Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.085378

ARTICLE

ExGAME: An Explainable Game Theoretic and Adaptive Intrusion Detection Framework for Human-Centric Medical IoT

Noha Alnazzawi[1], Nazik Alturki[2,*], Umar Mujahid[3], Fahad Masood[4] and Jawad Ahmad[5]

[1]Computer Science and Engineering Department, Yanbu Industrial College, Royal Commission for Jubail and Yanbu, Yanbu, Saudi Arabia

[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

[3]School of Science and Technology, Georgia Gwinnett College, Lawrenceville, GA, USA

[4]Department of Computer Science, CECOS University of IT and Emerging Sciences, Peshawar, Pakistan

[5]Cybersecurity Center, Prince Mohammad Bin Fahd University, Alkhobar, Saudi Arabia

*Corresponding Author: Nazik Alturki. Email: namalturki@pnu.edu.sa

Received: 10 May 2026; Accepted: 29 June 2026; Published: 28 August 2026

ABSTRACT: The rapid deployment of Internet of Medical Things (IoMT) devices in current healthcare systems has made it much easier to maintain patient monitoring, make diagnoses, and provide long-distance medical treatment. The interconnection of these devices also creates significant cybersecurity problems, including distributed denial-of-service attacks, data breaches, and network intrusions. High detection accuracy and interpretability are essential for a trustworthy intrusion detection system. This study presents ExGAME, an Explainable Game-Theoretic Artificial Intelligence framework intended for intrusion detection in human-centric IoT networks. The proposed framework combines machine-learning-based anomaly detection with explainable AI and a game-theoretic defense strategy to improve both detection performance and decision-making clarity. A game-theoretic perspective is used to explain the interaction of the attackers and defenders. Experiments have been performed using IoT-23, Bot-IoT, CICIDS2017, UNSW-NB15, WUSTL-EHMS-2020, and MedBIoT datasets. Results achieved an accuracy range of upto 98% in different attack scenarios. The results show that packet rate and traffic flow characteristics are very important for distinguishing between normal and abnormal network activities. The proposed ExGAME architecture makes detection more transparent while enabling effective intrusion localization. This research aids in the creation of secure, comprehensible, and flexible protection mechanisms for future healthcare IoT systems.

KEYWORDS: Cyber security; explainable AI; game theory; IoMT; machine learning

1 Introduction

The Internet of Things (IoT) has grown rapidly and given rise to several critical issues, including energy efficiency, cyber security, and resource optimization [1]. IoMT is a network of medical devices, sensors, and healthcare apps providing real-time monitoring, remote diagnostics, and smart medical services [2]. These technologies have made healthcare more efficient by enabling continuous monitoring and automated data transmission between medical devices and healthcare infrastructure, improving patient outcomes and making medical services more accessible [3]. Wearable devices, smart sensors, and connected medical equipment transmit large volumes of sensitive data over hospital networks and cloud-based platforms. These

improvements make healthcare delivery much better, but they also make cybersecurity harder because there are so many different types of devices connected to one another and operating across different networks [4].

IoMT systems offer several benefits, yet their widespread use has made healthcare systems more vulnerable to cyber threats to sensitive patient information [5]. Most IoMT devices lack sufficient processing power and rely on outdated security measures, making them easy targets for hackers. Numerous studies have examined machine learning deep learning, and federated learning methodologies for intrusion detection in IoT and IoMT networks [6,7]. Support Vector Machines, Random Forests, Artificial Neural Networks, and Long Short-Term Memory networks are among the most common methods for detecting unusual traffic patterns. When tested on publicly available datasets such as IoT-23 [8], Bot-IoT [9], CICIDS2017 [10], UNSW-NB15 [11], WUSTL-EHMS-2020 [12], and MedBIoT [13], these models have demonstrated strong detection performance. Most current methods, on the other hand, focus primarily on improving detection accuracy. It is essential not only to identify cyber threats but also to understand the rationale behind specific decisions [14,15].

A limitation of existing intrusion detection systems is their failure to account for attackers' strategic behavior. Cyber attackers are continually devising new methods to bypass detection systems, keeping attackers and defenders in a constant state of conflict. These systems often treat intrusion detection as a static categorization problem, overlooking how malicious agents and security measures interact in real time. These limitations underscore the need for adaptive, intelligent protection systems that can respond to evolving threats while maintaining robust detection performance [16].

This study introduces ExGAME, an Explainable Game-Theoretic Artificial Intelligence framework for detecting intrusions in the IoMT environment. The proposed framework integrates machine-learning detection, explainable artificial intelligence techniques, and a game-theoretic defense strategy to enhance detection precision and decision-making clarity. The model analyzes network traffic metrics such as packet rate, flow duration, payload size, source bytes, and TCP flags, and analyzes traffic behavior to identify normal or malicious network patterns. A game-theoretic framework has also been shown to model interactions between attackers and defenders, allowing the system to adjust its defenses as cyber threats evolve. The explainability module employs Shapley Additive explanations (SHAP) based feature contribution analysis to explain how the model makes decisions in a comprehensible way. These sections work together to make the intrusion detection process clearer and more effective.

This unique method uses explainable artificial intelligence and a game-theoretic defense framework to find IoMT incursions. The ExGAME approach suggested puts more emphasis on adaptability, transparency, and strategic defense. By merging interpretable machine learning methods with adversarial modeling, the framework provides a more complete security solution. It can detect cyber threats and explain why it made each choice. This integrated approach helps create smart and reliable cybersecurity systems for the next generation of healthcare IoT infrastructure.

Section 2 of the presented work highlights the details of the literature review. Section 3 presents complete methodology details. Results and discussion have been presented in Section 4, and conclusion and future work are discussed in Section 5.

2 Literature Review

A new framework has been proposed that uses a human-centric approach and a quantum random forest (QRF) to detect cyberattacks while protecting patient data [17]. The framework is evaluated using performance metrics such as accuracy, detection rate, time, and memory complexity. The experimental

results show that the proposed framework excelled in attack detection using the ICU and WUSTL-EHMS-2020 datasets. An ensemble-based ML-envisioned scheme has been proposed to detect various types of intrusions in the Industry 5.0-driven healthcare system (in short, EIDS-HS) [18]. The proposed EIDS-HS has been tested on a standard data set and its performance has been evaluated using key performance parameters. The security analysis of the proposed EIDS-HS demonstrates its resistance to various potential attacks. Furthermore, EIDS-HS outperforms existing intrusion detection schemes across key performance metrics.

The development of SENTRY-AI, an explainable, multimodal anomaly-detection framework that integrates deep learning and computer vision to improve cybersecurity defenses in cyberspace, has been discussed in [19]. A Variational Autoencoder (VAE) has been applied to conduct unsupervised anomaly detection on tabular network features with a Convolutional Neural Network (CNN). An analysis is performed on time-series traffic data, transformed into a Gramian Angular Field (GAF), using a late-fusion approach. Outputs from the two models are used to evaluate the efficacy and robustness of the hybrid approach. Three intrusion detection datasets (NSL-KDD, CICIDS2017, and UNSW-NB15) have been used in the experimental work. SENTRY-AI outperformed traditional machine learning intrusion detection system (IDS) approaches, achieving an F1-score of 98.92% on NSL-KDD, 100.00% on UNSW-NB15, and an AUC-ROC score exceeding 99% on all datasets.

The Fused Federated Learning (FFL) framework has been introduced and integrated with IoMT devices to securely monitor patient health data in a decentralized manner [20]. Real-Time Sequential Deep Extreme Learning Machine (RTS-DELM) has been combined with the Fused Federated Learning (FFL), securing chronic kidney disease diagnosis within Healthcare 5.0. Experimental validation demonstrates noteworthy progress, achieving an accuracy of 98.21%, thereby showcasing superior performance over current federated learning methodologies. Split Federated Learning (SFL) using a hybrid of Deep Learning (DL) and Blockchain has been proposed, with the split occurring between the IoMT and Edge layers [21]. Edge-based bidirectional Long Short-Term Memory (BiLSTM) networks identify threats and temporal patterns, while a lightweight CNN at the IoT layer extracts spatial features from patient data. SFL facilitates safe decentralized training, and Blockchain uses the Practical Byzantine Fault Tolerance (PBFT) for trust and authentication. Results show improved privacy, robustness against attacks, block commit rate (450 b/s), and reduced consensus time (250 ms) in addition to superior accuracy (99.95%).

