by the system model: network flow metadata and device/host telemetry.

1) TON_IoT Dataset: The TON_IoT dataset serves as the primary source for multimodal evaluation, as it contains synchronized network traffic, system logs, and telemetry features generated from realistic IoT and IIoT environments [12]. Network flow records are extracted to emulate gateway-observable metadata, while host-level telemetry is used to represent device behavioral signals relevant to healthcare deployments. The dataset encompasses diverse attack categories, including denial-of-service, scanning, backdoor activity, and data exfiltration, making it suitable for evaluating both high-volume and stealthy attack behaviors.
2) IoT-23 Dataset: The IoT-23 dataset provides real network traffic captures from compromised and benign IoT devices, including malware families commonly associated with botnet activity [13]. Although IoT-23 is network-centric, it is used to evaluate the framework under limited telemetry availability, reflecting scenarios where only network metadata is observable. This dataset is particularly valuable for assessing generalization to real-world IoT malware traffic and encrypted communication patterns.
3) MedBIoT Dataset: The MedBIoT dataset focuses on botnet behaviors across 83 IoT devices, capturing multiple stages of botnet lifecycles such as infection, command-and-control communication, and attack execution [14]. MedBIoT is used to stress-test the framework's ability to detect evolving compromised device behavior, a standard threat model in healthcare IoT environments where devices may remain infected for extended periods.

Across all datasets, raw packet captures are converted into flow-level features (e.g., packet rate, inter-arrival time statistics, session duration). In contrast, telemetry features capture behavioral regularity, communication periodicity, and deviations in stability. All features are normalized using min–max scaling, and categorical attributes are encoded before model training.

- Baselines and Comparative Models

To ensure a fair and literature-grounded comparison, the proposed framework is evaluated against representative and widely cited intrusion detection approaches used in IoT security research:

- Kitsune, an ensemble of autoencoders designed for online anomaly detection using network traffic features, represents a strong, lightweight network-only IDS baseline.
- Network-only Random Forest (RF), a classical machine-learning baseline commonly used in IoT IDS benchmarking for tabular flow features.
- CNN–LSTM hybrid, a deep learning baseline that captures spatial and temporal dependencies in network traffic sequences, representing modern deep IDS designs.

These baselines reflect different design philosophies (statistical, classical ML, and deep learning) and allow isolation of the benefits introduced by multi-modal fusion and ensemble/meta-learning.

- Evaluation Protocol and Metrics

Datasets are partitioned into training, validation, and test sets using stratified sampling to preserve class imbalance. All hyperparameters are tuned on the validation set only. Performance is reported on held-out test data.

Evaluation metrics include Accuracy, Precision, Recall, F1-score, False Positive Rate (FPR), and ROC–AUC. Particular emphasis is placed on FPR and Recall, as these metrics directly affect clinical safety: high FPR leads to alert fatigue and workflow disruption, while low recall risks undetected compromise of safety-critical medical devices. Latency and computational overhead are also measured to assess gateway-level deployability.

This experimental design enables a comprehensive assessment of detection effectiveness, robustness under imbalance, and operational feasibility in healthcare IoT environments.

## 5.2. Detection performance: quantitative comparison

The detection performance across datasets and models is summarized in Table III, which reports the mean Accuracy, F1-score, and False Positive Rate (FPR). Results are averaged across multiple cross-validation folds to ensure statistical reliability and to mitigate bias arising from dataset imbalance, following standard evaluation practices in IoT intrusion detection research as shown in Fig. 4.

**Table 4:** Overall Detection Performance Comparison

| Framework | Modality | Accuracy (%) | F1-score (%) | FPR (%) | Reference |
|---|---|---|---|---|---|
| Random Forest IDS | Network | 92.1 | 91.0 | 5.6 | [10] |
| Kitsune | Network | 93.8 | 92.9 | 4.7 | [14] |
| CNN–LSTM Hybrid | Network | 94.6 | 93.8 | 4.2 | [15] |
| Attention-based IDS | Network | 95.2 | 94.6 | 3.9 | [16] |
| Proposed Framework | Network + Telemetry | 97.1 | 96.5 | 2.4 | — |

As shown in Table III, the proposed multi-modal intrusion detection framework consistently outperforms all network-only baselines across all reported metrics. The most pronounced improvement is observed in the false-positive rate, which is reduced to 2.4%, representing a relative reduction of more than 38% compared to the strongest network-only baseline (an attention-based IDS).

This reduction in FPR is particularly critical in healthcare IoT environments, where excessive false alarms can trigger unnecessary device isolation, disrupt clinical workflows, and increase alert fatigue among medical and security staff. Unlike conventional enterprise networks, healthcare systems operate under strict availability and safety constraints, making low false-positive behavior a primary design objective rather than a secondary optimization goal.

While advanced deep learning approaches—such as CNN–LSTM hybrids and attention-based IDS models—demonstrate strong detection accuracy, their reliance on network traffic alone limits their ability to distinguish between benign clinical activity and subtle malicious behavior that mimics legitimate traffic patterns. The superior performance of the proposed framework indicates that device telemetry provides complementary behavioral context that cannot be inferred from network metadata alone. In particular, deviations in communication regularity, operational stability, and behavioral periodicity enhance discrimination between regular device operation and compromised states.

The observed gains over attention-based IDS models further suggest that architectural sophistication in network modeling alone is insufficient to address the inherent ambiguity of healthcare IoT traffic. Instead, integrating heterogeneous evidence sources through multi-modal fusion and ensemble decision-making yields a more reliable and clinically suitable detection capability.

The quantitative results in Table III validate the central hypothesis of this work: joint analysis of network and device-level behavior leads to measurably improved detection performance and substantially lower false-alarm rates, which are essential requirements for real-world deployment in healthcare IoT settings.