A dynamic Trust-Aware Controller (TAC) is presented on benchmark datasets and proprietary H-IoT signals under various attack and noise circumstances [22]. Real-time trust scores have been computed using anomaly likelihood, context entropy, and historical behavior. The suggested framework attained an average F1-score of 94.3% for anomaly detection and a 96.1% accuracy in access decision classification. Comparative results against rule-based and statistical baselines showed a 12%–18% improvement in detection sensitivity. AIDA (Awareness, Integration, Detection, and Adaptation) has been proposed as a clinically aware security analytics framework for healthcare information environments [23]. AIDA combines Random Forest, Autoencoder, and XGBoost models, along with multi-objective optimization and a Patient Care Disruption Index (PCDI), to support clinically informed response selection. AIDA achieved an AUC of 0.961, an Area Under the Precision–Recall Curve (AUCPR) of approximately 0.955, and an average detection latency of 18 s. In the simulated hospital environment, the framework reduced PCDI by approximately 37%.

Pseudo-random sequences have been proposed to ensure the confidentiality of medical images by using a novel 3D hyperchaotic map [24]. The approximation components of the images are extracted, followed by a novel diffusion algorithm that masks critical information. Performance analysis reveals that the decrypted medical images exhibit high visual dependability, consistently achieving values above 35 dB. A secure magnetic resonance imaging (MRI) image classification system has been proposed using chaotic and Arnold encryption techniques [25]. High classification performance has been achieved for chaotic and

Arnold encryption, with accuracies of 93.75% and 94.1%, precisions of 94.38% and 96.9%, recalls of 93.75% and 94.1%, and F-scores of 93.67% and 96.6%, respectively. A machine learning-enabled Cognitive Cyber-Physical System (ML-CCPS) has been designed for cyber threats identifications in MIoT environments [26]. Extreme Learning Machine (ELM)-based classification has been used for adaptive access control and reliable intrusion detection. ML-CCPS outperformed standard classifiers by F1-score of 97.8% and an AUC of 99.1%.

A hybrid deep learning-based intrusion detection system has been proposed for IoMT networks (HIDS-IoMT) using Convolutional Neural Network (CNN) and the Long Short Term Memory neural network (LSTM) [27]. The IoTID20 and the Edge-IIoTset datasets have been used for analysis and assessment. The model attained an accuracy of 99.92%, a precision of 99.91%, a recall rate of 99.99%, and an F1-score of 99.95%. The importance of domain-specific datasets has been analyzed and compared for various ML models [28]. The dataset's limitations have been analyzed to underscore the importance of domain-specific data.

3 Methodology

A human-centric and adaptive intrusion detection system for Healthcare IoT environments has been proposed, integrating explainable artificial intelligence with dynamic game-theoretic defense modeling shown in Fig. 1. Relevant communication and behavioral features are extracted, generating network traffic by medical IoT devices. Both known and anomalous attack patterns are identified by a machine-learning-based detection engine. An explainability module is encompassed, ensuring a transparent understanding for the healthcare administrators. A dynamic game-theoretic-based attacker-defender interaction is incorporated to further enhance resilience against adaptive adversaries. The attacker's behavior is continuously monitored to adjust the defender's strategies, rather than relying on static detection thresholds. It balances detection sensitivity, false alarm rates, and computational efficiency. The human decision-making requirements are fully aligned with the explainability and strategic adaptation components, ensuring interpretability of the defensive adjustments.

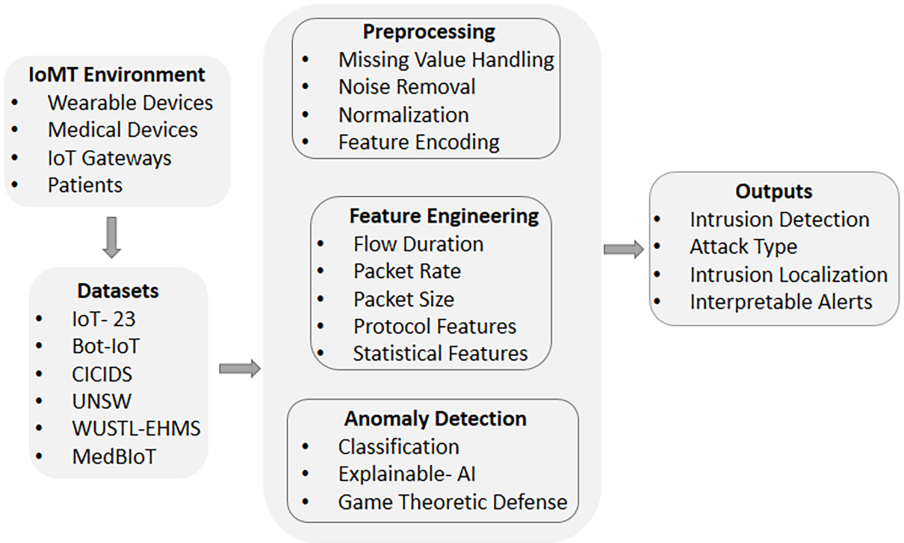

Figure 1: Work Flow of the proposed Ex-GAME framework.

3.1 Dataset Description

The framework implements a three-layer architecture that comprises device, edge, and core/cloud layers. The network environment is established using Python and MATLAB to combine deep-learning-based

intrusion detection with game-theoretic defense adaptation. Several publicly available cybersecurity datasets, including IoT-23 [8], Bot-IoT [9], CICIDS-2017 [10], and UNSW-NB15 [11], WUSTL-EHMS-2020 [12], and MedBIoT [13] are utilized for model implementation. These datasets provide diverse network traffic patterns across multiple attack scenarios. Table 1 shows dataset characteristics for different attack types. The sample class distribution has been shown in Table 2.

Table 1: Composition of the multi-dataset used for ExGAME evaluation.

Dataset	Instances	Features	Attack Categories	Domain
[8]	20M+	23	Botnet, DDoS, Malware	General IoT
[9]	72M+	46	DDoS, DoS, Data Theft	IoT Security
[10]	2.8M	80	Brute Force, DoS, Web Attacks	Networks
[11]	2.5M	49	Exploits, Fuzzers, DoS, Worms	Cybersecurity
[12]	16,318	44	Intrusion and Anomaly Events	Healthcare IoMT
[13]	Multiple	83+	Botnet, Malware Propagation	Medical IoT
Multi-Dataset (Proposed Work)	≈97M+	Unified	Comprehensive Attack Coverage	IoT + IoMT

Table 2: Sample class distribution (After Preprocessing).

Class	Instances	Percentage
Benign	60,000	60%
DoS	15,000	15%
Botnet	10,000	10%
Spoofing	8000	8%
Other Attacks	7000	7%

Table 3 shows the feature categories used for the training. The details are given for the categories and their examples.

Table 3: Feature categories used for training.

Category	Examples
Basic Features	Duration, Protocol, Packet Size
Content Features	Payload Length, Flags
Time-based Features	Flow Interval, Rate
Statistical Features	Mean, Variance, Entropy

3.2 System Model

Consider a Healthcare IoT network consisting of N medical devices. Each device generates network traffic samples $\mathbf{x}_i \in \mathbb{R}^m$ represented as feature vectors and m denotes the number of extracted features. The classifier output for enhanced learning objectives is defined as,

$$\hat{y}_i = f_\theta(\mathbf{x}_i) \tag{1}$$

The empirical risk over dataset $\mathcal{S}$ is given as,

$$\mathcal{R}(\theta) = \frac{1}{M}\sum_{i=1}^{M}\ell(f_\theta(\mathbf{x}_i), y_i) \tag{2}$$

where $\ell(\cdot)$ denotes binary cross-entropy and is represented as,

$$\ell(\hat{y}_i, y_i) = -y_i \log(\hat{y}_i) - (1 - y_i)\log(1 - \hat{y}_i) \tag{3}$$

The L_2 regularization is introduced to prevent overfitting as,

$$\Omega(\theta) = \frac{\lambda}{2}\|\theta\|_2^2 \tag{4}$$

The regularized learning objective becomes,

$$\mathcal{L}_{\text{reg}}(\theta) = \mathcal{R}(\theta) + \Omega(\theta) \tag{5}$$

The cost-sensitive learning is introduced to penalize false negatives more heavily in healthcare settings as,

$$\ell_{\text{cost}} = -\alpha y_i \log(\hat{y}_i) - \beta(1 - y_i)\log(1 - \hat{y}_i) \tag{6}$$

The total cost-sensitive risk can be given as,

$$\mathcal{L}_{\text{cost}} = \frac{1}{M}\sum_{i=1}^{M}\ell_{\text{cost}} \tag{7}$$

An adversarial robustness term is utilized to enhance robustness against adversarial perturbations using $\mathbf{x}_i^{adv}$,

$$\mathbf{x}_i^{adv} = \mathbf{x}_i + \varepsilon \cdot \text{sign}(\nabla_{\mathbf{x}}\ell) \tag{8}$$

where the robust objective can be given as,

$$\mathcal{L}_{\text{robust}} = \frac{1}{M}\sum_{i=1}^{M}\ell(f_\theta(\mathbf{x}_i^{adv}), y_i) \tag{9}$$

The final unified learning objective in terms of $\mathcal{L}_{\text{total}}, \mathcal{L}_{\text{cost}}, and\, \mathcal{L}_{\text{robust}}$ and optimal parameters $\theta^* = \arg\min_\theta \mathcal{L}_{\text{total}}$ is given as,

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cost}} + \Omega(\theta) + \gamma\mathcal{L}_{\text{robust}} \tag{10}$$

3.3 Explainability Modeling

The Shapley value for feature j and full feature set F is defined as,

$$\phi_j(\mathbf{x}) = \sum_{S\subseteq F\setminus\{j\}} \frac{|S|!(m - |S| - 1)!}{m!}\left[f(S\cup\{j\}) - f(S)\right] \tag{11}$$

The additive feature attribution with the simplified binary representation of input $\mathbf{x}'$ is given as,

$$g(\mathbf{x}') = \phi_0 + \sum_{j=1}^{m} \phi_j x'_j \tag{12}$$

The local fidelity loss is used to measure explanation fidelity in terms of a locality weighting kernel $\pi_{\mathbf{x}}(z)$ as

$$\mathcal{L}_{\text{fidelity}} = \sum_{z \in Z} \pi_{\mathbf{x}}(z) \left[f(z) - g(z) \right]^2 \tag{13}$$

The explanation sparsity constraint is defined to enforce human interpretability as,

$$\mathcal{R}_{\text{sparse}} = \lambda_1 \sum_{j=1}^{m} |\phi_j| \tag{14}$$

Algorithm 1 presents the training of the intrusion detection model using the optimized model parameters.

Algorithm 1: Intrusion detection model training

Require: Training dataset $D = (X, Y)$, learning rate λ, epochs T
Ensure: Optimized model parameters θ^*
  Initialize model parameters θ
  Define attack penalty weight w_a
  Define benign penalty weight w_b
  for $t = 1$ to T **do**
    for each training sample (x_i, y_i) in D **do**
      Predict label $\hat{y}_i \leftarrow f(x_i, \theta)$
      if $y_i = attack$ **then**
        Assign class weight $w_i \leftarrow w_a$
      else
        Assign class weight $w_i \leftarrow w_b$
      end if
      Compute weighted detection loss
      $\mathcal{L}_{det} \leftarrow w_i \cdot \ell(\hat{y}_i, y_i)$
      Update model parameters
      $\theta \leftarrow \theta - \lambda \nabla_\theta \mathcal{L}_{det}$
    end for
  end for
  $\theta^* \leftarrow \theta$
  return θ^*

The explanation stability under perturbation is given as,

$$\mathcal{R}_{\text{stable}} = \mathbb{E}_\delta \left[\| \Phi(\mathbf{x}) - \Phi(\mathbf{x} + \delta) \|_2^2 \right] \tag{15}$$

where the global explanation consistency is given as,

$$I_j = \frac{1}{M} \sum_{i=1}^{M} |\phi_j(\mathbf{x}_i)| \tag{16}$$

The integrated explainability objective in terms of $\mathcal{L}_{\text{exp}}$, $\mathcal{L}_{\text{fidelity}}$, $\mathcal{R}_{\text{sparse}}$, and $\mathcal{R}_{\text{stable}}$ is expressed as,

$$\mathcal{L}_{\text{exp}} = \mathcal{L}_{\text{fidelity}} + \mathcal{R}_{\text{sparse}} + \lambda_2 \mathcal{R}_{\text{stable}} \tag{17}$$

The explainability generation using SHAP has been presented in Algorithm 2. Table 4 shows the details of the attack type and its impact on different target layers.

Algorithm 2: Explainability generation using SHAP

Require: Trained model $f(\theta)$, input dataset X, perturbation set Δ, sparsity weight λ_1, stability weight λ_2
Ensure: Explanation matrix E and explainability loss $\mathcal{L}_{\text{exp}}$
Initialize explanation matrix $E \leftarrow \varnothing$
Initialize total fidelity loss $\mathcal{L}_{\text{fidelity}} \leftarrow 0$
Initialize sparsity penalty $\mathcal{R}_{\text{sparse}} \leftarrow 0$
Initialize stability penalty $\mathcal{R}_{\text{stable}} \leftarrow 0$
for each instance x_i in dataset X **do**
Generate perturbed samples $Z_i \leftarrow \{x_i + \delta \mid \delta \in \Delta\}$
Compute locality weights $\pi_{x_i}(z)$ for all $z \in Z_i$
Compute SHAP explanation $e_i \leftarrow SHAP(f, x_i)$
Append e_i to explanation matrix E
Train local surrogate model g_i using Z_i, $f(z)$, and $\pi_{x_i}(z)$
Compute local fidelity loss
$\mathcal{L}_{\text{fidelity}} \leftarrow \mathcal{L}_{\text{fidelity}} + \sum_{z \in Z_i} \pi_{x_i}(z) \left(f(z) - g_i(z)\right)^2$
Compute local sparsity penalty
$\mathcal{R}_{\text{sparse}} \leftarrow \mathcal{R}_{\text{sparse}} + \lambda_1 \sum_{j=1}^{d} |e_{ij}|$
Compute local stability penalty
$\mathcal{R}_{\text{stable}} \leftarrow \mathcal{R}_{\text{stable}} + \sum_{\delta \in \Delta} \left\| SHAP(f, x_i) - SHAP(f, x_i + \delta) \right\|_2^2$
for each feature j **do**
if $e_{ij} > 0$ **then**
Mark feature j as positive contributor
else
Mark feature j as negative contributor
end if
end for
end for
Compute total explainability loss
$\mathcal{L}_{\text{exp}} \leftarrow \mathcal{L}_{\text{fidelity}} + \mathcal{R}_{\text{sparse}} + \lambda_2 \mathcal{R}_{\text{stable}}$
return E, $\mathcal{L}_{\text{exp}}$

Table 4: Threat model in healthcare IoT.

Attack Type	Impact	Target Layer
Denial-of-Service (DoS)	Service disruption	Gateway/Core
Spoofing	Device impersonation	Device Layer
Botnet Activity	Coordinated attack	Network Layer
Data Injection	Medical record tampering	Core Server
Unauthorized Access	Privacy breach	All Layers

3.4 Integrated Learning–Explainability Optimization

A multi-objective learning function has been formulated that integrates classification loss, explainability regularization, and defender utility as,

$$\mathcal{J} = \mathcal{L}_{\text{total}} + \kappa \mathcal{L}_{\text{exp}} - \rho U_D \tag{18}$$

where $\mathcal{L}_{\text{total}}$ is the primary detection loss of the intrusion detection model, $\mathcal{L}_{\text{exp}}$ denotes the explainability loss, and U_D corresponds to the defender's utility obtained from the game-theoretic interaction. The trade-off between explainability and strategic defense effectiveness is controlled with the parameters κ and ρ. The joint objective is minimized to obtain the optimal model parameters as,

$$\theta^* = \arg\min_{\theta} \mathcal{J} \tag{19}$$

where θ represents the learnable parameters of the detection model. The attacker best response strategy has been presented in Algorithm 3.

Algorithm 3: Attacker best response strategy

Require: Defender strategy s_D, attacker strategies $\mathcal{S}_A$
Ensure: Optimal attacker strategy s_A^*
  $U_A^{max} \leftarrow -\infty$
  for each strategy s_A in $\mathcal{S}_A$ **do**
    Compute attacker success rate ASR
    Compute attacker cost C_A
    $U_A \leftarrow \delta \cdot ASR - \eta \cdot C_A$
    if $U_A > U_A^{max}$ **then**
      $U_A^{max} \leftarrow U_A$
      $s_A^* \leftarrow s_A$
    end if
  end for
  return s_A^*

3.4.1 Game-Theoretic Model

The strategic interaction between attackers and defenders in the medical IoT environment is modeled as a Stackelberg game, in which the defender acts as the leader and the attacker responds strategically. The defender's and the attacker's strategy space is defined as,

$$\mathcal{S}_D = \{s_D^1, s_D^2, \ldots, s_D^k\} \tag{20}$$

$$\mathcal{S}_A = \{s_A^1, s_A^2, \ldots, s_A^l\} \tag{21}$$

where each strategy s_D^i corresponds to a defensive configuration, and each attacker strategy represents possible intrusion behaviors.