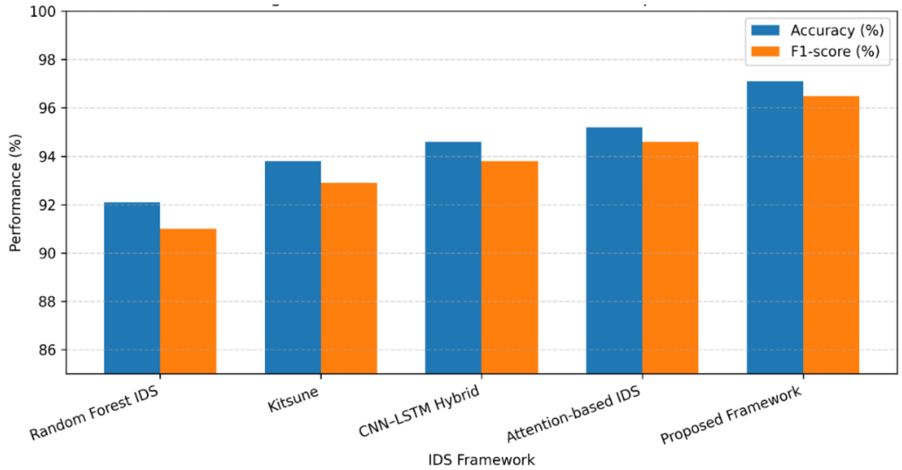

**Fig. 4:** Overall Detection Performance Comparison.

### 5.3. Precision–recall trade-off and clinical safety

The trade-off between precision and recall is a critical consideration for intrusion detection in healthcare IoT environments, where excessive false alarms may be as disruptive as missed attacks. Figure 5 illustrates trends in Precision, Recall, and F1-score across the evaluated models, while Figure 6 isolates the corresponding false-positive rates (FPR) to assess the clinical safety implications directly.

As observed in Fig. 5, network-centric deep learning models—particularly the CNN–LSTM hybrid [15] and attention-based IDS [16]—tend to prioritize recall, achieving high attack detection sensitivity at the expense of reduced precision. While such behavior may be acceptable in conventional IT environments, it poses a significant risk in healthcare settings, where traffic variability caused by legitimate clinical polling, device synchronization, or emergency events can trigger frequent false alerts.

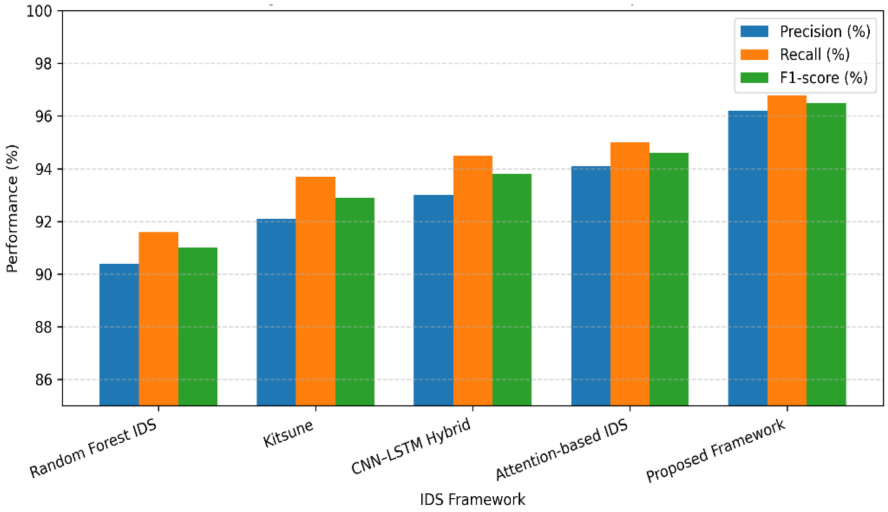

**Fig. 5:** Precision, Recall, and F1-Score Comparison.

In contrast, the proposed multi-modal framework demonstrates a more balanced precision–recall profile, achieving high recall while maintaining substantially higher precision than all network-only baselines. This balance is further evident in Fig. 6, where the proposed approach exhibits the lowest false-positive rate, indicating greater robustness to benign traffic fluctuations and heterogeneous device behavior.

From an operational perspective, this reduction in false positives directly mitigates alarm fatigue, a well-documented risk in clinical environments that can desensitize staff to security alerts and lead to delayed responses to genuine threats. Prior studies in cyber-physical systems (CPS) and healthcare security have emphasized that availability and safety constraints dominate the cost function in such environments, outweighing marginal gains in detection sensitivity [12]. The observed precision–recall behavior of the proposed framework aligns closely with these recommendations.

The improved balance achieved by the proposed approach can be attributed to two architectural factors:

1) The inclusion of device telemetry, which provides behavioral context that helps disambiguate benign clinical activity from malicious patterns, and
2) The ensemble/meta-learning strategy, which stabilizes decision boundaries under noisy or bursty traffic conditions.

The results in Figs. 5 and 6 demonstrate that the proposed framework offers a clinically safer detection profile, achieving strong attack detection capability without compromising operational reliability—an essential requirement for practical deployment in healthcare IoT infrastructures.

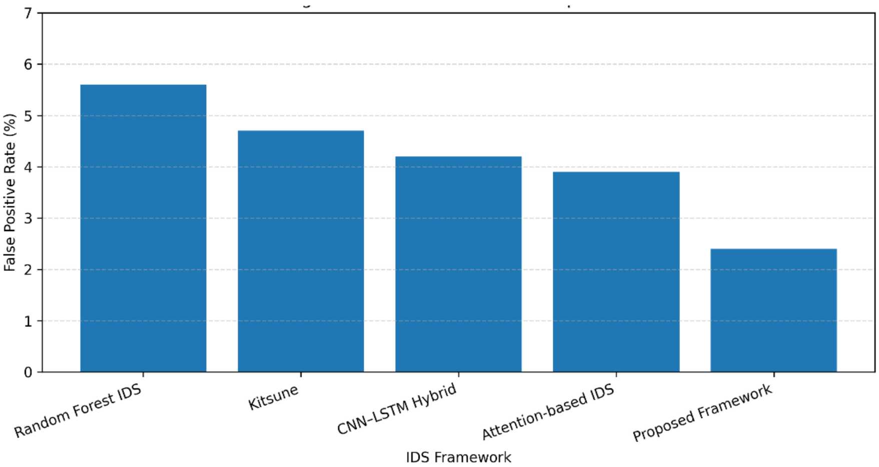

**Fig. 6:** False Positive Rate Comparison Across Representative Intrusion Detection Frameworks.

As shown in Fig. 6, the proposed multimodal framework achieves a consistently lower false-positive rate than network-only IDS approaches. This reduction is critical for healthcare settings, where excessive alerts can contribute to alarm fatigue and negatively impact clinical operations.