3.4.2 Utility Functions

The benefits and costs of the strategies are captured using utility functions that model the interaction between the attacker and the defender. The defender and attacker utilities are formulated as

$$U_D = \alpha \cdot DR - \beta \cdot FPR - \gamma \cdot C_D \tag{22}$$

$$U_A = \delta \cdot ASR - \eta \cdot C_A \tag{23}$$

where DR denotes the detection rate, FPR represents the false positive rate, and C_D is the operational cost incurred by defensive mechanisms, and α, β, and γ are the weighting parameters. ASR is the attack success rate, and C_A denotes the cost associated with executing attacks. δ and η are the parameters that determine the attacker's preference between attack effectiveness and cost. The defender strategy optimization has been presented in Algorithm 4.

Algorithm 4: Defender strategy optimization

Require: Defender strategy set $\mathcal{S}_D$, attacker strategy s_A^*
Ensure: Optimal defender strategy s_D^*

```
U_D^max ← −∞
for each strategy s_D in S_D do
    Compute detection rate DR
    Compute false positive rate FPR
    Compute defense cost C_D
    U_D ← α · DR − β · FPR − γ · C_D
    if U_D > U_D^max then
      U_D^max ← U_D
      s_D^* ← s_D
    end if
end for
return s_D^*
```

3.4.3 Stackelberg Game Formulation

The interaction between the attacker and defender strategies is modeled as a Stackelberg game. The attacker's best-response strategy given the defender's chosen strategy is defined as

$$s_A^*(s_D) = \arg\max_{s_A \in \mathcal{S}_A} U_A(s_D, s_A) \tag{24}$$

The defender anticipates the response, and an optimal defensive strategy is selected to minimize the attacker's influence as,

$$\frac{dU_D}{ds_D} = \frac{\partial U_D}{\partial s_D} + \frac{\partial U_D}{\partial s_A^*}\frac{\partial s_A^*}{\partial s_D} \tag{25}$$

The dynamic Game-Theoretic strategy adaptation has been presented in Algorithm 5.

Algorithm 5: Dynamic stackelberg-based strategy adaptation

Require: Initial defender strategy s_D, attacker distribution p_A, iterations T
Require: Attack threshold τ_A, defense threshold τ_D, learning rates λ and μ
Ensure: Updated defender strategy s_D^* and attacker distribution p_A^*

for $t = 1$ to T **do**
  Determine attacker best response
  $s_A^*(s_D^t) \leftarrow \arg\max_{s_A \in \mathcal{S}_A} U_A(s_D^t, s_A)$
  Compute attacker utility
  $U_A \leftarrow U_A(s_D^t, s_A^*)$
  if $U_A > \tau_A$ **then**
    Attacker launches optimized strategy s_A^*
  else
    Attacker remains inactive
  end if
  Compute defender utility
  $U_D \leftarrow U_D(s_D^t, s_A^*(s_D^t))$
  Compute Stackelberg-aware defender gradient
  $G_D \leftarrow \frac{\partial U_D}{\partial s_D} + \frac{\partial U_D}{\partial s_A^*} \frac{\partial s_A^*}{\partial s_D}$
  if $U_D < \tau_D$ **then**
    Update defense configuration
    Retrain detection model if required
  else
    Maintain current defense configuration
  end if
  Update defender strategy
  $s_D^{t+1} \leftarrow s_D^t + \lambda G_D$
  Compute intermediate attacker distribution
  $\tilde{p}_A^{t+1} \leftarrow p_A^t + \mu \nabla U_A(s_D^t, s_A^*)$
  Project negative probabilities to zero
  $\tilde{p}_{A,k}^{t+1} \leftarrow \max(0, \tilde{p}_{A,k}^{t+1}), \quad \forall k$
  Normalize attacker distribution
  $p_{A,k}^{t+1} \leftarrow \frac{\tilde{p}_{A,k}^{t+1}}{\sum_{m=1}^{|\mathcal{S}_A|} \tilde{p}_{A,m}^{t+1}}, \quad \forall k$
end for
$s_D^* \leftarrow s_D^T$
$p_A^* \leftarrow p_A^T$
return s_D^*, p_A^*

3.4.4 Dynamic Adaptation

The proposed ExGAME framework incorporates dynamic strategy updates based on observed utilities to address evolving cyber threats in healthcare IoT networks. A gradient-based learning is used to iteratively update the defender strategy as,

$$s_D^{(t+1)} = s_D^{(t)} + \lambda \nabla U_D^{(t)} \tag{26}$$

where λ is the learning rate controlling the adaptation speed and $\nabla U_D^{(t)}$ denotes the gradient of the defender utility at iteration t. The strategy distribution is updated by the attacker in terms of the probability distribution p_A and attacker adaption control rate μ as

$$p_A^{(t+1)} = p_A^{(t)} + \mu \nabla U_A^{(t)} \tag{27}$$

3.4.5 Integrated Optimization Objective

The ExGAME framework employs a global objective function unifying learning, explainability, and strategic defense. The formulation is done using the intrusion detection loss $\mathcal{L}_{\text{det}}$, U_D to capture defender effectiveness in the strategic interaction, and the explainability regularization term $\mathcal{R}_{\text{exp}}$ that promotes interpretable model behavior as

$$\mathcal{J} = \mathcal{L}_{\text{det}} - \rho U_D + \kappa \mathcal{R}_{\text{exp}} \tag{28}$$

3.4.6 Equilibrium Conditions

Equilibrium conditions are used to characterize the optimal interaction between the attacker and the defender. The attacker and defender equilibrium is satisfied as,

$$U_A(s_D^*, s_A^*) \geq U_A(s_D^*, s_A), \quad \forall s_A \in \mathcal{S}_A \tag{29}$$

$$U_D(s_D^*, s_A^*) \geq U_D(s_D, s_A^*), \quad \forall s_D \in \mathcal{S}_D \tag{30}$$

The strategic interaction in the ExGAME framework is governed by these conditions in the Stackelberg equilibrium.

3.5 Unified Multi-Objective Optimization Framework

Detection accuracy, robustness, explainability, and strategic defense are optimized through a unified multi-objective optimization problem. The detection risk minimization is performed considering the empirical classification risk as,

$$\mathcal{R}_{\text{det}}(\theta) = \frac{1}{M} \sum_{i=1}^{M} \ell(f_\theta(\mathbf{x}_i), y_i) \tag{31}$$

where $\ell(\cdot)$ denotes the cost-sensitive cross-entropy loss. Generalization is improved using regularization as,

$$\mathcal{R}_{\text{reg}}(\theta) = \frac{\lambda}{2} \|\theta\|_2^2 \tag{32}$$

The regularized detection objective becomes,

$$\mathcal{L}_{\text{det}} = \mathcal{R}_{\text{det}} + \mathcal{R}_{\text{reg}} \tag{33}$$

Robustness regularization is performed to defend against adversarial perturbations. Worst-case risk minimization is done to enforce robustness as,

$$\mathcal{L}_{\text{rob}} = \frac{1}{M} \sum_{i=1}^{M} \max_{\|\delta_i\| \leq \varepsilon} \ell(f_\theta(\mathbf{x}_i + \delta_i), y_i) \tag{34}$$

The explanation fidelity term ensures explainability optimization to surrogate explanation model $g(\cdot)$ approximates the original model as,

$$\mathcal{L}_{\text{fid}} = \mathbb{E}_{\mathbf{z}\sim\pi_{\mathbf{x}}}\left[\left(f_\theta(\mathbf{z}) - g(\mathbf{z})\right)^2\right] \tag{35}$$

To promote sparsity and stability under input perturbations, we get,

$$\mathcal{R}_{\text{sparse}} = \lambda_1 \sum_{j=1}^{m} |\phi_j| \tag{36}$$

$$\mathcal{R}_{\text{stable}} = \mathbb{E}_\delta\left[\|\Phi(\mathbf{x}) - \Phi(\mathbf{x}+\delta)\|_2^2\right] \tag{37}$$

Hence, the total explainability objective becomes,

$$\mathcal{L}_{\text{exp}} = \mathcal{L}_{\text{fid}} + \mathcal{R}_{\text{sparse}} + \lambda_2 \mathcal{R}_{\text{stable}} \tag{38}$$

The strategic defense utility is defined as,

$$U_D = \alpha DR - \beta FPR - \gamma C_D \tag{39}$$

A penalty term $\mathcal{L}_{\text{game}} = -U_D$ is introduced to incorpor utility. The complete multi-objective function in terms od trade-off coefficients η_1, η_2, η_3 is defined as,

$$\mathcal{J}(\theta) = \mathcal{L}_{\text{det}} + \eta_1 \mathcal{L}_{\text{rob}} + \eta_2 \mathcal{L}_{\text{exp}} + \eta_3 \mathcal{L}_{\text{game}} \tag{40}$$

The optimal solution is obtained as,

$$\theta^* = \arg\min_\theta \mathcal{J}(\theta) \tag{41}$$

The constrained optimization form in terms of predefined thresholds controlling interpretability, robustness, and strategic utility τ_1, τ_2, τ_3 is expressed as,

$$\min_\theta \mathcal{L}_{\text{det}} \tag{42}$$

subject to:

$$\mathcal{L}_{\text{exp}} \le \tau_1, \quad \mathcal{L}_{\text{rob}} \le \tau_2, \quad U_D \ge \tau_3 \tag{43}$$

The constrained problem can be converted into Lagrangian form as,

$$\mathcal{L}_{\text{Lag}} = \mathcal{L}_{\text{det}} + \mu_1(\mathcal{L}_{\text{exp}} - \tau_1) + \mu_2(\mathcal{L}_{\text{rob}} - \tau_2) - \mu_3(U_D - \tau_3) \tag{44}$$

The saddle-point solution satisfies,

$$\theta^* = \arg\min_\theta \max_{\mu_1,\mu_2,\mu_3\ge 0} \mathcal{L}_{\text{Lag}} \tag{45}$$

Algorithm 6 shows the joint learning–explainability optimization. The computational complexity of the proposed Algorithms (1–6) has been presented in Table 5. The time and space complexity have been mentioned for all the algorithms. T is the number of epochs, N represents the number of samples, and d denotes the feature dimension and M represents the number of background samples.

Algorithm 6: Joint learning–explainability optimization

Require: Detection loss $\mathcal{L}_{det}$, explainability loss $\mathcal{L}_{exp}$, defender utility U_D
Ensure: Optimized objective $\mathcal{J}$
Compute explainability contribution
$E \leftarrow \kappa \cdot \mathcal{L}_{exp}$
Compute defense contribution
$D \leftarrow \rho \cdot U_D$
Compute joint objective
$\mathcal{J} \leftarrow \mathcal{L}_{det} + E - D$
if $\mathcal{J}$ decreases **then**
Accept model update
else
Re-adjust model parameters
end if
return $\mathcal{J}$

Table 5: Computational complexity analysis of the algorithms.

Algorithm	Time Complexity	Space Complexity
1	$O(T \times N \times d)$	$O(N \times d)$
2	$O(N \times d \times M)$	$O(N \times d)$
3	$O(\lvert\mathcal{S}_A\rvert)$	$O(1)$
4	$O(\lvert\mathcal{S}_D\rvert)$	$O(1)$
5	$O(T \times (\lvert\mathcal{S}_A\rvert + \lvert\mathcal{S}_D\rvert))$	$O(\lvert\mathcal{S}_A\rvert + \lvert\mathcal{S}_D\rvert)$
6	$O(N \times d)$	$O(d)$

4 Discussion and Analysis

This section provides a detailed discussion of the performance assessment for the proposed framework. The network architecture and simulation environment have been discussed, along with descriptions of parameters and values. The network configuration has been presented, detailing the number of devices, gateway nodes, network topology, and communication protocols. The model training parameters, including learning rate, batch size, and activation function, have also been mentioned. Assessment has been done for various performance metrics, including anomaly detection score, packet rate distribution, traffic intensity, and latency analysis.

4.1 Network Architecture and Simulation Setup

A three-layered healthcare IoT environment (device, edge, and cloud) has been considered within a smart hospital infrastructure. It consists of wearable sensors, implantable medical devices, bedside monitoring systems, imaging equipment, and centralized electronic health record (EHR) servers. Device

communication has been done with edge gateways and cloud infrastructure. The network configuration parameters have been presented in Table 6.

Table 6: Healthcare IoT network configuration.

Parameter	Value
Number of IoT Devices	100–500
Gateway Nodes	5
Core Server	1 Central Security Server
Network Topology	Hierarchical Star
Communication Protocols	MQTT, HTTP, TCP/IP
Bandwidth	100 Mbps–1 Gbps
Traffic Type	Periodic + Event-driven
Security Layer	TLS/SSL Encryption
IDS Deployment	Edge + Core Hybrid

The experiments were conducted using a standardized implementation pipeline consisting of data preprocessing, feature extraction, intrusion detection training, explainability analysis, and game-theoretic adaptation. The hardware includes Intel Core i9/Xeon processor, 32–64 GB RAM, and NVIDIA RTX 3080/A100 GPU. The model training and Game-Theoretic Simulation parameters are shown in Table 7.

Table 7: Model training and game-theoritic parameters.

Parameter	Value
Learning Rate	0.001
Batch Size	64
Epochs	50–100
Optimizer	Adam
Activation Function	ReLU/Sigmoid
Train-Test Split	70%–30%
Cross-Validation	5-fold
Game Type	Stackelberg Dynamic Game
Iterations	100 Time Steps
Defender Strategies	5
Attacker Strategies	5
Utility Weights (α, β, γ)	Tuned Experimentally
Adaptation Rate	0.01–0.1

4.2 Results Analysis

Fig. 2 presents the accuracy comparison of the proposed framework and a traditional IDS. The resilience of the proposed ExGAME intrusion detection system to adversarial perturbations in network traffic data has been measured. It is noticeable that the detection accuracy of both systems progressively diminishes as the perturbation strength increases. The proposed framework still exhibits markedly greater stability and maintains superior accuracy across all perturbation levels. The shaded area around the ExGAME curve shows the confidence interval, or the range of possible performance. These findings demonstrate that integrating

explainability and game-theoretic defense mechanisms strengthens the detection system's resilience against adversarial manipulations.

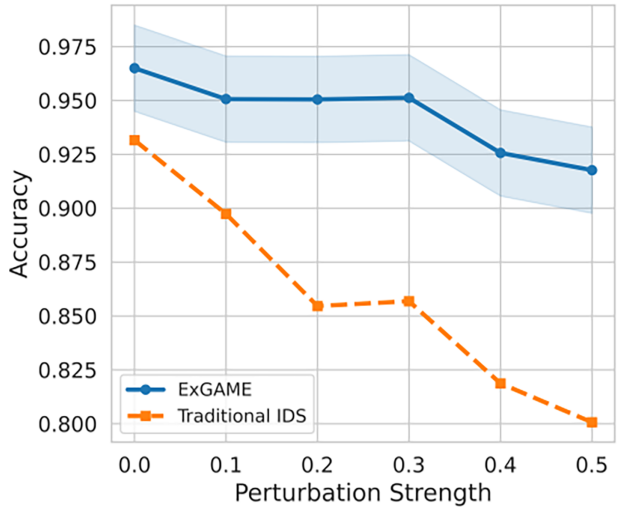

Figure 2: Robustness in terms of accuracy and perturbation strength.

Fig. 3 shows the attacker and defense utilities for different iterations. The defender's utility gradually increases as the number of iterations increases. The attacker's utility, on the other hand, decreases when the number of iterations increases. This indicates the convergence behavior of the proposed system. The game-theoretic model adopts a stable defense plan that minimizes the attacker's influence on the network.

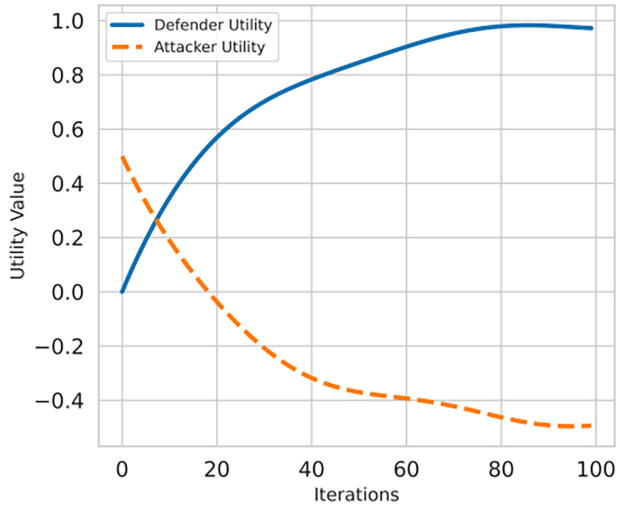

Figure 3: Attacker and defender utility for different iterations.

The consistency score and the dependability of the explanations over time have been shown in Fig. 4. The model processes diverse traffic data, gradually improving and stabilizing the consistency of the explanations. Small differences in explanation scores are presented by the darkened area around the curve. This result indicates that the explainability mechanism consistently provides consistent interpretations of model decisions.