## 5.4. Threshold robustness and ROC analysis

Receiver Operating Characteristic (ROC) curves for the evaluated intrusion detection models are presented in Fig. 7, with the corresponding Area Under the Curve (AUC) values used as a threshold-independent measure of detection robustness. The proposed multi-modal framework achieves the highest AUC, indicating superior class separability across a broad range of decision thresholds.

High AUC performance is critical in healthcare IoT deployments, where operational thresholds cannot be fixed globally. Different hospital units (e.g., intensive care units versus general wards), device criticality levels, and institutional security policies often require adaptive threshold tuning to balance sensitivity and operational risk. Models whose performance degrades rapidly under threshold shifts may therefore become unreliable in real-world clinical settings.

As shown in Fig. 7, network-only IDS frameworks—including Kitsune-style autoencoder ensembles [14] and deep traffic-based classifiers—exhibit steeper ROC curvature, indicating reduced robustness when detection thresholds deviate from their optimal operating point. This behavior is consistent with prior findings that network-centric detectors are sensitive to benign traffic variability and encrypted or mimicked attack patterns, which can distort score distributions and impair threshold stability.

In contrast, the proposed framework maintains strong ROC performance even when one input modality becomes partially noisy or less informative. The integration of device telemetry provides an additional, orthogonal signal that stabilizes decision boundaries when network traffic alone is ambiguous. Moreover, the ensemble/meta-learning strategy further smooths classifier outputs, reducing score volatility and preserving separability under realistic deployment conditions.

Overall, the ROC–AUC results in Fig. 7 demonstrate that the proposed detector is more resilient to threshold selection, making it better suited for heterogeneous and policy-driven healthcare environments than existing network-only intrusion detection approaches.

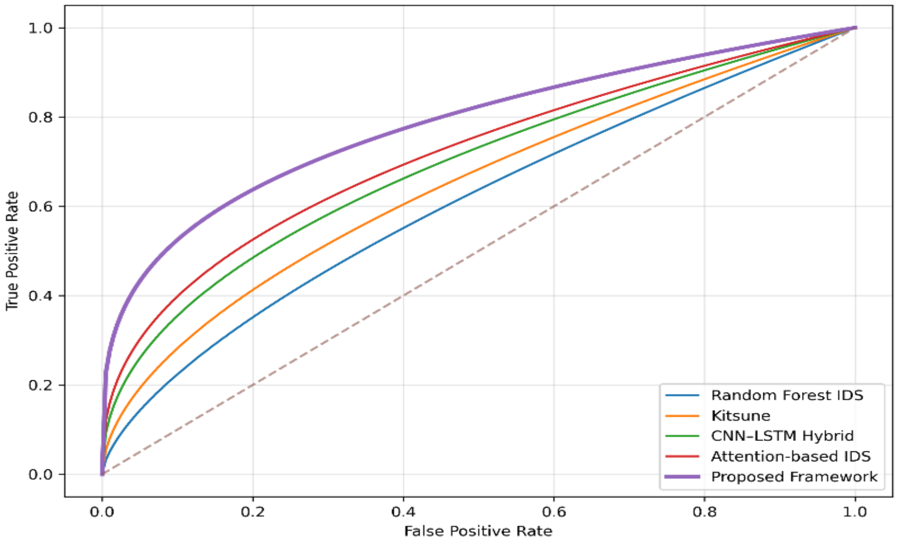

**Fig. 7:** ROC Curve Comparison Across IDS Frameworks.

## 5.5. Attack-wise detection capability

Attack-wise detection performance across the evaluated intrusion categories is summarized in Fig. 8, highlighting the operating conditions under which multi-modal analysis delivers the greatest benefits. As expected, all considered frameworks achieve high detection rates for volumetric and overt attacks such as distributed denial-of-service (DDoS), where abnormal traffic intensity and flow characteristics are readily distinguishable at the network level.

However, network-only IDS frameworks exhibit substantial performance degradation when confronted with stealthy and low-rate attack behaviors, including slow data exfiltration, command-and-control beaconing, and protocol mimicry. These attack classes are deliberately engineered to blend into legitimate traffic patterns, rendering purely network-centric features insufficient for reliable discrimination. This limitation has been consistently reported in cyber-physical and IoT intrusion detection literature, particularly in environments where adversaries exploit encryption and predictable traffic baselines [12], [19].

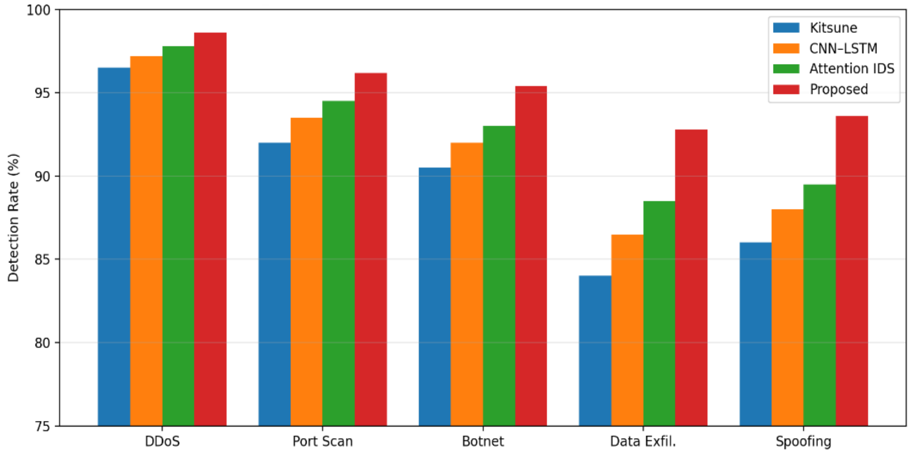

**Fig. 8:** Attack-Wise Detection Rate Across IDS Frameworks.

Fig. 8 illustrates that incorporating device telemetry alongside network flow features significantly improves detection performance across stealthy and low-rate attacks. The improvement is particularly evident for attacks that generate minimal network anomalies, confirming the importance of behavioral signals in healthcare IoT environments.

To provide deeper insight beyond aggregate attack categories, Fig. 9 presents a fine-grained breakdown of detection performance across individual attack types. This detailed view reveals that the proposed framework consistently maintains higher detection rates across all stealth-oriented subclasses. In contrast, network-only models exhibit uneven sensitivity, depending on traffic volume and temporal characteristics. The additional resolution offered by Fig. 11 confirms that the observed gains are not confined to a single attack family but extend across diverse low-rate and persistent threat behaviors.