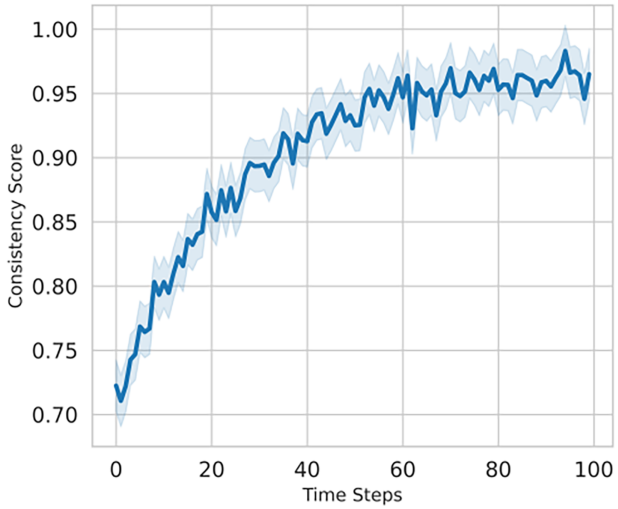

Figure 4: Explanation stability for various steps.

Fig. 5 presents a comparison of the proposed ExGAME and the IDS in terms of latency as the number of IoT devices increases. Both systems experience longer processing times as more devices connect due to increased traffic. The proposed framework shows lower latency growth, indicating that it performs well as more IoMT devices are connected. This improvement is due to efficient strategy-adaptation mechanisms across the edge and core layers.

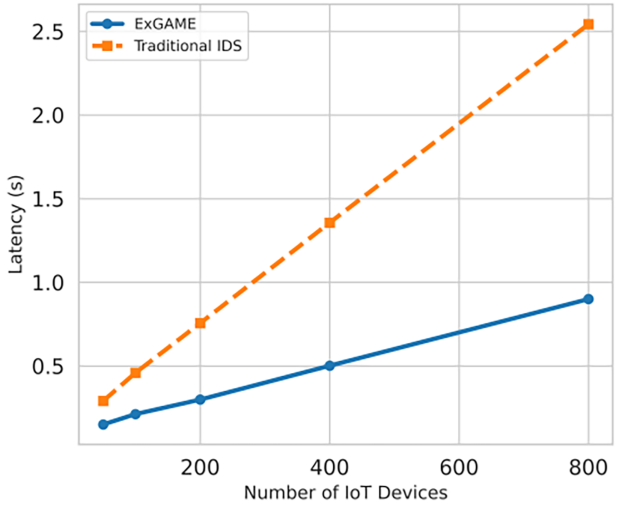

Figure 5: Comparison of latency for the number of IoT devices.

Fig. 6 shows the difference between normal and attack traffic for different packet rates. The results show that attack traffic sends more packets than normal traffic. It indicates the wider range of attack distributions during abnormal situations. The large number of packets transmitted in a short period is a common characteristic of distributed denial-of-service (DDoS) and botnet-driven traffic-flooding attacks.

A kernel density estimate (KDE) has been shown in Fig. 7 for both normal and attack traffic. Each curve indicates the probability distribution of packet rates. The filled density regions show clear differences between the two traffic classes. The normal distribution of packet rate values is more left-skewed than the

attack distribution. A reasonable overlap between the two distributions can be monitored in a certain range of packet rates. This observation reflects the importance of packet rate as a critical characteristic within the intrusion detection paradigm.

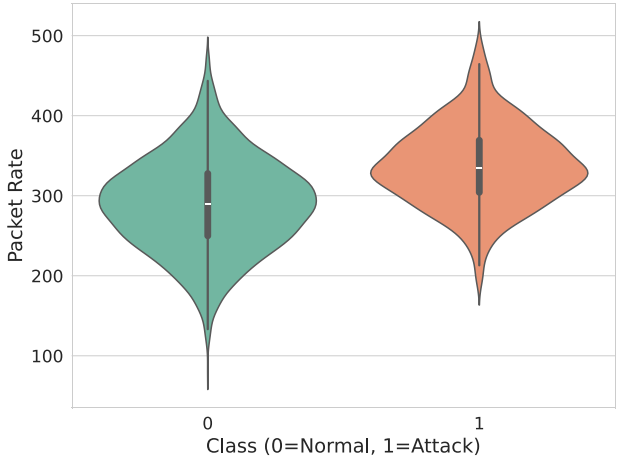

Figure 6: Packet rates for regular and attack traffic patterns.

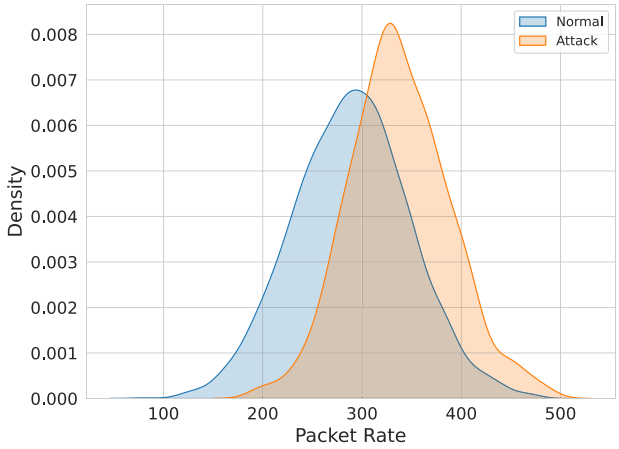

Figure 7: Kernel density estimate (KDE) for both normal and attack traffic.

Fig. 8 shows the SHAP summary plot of feature contributions. SHAP values for the data instances have been shown on the plot. These values show a characteristic contribution to the prediction outcome. The feature's overall importance is ranked by the model. The most important factor is the packet rate, showing the traffic nature, whether normal or attacked. source bytes and payload size also show their importance, but they have a smaller effect on the decision-making process.

Fig. 9 shows the SHAP dependence plot, which solely illustrates the connection between packet rate and SHAP values. The packet rate is on the horizontal axis, and the feature's effect on the model output is on the vertical axis. Each point represents one data point. The picture indicates that when packet rates increase, SHAP values frequently increase as well. This makes it more likely that people will view traffic as an attack. On the other hand, lower packet rates often lead to negative SHAP values, which push forecasts toward the normal traffic class.

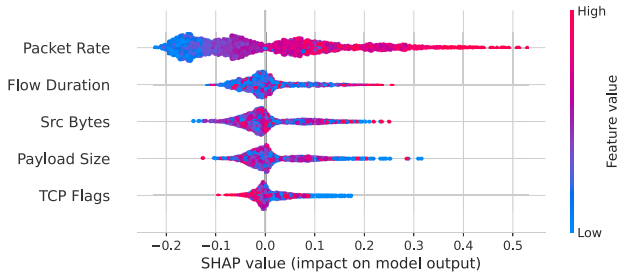

Figure 8: SHAP summary for various parameters.

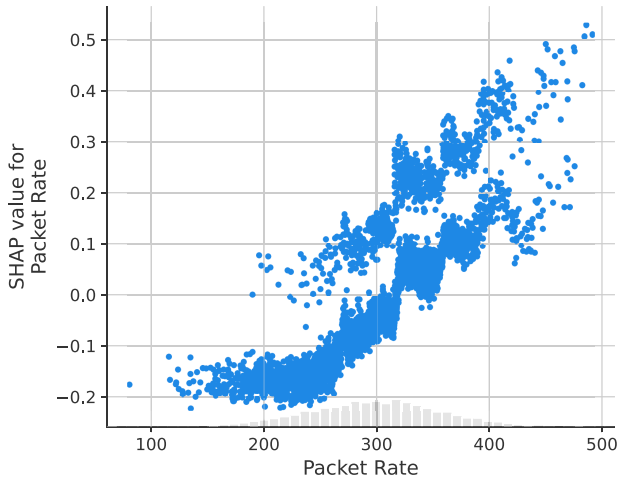

Figure 9: SHAP dependence plot for packet rates.

Fig. 10 shows anomaly scores with respect to time. The traffic behavior has been monitored around a threshold line. The periodic oscillations in the curve highlight the network traffic flows. The unusual activities or events are indicated by intermittent surges that come close to or exceed the threshold. This analysis plays an important role in identifying new attacks or unusual communication patterns.

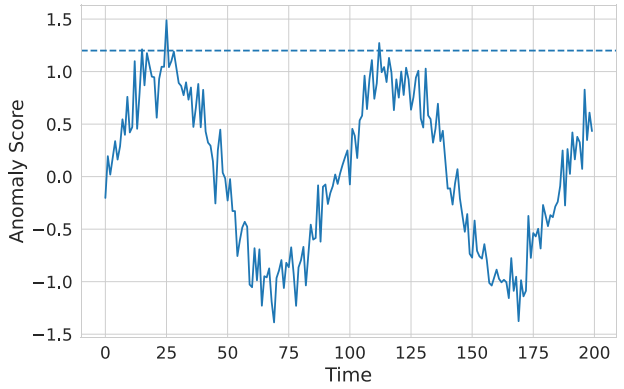

Figure 10: Anomaly score for various time slots.

Fig. 11 shows a heatmap of feature correlations and the relation of network traffic features in the dataset. Strong positive correlations are shown by the warmer colors, while weak or negative correlations are highlighted by the cooler colors. The features exhibit relatively low correlations with one another.

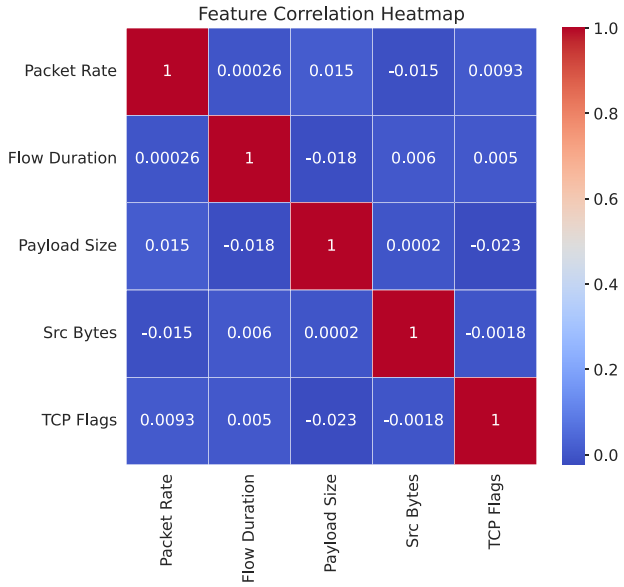

Figure 11: Correlation Heatmap of feature correlations.

Table 8 shows the comparison of the proposed framework with different state-of-the-art (SOTA) methods for various performance metrics. The Support Vector Machine (SVM) tested on the UNSW-NB15 dataset achieved 91.4% accuracy. The CICIDS2017 dataset was tested with a Random Forest and an XAI-Based IDS method, achieving 94.7% and 96.5% accuracy, respectively. Deep Neural Networks (DNN) and Long Short-Term Memory achieved accuracies of 96.2% and 97.1%, respectively. The proposed ExGAME framework outperformed all methods across multiple datasets, achieving 98.3% accuracy.

Table 8: Comparison with state-of-the-art intrusion detection methods.

Method	Dataset	Accuracy (%)	Precision (%)	Recall (%)
SVM-Based IDS	UNSW-NB15	91.4	90.2	89.8
Random Forest IDS	CICIDS2017	94.7	93.5	94.1
Deep Neural Network (DNN)	Bot-IoT	96.2	95.6	95.8
LSTM-Based IDS	IoT-23	97.1	96.8	96.5
XAI-Based IDS	CICIDS2017	96.5	95.9	95.7
ML/DL-IDS	MedBIoT	97.3	96.7	96.5
ML/DL-IDS	WUSTL-EHMS-2020	96.8	96.1	95.9
Proposed ExGAME	Multi-Dataset	98.3	97.9	97.6

The traffic intensity heatmap has been shown in Fig. 12 for different IoMT devices. Each device generates traffic over a different time period. Higher activity is indicated by the bright regions, while lower communication periods are indicated by the dark regions. The heatmap also shows changes in time-to-time

patterns from one device to another. The traffic-intensity heatmap is crucial in an IoT healthcare environment for patient monitoring and treatment planning.

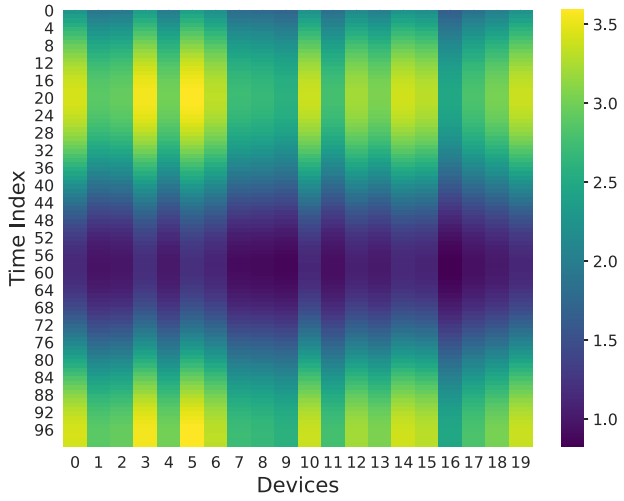

Figure 12: Traffic intensity heatmap for IoMT devices.

Table 9 shows the comparison of the real time performance with existing IDS framework. Different methods represented by various datasets have been compared for accuracy, latency, and throughput. Results show that the proposed framework outperformed several methods for the given quality of service parameters.

Table 9: Real-time performance comparison with existing IDS frameworks.

Method	Accuracy (%)	Latency (s)	Throughput (pkt/s)	Explainable
SVM-Based IDS	91.4	0.85	4200	No
Random Forest IDS	94.7	0.63	5100	No
DNN-Based IDS	96.2	0.58	5600	No
LSTM-Based IDS	97.1	0.71	4900	No
XAI-Based IDS	96.5	0.67	5200	Yes
Proposed ExGAME	98.3	0.42	6700	Yes

Fig. 13 shows a boxplot representation of the packet rate for normal and attack traffic types. The data's interquartile range is shown by the box parts. The median packet rate is shown by the middle line, the range of normal values is presented by the whiskers, and outliers are shown as single points. A higher median packet rate has been observed in the attacked traffic than in the normal traffic.

Table 10 presents a detailed evaluation of the proposed ExGAME framework for various performance metrics. The overall model accuracy was 98.3%, while the precision and recall scores were 97.9% and 97.6%, respectively. Results show that the model can successfully detect the abnormal traffic patterns, reducing false alarms in the IoMT environment. The F1-score and the false positive rate were 97.7% and 1.8%, respectively, indicating the model's robustness.

The features influencing intrusion detection are presented in Table 11. The packet rate reflects the most important feature, having a 0.41 score. It also indicates that abnormal network activity can be identified using this factor. Source bytes is the second most important factor, with a 0.23 score, followed by payload

size at 0.17. The rest of the features, including flow duration and TCP flags, are also important but with less importance values.

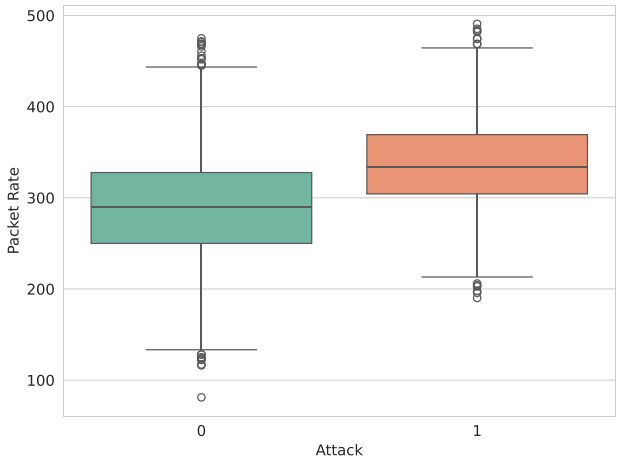

Figure 13: Boxplot results for normal and attack traffic.

Table 10: Performance evaluation of the proposed ExGAME framework.

Metric	Score
Accuracy	98.3%
Precision	97.9%
Recall	97.6%
F1-Score	97.7%
False Positive Rate	1.8%
Detection Latency	0.42 s
Inference Time	0.18 s
ROC-AUC	0.991
Throughput	6700 pkt/s

Table 11: Top features influencing intrusion detection.

Feature	Relative Importance
Packet Rate	0.41
Src Bytes	0.23
Payload Size	0.17
Flow Duration	0.12
TCP Flags	0.07

Table 12 shows the explainability evaluation of ExGAME. Various metrics, including explanation consistency, feature stability, interpretability score, and SHAP agreement, have been assessed. The SHAP agreement was observed to be higher than the other metrics. Table 13 presents the scalability analysis as IoMT devices increase. It has been observed that the accuracy drops by 0.6% as the number of devices increases from 50 to 800.

Table 12: Explainability evaluation of ExGAME.

Metric	Value
Explanation Consistency	96.4%
Feature Stability	95.8%
Interpretability Score	94.9%
SHAP Agreement	97.1%

Table 13: Scalability analysis under increasing IoMT devices.