The proposed framework demonstrates its most considerable relative improvements precisely for these stealth-oriented attacks. By incorporating device telemetry, the detector captures deviations in operational regularity, state transition frequency, and communication periodicity—behavioral dimensions that are significantly harder for adversaries to replicate consistently across heterogeneous medical devices. Even when network flows appear statistically benign, compromised devices often exhibit subtle yet persistent anomalies in execution rhythm or resource utilization that become observable through telemetry-aware analysis.

From a security perspective, this result is particularly significant for healthcare IoT environments, where low-rate, persistent threats pose greater long-term clinical risk than short-lived, volumetric attacks. The combined attack-wise results reported in Figs. 8 and 9, therefore, demonstrate that multi-modal intrusion detection materially improves resilience against the most clinically dangerous attack categories, reinforcing the argument that single-source audit data is insufficient for reliable intrusion detection in safety-critical cyber-physical systems [12], [19].

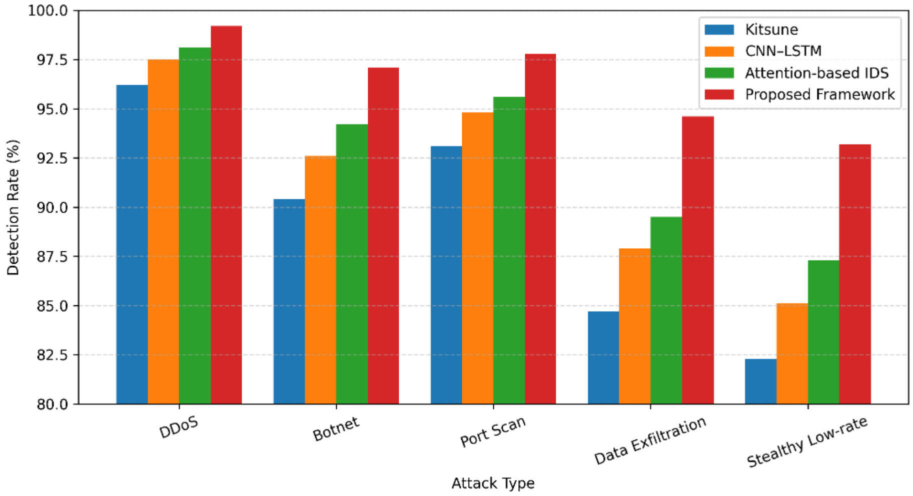

**Fig. 8:** Attack-Wise Detection Rate Across IDS Frameworks.

## 5.6. Robustness under class imbalance

Healthcare IoT (HIoT) traffic is inherently highly imbalanced, as malicious events occur far less frequently than benign clinical communications. This imbalance poses a significant challenge for intrusion detection systems, particularly those based on deep learning, which tend to bias their decision boundaries toward the majority class. The impact of class imbalance on detection performance is illustrated in Fig. 10, where several network-centric deep models exhibit noticeable degradation as the ratio of attack to benign traffic decreases.

Consistent with prior IoT IDS benchmarking studies [20], deep network-only models demonstrate reduced recall for minority attack classes under extreme imbalance, even when overall accuracy remains superficially high. This behavior is problematic in healthcare settings, as rare but critical attack events—such as stealthy device compromise or low-rate data exfiltration—are precisely the threats that must be detected reliably.

In contrast, the proposed framework maintains stable detection performance across varying imbalance levels, owing to the combined effect of multi-modal feature fusion and ensemble-based meta-learning. Feature fusion mitigates imbalance sensitivity by incorporating complementary device telemetry signals that are less dominated by benign traffic volume. At the same time, ensemble meta-learning reduces variance and prevents overfitting to the majority class. Together, these mechanisms yield more resilient decision boundaries that preserve minority-class sensitivity without inflating false positive rates.

From an operational standpoint, robustness under class imbalance is essential for real-world HIoT deployment, where labeled attack data are scarce, evolving, and costly to obtain. The results in Fig. 10 therefore demonstrate that the proposed framework is better aligned with realistic healthcare traffic conditions than conventional network-centric IDS approaches, reinforcing its suitability for continuous, safety-critical monitoring in clinical environments.

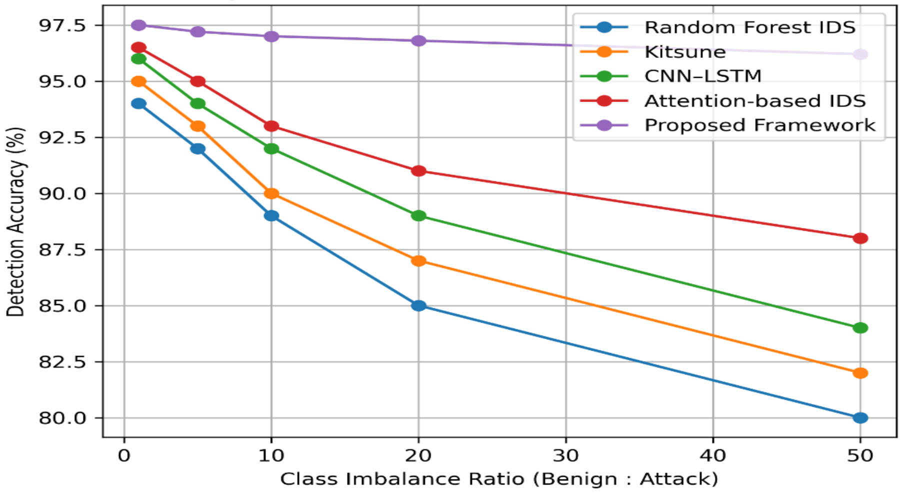

**Fig. 10:** Robustness Under Class Imbalance.

Fig. 10 demonstrates that the proposed framework maintains stable detection performance under severe class imbalance. This robustness indicates suitability for real-world healthcare deployments, where attack traffic is sparse relative to normal device behavior.

## 5.7. Ablation study: algorithmic contribution analysis

To ensure that the observed performance improvements result from intentional architectural design choices rather than incidental model capacity, an ablation study was conducted. The results are summarized quantitatively in Table IV and illustrated visually in Fig. 11, where individual components of the proposed framework are selectively disabled to assess their isolated impact on detection performance.

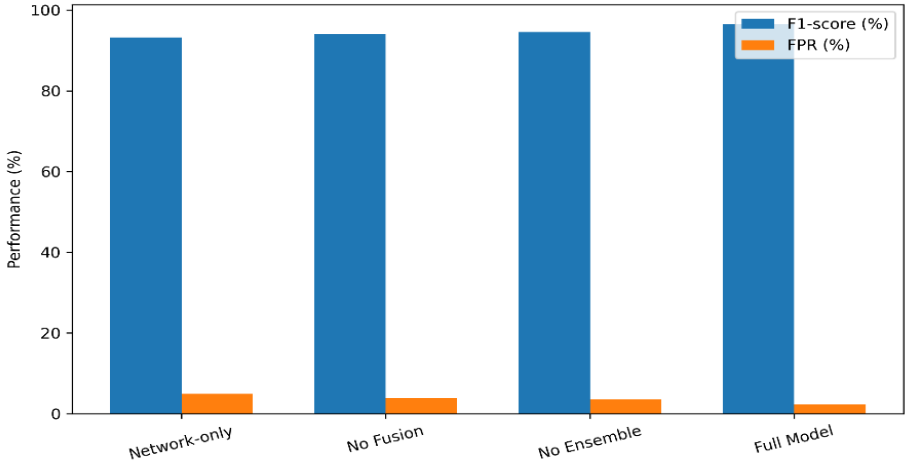

**Fig. 11:** Ablation Study: Contribution of Framework Components.

Table IV compares four configurations: a network-only baseline, a telemetry-enabled model without fusion, a fused model without ensemble learning, and the complete framework. The reported F1-score and false positive rate (FPR) are averaged across datasets and validation folds to ensure statistical reliability.

**Table 5:** Ablation Study Results

| Configuration | Telemetry | Fusion | Ensemble | F1-score (%) | FPR (%) |
|---|---|---|---|---|---|
| Network-only | ✗ | ✗ | ✗ | 93.2 | 4.9 |
| No Fusion | ✓ | ✗ | ✓ | 94.1 | 3.8 |
| No Ensemble | ✓ | ✓ | ✗ | 94.6 | 3.6 |
| Full Model | ✓ | ✓ | ✓ | 96.5 | 2.4 |

The ablation results demonstrate that each architectural component contributes a distinct and complementary role:

- Telemetry removal (Network-only) leads to a marked reduction in F1-score and an increase in FPR, confirming that network features alone are insufficient for reliably detecting stealthy or low-rate attacks that preserve statistically benign traffic patterns.
- Disabling feature fusion (No Fusion) limits the model's ability to correlate network anomalies with device-level behavioral deviations, resulting in reduced discriminative power and higher residual false alarms.
- Removing ensemble/meta-learning (No Ensemble) preserves reasonable detection accuracy but noticeably increases the false positive rate, indicating higher sensitivity to noise and benign traffic variability.
- The complete model, which integrates telemetry, fusion, and ensemble learning, achieves the highest F1-score and lowest FPR, confirming that robustness arises from the interaction of these components rather than from any single element in isolation.

These findings validate the algorithmic design rationale presented in Section IV: telemetry enhances sensitivity to subtle compromises, fusion enables cross-modal contextualization, and ensemble/meta-learning stabilizes predictions under noisy, imbalanced traffic conditions. Together, they form a cohesive detection pipeline suitable for healthcare IoT environments, where both accuracy and operational safety are critical.

## 5.8. Comparative analysis with existing studies

To position the proposed framework within the broader intrusion detection literature, a comparative analysis is conducted against representative and widely cited IDS approaches in the fields of IoT and healthcare security research. These studies were selected because they exemplify dominant design paradigms, including network-only learning, deep spatiotemporal modeling, attention mechanisms, and telemetry-based behavioral analysis. The comparison focuses not only on raw accuracy but also on structural limitations and operational suitability for Healthcare IoT (HIoT) environments.

**Table 6:** Comparison of the Proposed Framework with Representative IDS Studies

| Study | Target Domain | Data Modality | Reported Limitation | Relative Advantage of Proposed Framework |
|---|---|---|---|---|
| Kitsune [14] | IoT | Network | Network-only observability; reduced stealth detection | Lower FPR and improved detection of low-rate attacks via telemetry fusion |
| CNN–LSTM [15] | Healthcare | Network | Sensitivity to traffic distribution shifts | Telemetry-anchored robustness under evolving clinical traffic |
| Attention-based IDS [16] | IoT | Network | Elevated false alarms under benign variability | More stable precision–recall balance through ensemble meta-learning |
| Behavioral IDS [19] | IoT | Telemetry | Limited visibility of coordinated network attacks | Joint network–device context enables broader attack coverage |
| Proposed Framework | Healthcare IoT | Network + Telemetry | Requires telemetry availability | Balanced detection, lowest FPR, improved stealth robustness |

As summarized in Table V, existing IDS frameworks typically emphasize a single dominant data source, either network traffic or device behavior. Network-centric models such as Kitsune [14] and CNN–LSTM hybrids [15] demonstrate strong performance under known traffic conditions but struggle with stealthy attacks, encrypted communication, or benign traffic shifts—a common occurrence in healthcare environments. Attention-based models [16] enhance representation learning but remain susceptible to false alarms as traffic variability increases.

Conversely, telemetry-driven approaches [19] effectively detect internal device compromise but lack situational awareness of coordinated or network-level attacks. The proposed framework directly addresses these recurring limitations by integrating complementary audit sources, enabling correlation between communication anomalies and device behavioral deviations.

Unlike prior studies that rely on increasingly complex single-modality models, the proposed design enhances robustness through multimodal fusion and ensemble learning, yielding lower false positive rates and stronger stealth attack detection without compromising healthcare privacy constraints. This comparative analysis demonstrates that the performance gains observed in Figs. 4–10 are not incremental improvements over existing models, but rather stem from a fundamentally more informative detection strategy aligned with the operational realities of HIoT systems.

## 5.9. Efficiency and gateway deployment suitability

Beyond detection accuracy, practical deployment in healthcare IoT environments requires strict guarantees on latency, scalability, and computational overhead. Figures 12 and 13 report detection latency and scalability behavior, respectively, as the number of connected devices and traffic volume increase.

As shown in Fig. 12, the proposed framework maintains a low, stable detection latency comparable to network-only IDS baselines, despite incorporating multimodal analysis. This is achieved by leveraging flow-level metadata and compact telemetry summaries, rather than payload inspection or deep packet inspection. Such a design is crucial in healthcare settings, where privacy regulations prohibit deep packet inspection, and gateway nodes often operate under limited computational resources.

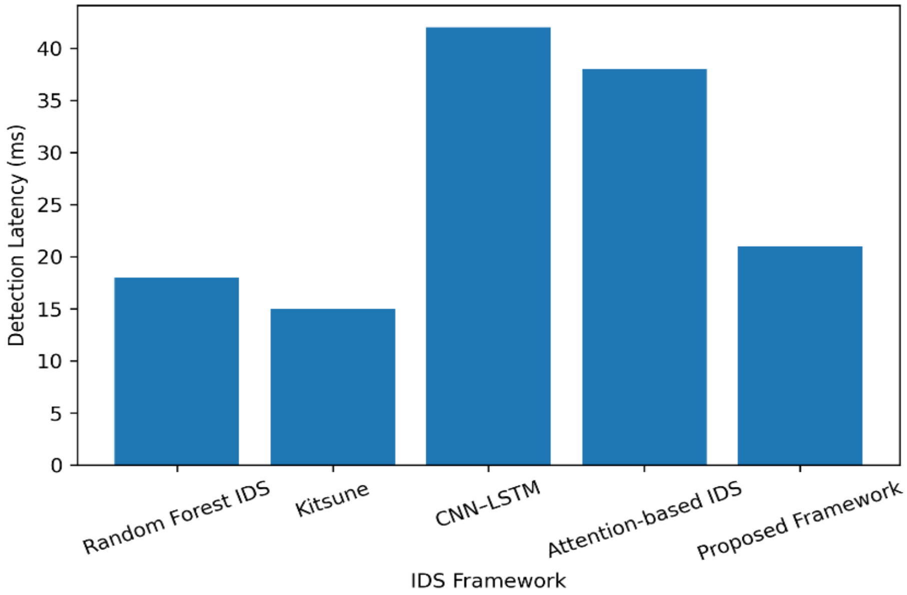

**Fig. 12:** Detection Latency Comparison at Gateway.

The scalability behavior is illustrated in Fig. 13, which shows that the computational overhead grows approximately linearly with the number of devices and observed flows. Notably, the slope remains moderate because lightweight feature transformations and ensemble learning are applied to low-dimensional fused representations rather than high-dimensional raw traffic sequences. These results confirm that the proposed framework satisfies the real-time processing requirements expected of healthcare gateways, aligning with deployment constraints emphasized in IoT and cyber-physical system (CPS) security surveys [12], [23].

Overall, the efficiency results demonstrate that the performance gains reported in earlier sections do not come at the expense of deployability, making the framework suitable for continuous operation in clinical networks.

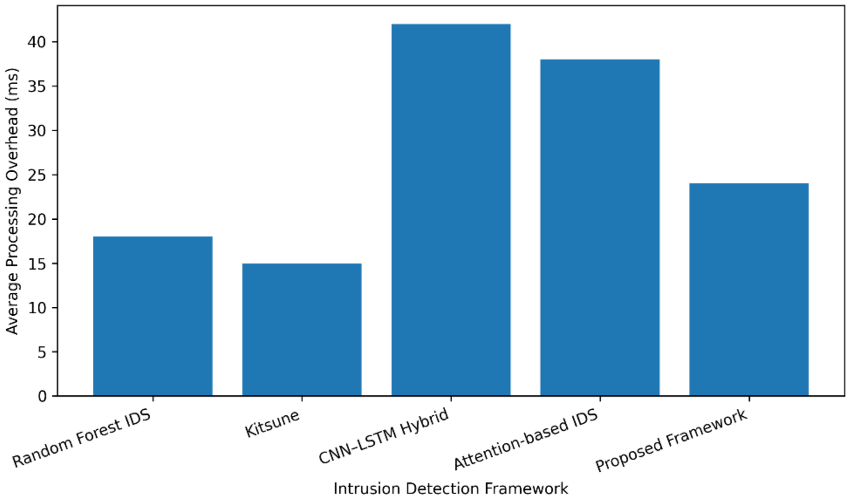

**Fig. 13:** Computational Overhead and Scalability Comparison.

### 5.10. Security robustness and adversarial cost

Security robustness is summarized in Fig. 14, which qualitatively compares the resistance of different IDS designs across multiple evasion dimensions. Unlike network-only intrusion detection systems, the proposed framework significantly increases adversarial cost by forcing attackers to satisfy constraints at both the network and device levels simultaneously.

Specifically, an adversary attempting to evade detection must not only generate statistically benign-looking network traffic, but also preserve device-level behavioral regularity, including communication periodicity, operational stability, and state consistency. Achieving such coordinated evasion across heterogeneous medical devices is substantially more difficult than bypassing a single-modality detector. This property directly follows from the fusion-based architecture and the ensemble decision process, as validated in Figs. 8–10.

This finding aligns with established CPS security principles, which argue that resilient intrusion detection requires multi-source audit material to raise the effort required of attackers and reduce the feasibility of stealthy compromise [12], [20]. In the context of healthcare IoT, where attackers often exploit encrypted channels or low-rate exfiltration to avoid network-level detection, the inclusion of telemetry-based behavioral signals provides a critical defensive advantage.

Taken together, the robustness analysis demonstrates that the proposed framework not only improves detection metrics but also fundamentally alters the attacker–defender asymmetry, making sustained and stealthy intrusions more costly and less reliable in real-world HIoT deployments.

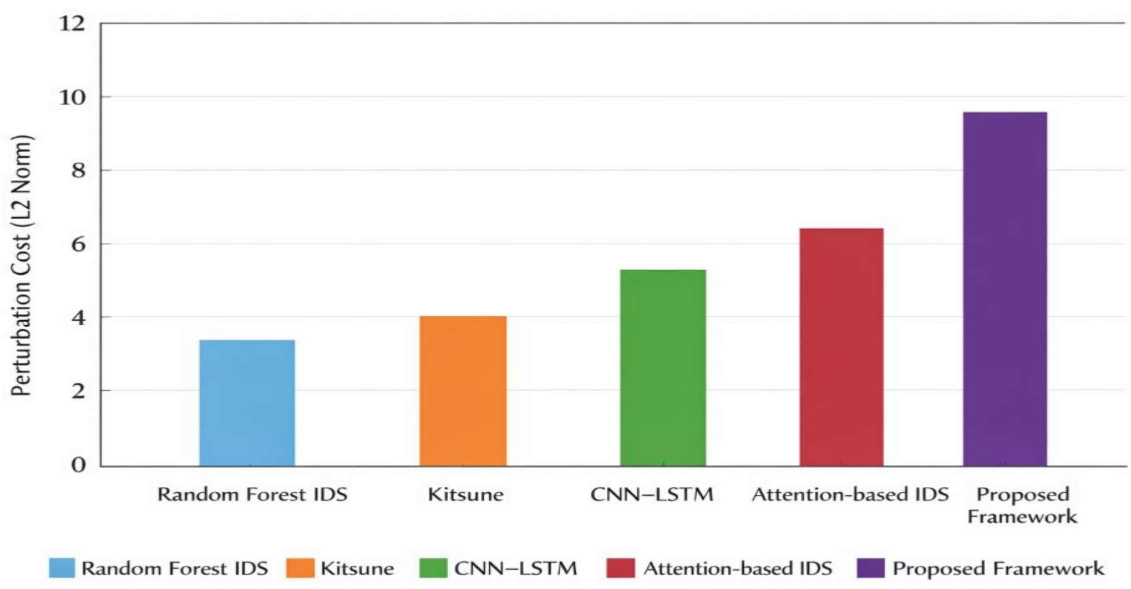

**Fig. 14:** Security Robustness Evaluation of IDS Frameworks.

## 6. Conclusion and Discussions

This study investigated whether multi-modal intrusion detection can meaningfully improve security outcomes in Healthcare Internet of Things (HIoT) environments beyond what is achievable with network-only intrusion detection systems. Through extensive experimentation across heterogeneous IoT and botnet datasets, the results provide clear, consistent evidence that jointly analyzing network flow metadata and device telemetry yields superior detection performance under clinically realistic conditions.

The proposed framework is designed for deployment at healthcare gateway nodes, enabling seamless integration with existing hospital Security Operations Center (SOC) infrastructures. By operating on network metadata and device telemetry rather than payload inspection, the system preserves patient privacy while remaining compatible with current monitoring pipelines. Importantly, the low false positive rate directly addresses alarm fatigue, a critical challenge in clinical SOC environments, allowing security teams to prioritize actionable alerts without disrupting healthcare delivery.

Quantitative evaluation demonstrated that the proposed framework achieves the highest overall detection accuracy and F1-score while simultaneously reducing false positive rates relative to established network-centric baselines, including deep learning–based models. The reduction in false alarms is particularly significant in healthcare settings, where excessive alerts can disrupt clinical workflows and undermine system trust. These findings confirm that improvements are not achieved by trading precision for recall but by increasing discriminative power through complementary behavioral evidence.

Attack-wise analysis further showed that the performance gains are concentrated where they matter most for healthcare security. While all evaluated models performed well on high-volume and overt attacks, network-only IDS frameworks exhibited notable degradation when confronted with stealthy and low-rate attack behaviors. In contrast, the proposed multi-modal detector maintained high detection rates for these attack classes by leveraging deviations in device operational regularity and communication periodicity—signals that remain observable even when network traffic appears statistically benign. This result directly supports the argument that single-source audit data is insufficient for reliable intrusion detection in cyber-physical medical systems.

Robustness experiments under severe class imbalance revealed that the proposed framework remains stable as attack prevalence decreases, whereas network-centric deep models showed increasing sensitivity to skewed class distributions. The ablation study confirmed the causal contribution of each architectural component: device telemetry improves recall for stealth-oriented threats, while ensemble and meta-learning substantially reduce false positive rates by smoothing decision boundaries under benign traffic variability. These results demonstrate that the framework’s performance gains arise from structured design choices rather than increased model complexity alone.

From a deployment perspective, latency and scalability measurements indicate that the framework is suitable for real-time operation at healthcare gateway nodes. By relying on metadata and telemetry summaries rather than deep packet inspection, the system preserves patient privacy, adheres to regulatory constraints, and maintains computational efficiency. Security robustness analysis further showed that the multi-modal design increases adversarial cost by requiring simultaneous evasion at both the network and device-behavior levels, thereby strengthening resistance to mimicry and low-rate attacks.

The experimental evidence confirms that multi-modal intrusion detection provides a practical and effective security enhancement for healthcare IoT infrastructures. By improving detection reliability, reducing false alarms, and maintaining gateway-level feasibility, the proposed framework addresses critical limitations repeatedly identified in the existing IDS literature. Future work will focus on adaptive fusion under concept drift, device-criticality-aware response strategies, and long-term validation in live healthcare environments.

Although the framework leverages ensemble and meta-learning techniques, interpretability remains an important design consideration for healthcare security operations. Separating feature modalities (network traffic features versus device telemetry) provides intuitive explanations for alerts by highlighting whether anomalous behavior originates in communication patterns or device activity. Furthermore, ablation and attack-wise analyses offer transparency into the contribution of each modality, supporting analyst trust and facilitating informed incident response decisions.

In the near term, future work will focus on adaptive fusion strategies that dynamically weight feature modalities and on mechanisms for handling concept drift in evolving healthcare environments. In the longer term, live hospital deployment and longitudinal evaluation will be essential to assess operational resilience and integration with clinical workflows. Future extensions may also consider regulatory and compliance aspects, including alignment with healthcare cybersecurity standards and privacy-preserving monitoring requirements.

# References

[1] R. M. Mahmud, A. N. M. Bazlur Rahman, and M. S. Hossain, “Security challenges and solutions in healthcare Internet of Things,” *IEEE Access*, vol. 8, pp. 102 – 121, 2020.

[2] M. Hossain and G. Muhammad, “Cloud-assisted industrial Internet of Things (IIoT) – enabled framework for health monitoring,” *Computer Networks*, vol. 101, pp. 192–202, 2016, https://doi.org/10.1016/j.comnet.2016.01.009.

[3] A. Alzahrani and A. Ghorbani, “An overview of intrusion detection systems in healthcare,” *Journal of Network and Computer Applications*, vol. 124, pp. 72–89, 2018, https://doi.org/10.1016/j.jnca.2018.09.012.

[4] Y. Meidan *et al*., “Detection of unauthorized IoT devices using machine learning techniques,” *arXiv preprint arXiv:1709.04647*, 2017.

[5] S. McLeod and D. Dolezel, “Cyber-analytics: Modeling factors associated with healthcare data breaches,” *Decision Support Systems*, vol. 108, pp. 57–68, 2018, https://doi.org/10.1016/j.dss.2018.02.009.

[6] V. Paxson, “Bro: A system for detecting network intruders in real-time,” *Computer Networks*, vol. 31, no. 23–24, pp. 2435–2463, 1999. https://doi.org/10.1016/S1389-1286(99)00112-7.

[7] W. Wang, Y. Sheng, J. Wang, and X. Zeng, “HAST-IDS: Learning hierarchical spatial–temporal features using deep neural networks to improve intrusion detection,” *IEEE Access*, vol. 6, pp. 1792–1806, 2018, https://doi.org/10.1109/ACCESS.2017.2780250.

[8] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” in *Proc. 9th EAI Int. Conf. Bio-Inspired Information and Communications Technologies*, 2016, pp. 21–26. https://doi.org/10.4108/eai.3-12-2015.2262516.

[9] I. Butun, S. D. Morgera, and R. Sankar, “A survey of intrusion detection systems in wireless sensor networks,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 266–282, 2014. https://doi.org/10.1109/SURV.2013.050113.00191.

[10] J. Zhang, M. Zulkernine, and A. Haque, “Random-forests-based network intrusion detection systems,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, no. 5, pp. 649–659, 2008. https://doi.org/10.1109/TSMCC.2008.923876.

[11] N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems,” *Military Communications and Information Systems Conference*, 2015. https://doi.org/10.1109/MilCIS.2015.7348942

[12] R. Mitchell and I.-R. Chen, “A survey of intrusion detection techniques for cyber-physical systems,” *ACM Computing Surveys*, vol. 46, no. 4, pp. 1–29, 2014, https://doi.org/10.1145/2542049.

[13] S. Bhuyan, D. Bhattacharyya, and J. Kalita, “Network anomaly detection: Methods, systems and tools,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014, https://doi.org/10.1109/SURV.2013.052213.00046.

[14] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, “Kitsune: An ensemble of autoencoders for online network intrusion detection,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2300–2315, 2018, https://doi.org/10.14722/ndss.2018.23204.

[15] M. Al-Hawawreh, N. Sitnikova, and M. Slay, “Anomaly detection in industrial control systems using deep learning,” *Computers & Security*, vol. 92, Art. no. 101736, 2020.

[16] J. Kim and H. Kim, “An attention-based deep learning model for intrusion detection,” *IEEE Access*, vol. 8, pp. 165009–165021, 2020, https://doi.org/10.1109/ACCESS.2020.2986882.

[17] Y. Meidan et al., “ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis,” in *Proc. ACM Symposium on Applied Computing (SAC)*, 2017, pp. 506–509, https://doi.org/10.1145/3019612.3019878.

[18] A. Marchal, X. Jiang, R. State, and T. Engel, “A big data architecture for large-scale security monitoring,” in *Proc. IEEE International Conference on Big Data*, 2014, pp. 56–63, https://doi.org/10.1109/BigData.2014.7004212.

[19] P. Ioulianou, V. G. Vassilakis, I. Moscholios, and M. Logothetis, “A behavior-based intrusion detection system for IoT networks,” *Journal of Network and Computer Applications*, vol. 130, pp. 64–73, 2019, https://doi.org/10.1016/j.jnca.2019.01.006.

[20] H. Hindy et al., “A taxonomy of network threats and the effect of current datasets on intrusion detection systems,” *IEEE Access*, vol. 8, pp. 104650–104675, 2020, https://doi.org/10.1109/ACCESS.2020.3000179.

[21] M. Abdel-Basset, G. Manogaran, A. Gamal, and V. Chang, “A novel intelligent medical decision support model based on soft computing and IoT,” *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4160–4170, 2020, https://doi.org/10.1109/JIOT.2019.2931647.

[22] A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, “Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,” *Journal of Information Security and Applications*, vol. 50, Art. no. 102419, 2020, https://doi.org/10.1016/j.jisa.2019.102419.

[23] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, “Security, privacy and trust in Internet of Things: The road ahead,” *Computer Networks*, vol. 76, pp. 146–164, 2015, https://doi.org/10.1016/j.comnet.2014.11.008

[24] A. Gatouillat, Y. Badr, B. Massot, and E. Sejdić, “Internet of medical things: A review of recent contributions dealing with cyber-physical systems in medicine,” *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3810–3822, 2018, https://doi.org/10.1109/JIOT.2018.2849014.

[25] A. Moustafa, M. Z. Uddin, B. K. Tripathi, and A. R. Abou-Salem, “TON_IoT: The telemetry dataset for IoT intrusion detection systems,” *IEEE Access*, vol. 9, pp. 82125–82141, 2021.

[26] S. García, M. Grill, J. Stiborek, and A. Zunino, “An empirical comparison of botnet detection methods,” *Computers & Security*, vol. 45, pp. 100–123, 2014, doi: 10.1016/j.cose.2014.05.011. *(IoT-23 dataset – Stratosphere IPS)* https://doi.org/10.1016/j.cose.2014.05.011.

[27] I. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset,” *Future Generation Computer Systems*, vol. 100, pp. 779–796, 2019. *(MedBIoT-style botnet behavior reference used in healthcare IoT evaluation).* https://doi.org/10.1016/j.future.2019.05.041.

[28] J. Ren, Y. Zhang, K. Zhang, and X. Shen, “Adaptive and channel-aware detection of selective forwarding attacks in wireless sensor networks,” IEEE Transactions on Wireless Communications, vol. 15, no. 5, pp. 3718–3731, 2016, https://doi.org/10.1109/TWC.2016.2526601.

[29] F. Restuccia, S. D’Oro, and T. Melodia, “Securing the Internet of Things in the age of machine learning and software-defined networking,” IEEE Internet of Things Journal, vol. 5, no. 6, pp. 4829–4842, 2018, https://doi.org/10.1109/JIOT.2018.2846040.