Devices	Accuracy (%)	Latency (s)
50	98.5	0.18
100	98.4	0.24
200	98.3	0.31
400	98.1	0.39
800	97.9	0.52

The statistical robustness analysis of the proposed framework is presented in Table 14. It has been noticed that the model achieved high detection performance with low variability across repeated experimental runs. The *p*-value comparison of ExGAME with SVM was (<0.001), RF (0.002), DNN (0.008), and LSTM (0.015).

Table 14: Statistical robustness analysis.

Metric	Mean	Std. Dev.	95% CI
Accuracy (%)	98.30	0.28	[98.12, 98.48]
Precision (%)	97.90	0.31	[97.71, 98.09]
Recall (%)	97.60	0.35	[97.38, 97.82]
F1-Score (%)	97.70	0.29	[97.52, 97.88]
Latency (s)	0.42	0.03	[0.40, 0.44]

Table 15 presents an analysis of ExGAME as IoMT devices increase, covering scalability, resource utilization, and communication overhead. It has been observed that the accuracy decreases marginally from 98.5 to 97.9 as the number of devices increases from 50 to 800. Latency increases from 0.18 to 0.52 s due to strategy optimization. The CPU and memory usage also increase with network size. The traffic volume and messages per minute also increase steadily.

Table 15: Analysis of ExGAME under Increasing IoMT devices.

Devices	Acc (%)	Lat (s)	CPU Usage (%)	Memory (MB)	Traffic Vol (MB/min)	Messages/min
50	98.5	0.18	24	512	8.5	45
100	98.4	0.24	29	648	15.2	81
200	98.3	0.31	34	812	28.7	153
400	98.1	0.39	42	1056	54.1	292
800	97.9	0.52	57	1480	103.6	568

5 Conclusion

This paper introduces an Explainable Game-Theoretic Artificial Intelligence framework, termed ExGAME, designed to improve intrusion detection in IoMT settings while ensuring transparency and interpretability.

- The proposed system combines machine-learning-based anomaly detection with explainable AI methods and a game-theoretic defense strategy.
- A Random Forest classifier was used to identify anomalous traffic patterns based on key network parameters, including packet rate, flow duration, payload size, source bytes, and TCP flags.
- SHAP-based explanations were added to show how each feature affected the detection outcome.

Experimental investigation showed that packet rate and flow behavior, two traffic parameters, are good signs of malicious activity in IoMT networks. The findings demonstrate that the suggested methodology not only facilitates precise intrusion detection but also enhances the clarity and comprehensibility of security judgments. The ExGAME framework is a good approach for safe, clear intrusion detection in healthcare IoT settings. Future endeavors may focus on assessing the framework using extensive IoMT datasets, incorporating deep learning-driven detection models, and developing adaptive protection mechanisms that can respond to evolving cyber threats in real-time healthcare systems.

Acknowledgement: The authors extend their appreciation to Princess Nourah bint Abdulrahman University for funding this project.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R333), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Conceptualization: Umar Mujahid and Fahad Masood; Methodology: Noha Alnazzawi and Fahad Masood; Software: Nazik Alturki and Jawad Ahmad; Validation: Noha Alnazzawi and Nazik Alturki; Formal analysis: Nazik Alturki and Umar Mujahid; Investigation: Umar Mujahid and Jawad Ahmad; Resources: Nazik Alturki and Jawad Ahmad; Data curation: Fahad Masood and Jawad Ahmad; Writing—original draft preparation: Fahad Masood and Jawad Ahmad; Ceptualization: Umar Mujahid and Fahad Masood; Methodology: Nazik Alturki and Jawad Ahmad; Writing—review and editing: Nazik Alturki and Umar Mujahid; Visualization: Noha Alnazzawi and Nazik Alturki; Supervision: Jawad Ahmad. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: IoT-23 [8], Bot-IoT [9], CICIDS2017 [10], UNSW-NB15 [11], WUSTL-EHMS-2020, [12], and MedBIoT [13].

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Masood F, Khan MA, Alshehri MS, Ghaban W, Saeed F, Albarakati HM, et al. AI-based wireless sensor IoT networks for energy-efficient consumer electronics using stochastic optimization. IEEE Trans Consum Electron. 2024;70(4):6855–62. doi:10.1109/tce.2024.3416035.
2. Ghubaish A, Salman T, Zolanvari M, Unal D, Al-Ali A, Jain R. Recent advances in the internet-of-medical-things (IoMT) systems security. IEEE Inter Things J. 2020;8(11):8707–18. doi:10.1109/jiot.2020.3045653.
3. Razdan S, Sharma S. Internet of medical things (IoMT): overview, emerging technologies, and case studies. IETE Tech Rev. 2022;39(4):775–88.
4. Koutras D, Stergiopoulos G, Dasaklis T, Kotzanikolaou P, Glynos D, Douligeris C. Security in IoMT communications: a survey. Sensors. 2020;20(17):4828. doi:10.3390/s20174828.

5. El-Saleh AA, Sheikh AM, Albreem MA, Honnurvali MS. The internet of medical things (IoMT): opportunities and challenges. Wirel Netw. 2025;31(1):327–44. doi:10.1007/s11276-024-03764-8.
6. Al Malwi W, Masood F, Ahmad J, Asiri F, Alturki N, Al Hamadi H, et al. Quantum enhanced federated edge intelligence for cyber resilience in IoT remote sensing. IEEE J Sele Top App Earth Observat Remote Sens. 2026;19:16506–16. doi:10.1109/jstars.2026.3690574.
7. Kumar M, Singh SK, Kim S. Hybrid deep learning-based cyberthreat detection and IoMT data authentication model in smart healthcare. Fut Generat Comput Syst. 2025;166(11):107711. doi:10.1016/j.future.2025.107711.
8. Garcia S, Parmisano MJE, Erquiaga MJ. IoT-23: a labeled dataset with malicious and benign IoT network traffic. Prague, Czech: Stratosphere IPS Project, Czech Technical University; 2020 [cited 2026 Jun 15]. Available from: https://www.stratosphereips.org/datasets-iot23.
9. Koroniotis N, Moustafa N, Sitnikova E, Turnbull B. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-IoT dataset. Future Gener Comput Syst. 2019;100(7):779–96. doi:10.1016/j.future.2019.05.041.
10. Sharafaldin I, Lashkari AH, Ghorbani AA. Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018). Setúbal, Portugal: SciTePress; 2018. p. 108–16.
11. Moustafa N, Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In: 2015 Military Communications and Information Systems Conference (MilCIS). Piscataway, NJ, USA: IEEE; 2015. p. 1–6.
12. WUSTL EHMS 2020 dataset for Internet of Medical Things (IoMT) cybersecurity research. St. Louis, MO, USA: Washington University; 2024 Sep 7 [cited 2026 Jun 15]. Available from: https://www.cse.wustl.edu/~jain/ehms/index.html.
13. Guerra-Manzanares A, Medina-Galindo J, Bahsi H, Nõmm S. MedBIoT: generation of an IoT botnet dataset in a medium-sized IoT network. In: Proceedings of 6th International Conference on Information Systems Security and Privacy, ICISSP 2020; 2020 Feb 25–27; Valletta, Malta. p. 207–18.
14. Benmalek M, Seddiki A, Haouam KD. SNN-IoMT: a novel AI-driven model for intrusion detection in Internet of Medical Things. Comput Model Eng Sci. 2025;143(1):1157.
15. Khan A, Rizwan M, Bagdasar O, Alabdulatif A, Alamro S, Alnajim A. Deep learning-driven anomaly detection for IoMT-based smart healthcare systems. Comput Model Eng Sci. 2024;141(3):2121. doi:10.32604/cmes.2024.054380.
16. Al Mazroa A, Masood F, Awaji BH, Alhefdi M, Aljohani A, Ahmad J. FedGNN: federated graph neural networks for privacy-preserving cyber-resilient energy optimization in IoT-based smart grids. Compu Model Eng Sci. 2026;147(2):1–10. doi:10.32604/cmes.2026.080134.
17. Al-Hawawreh M, Hossain MS. A human-centered quantum machine learning framework for attack detection in IoT-based healthcare industry 5.0. IEEE Internet Things J. 2025;12(22):46065–74. doi:10.1109/jiot.2025.3565687.
18. Wazid M, Singh J, Das AK, Rodrigues JJPC. An Ensemble-based machine learning-envisioned intrusion detection in industry 5.0-driven healthcare applications. IEEE Trans Consum Electron. 2024;70(1):1903–12. doi:10.1109/tce.2023.3318850.
19. Oluwasusi VA, Modupeola TO, Azar NAN. Explainable AI-powered anomaly detection: a computer vision approach to strengthening human-centric cybersecurity. In: Proceedings of the International Conference on Artificial Intelligence and Cybersecurity (ICAIC 2025); 2025 Oct; Singapore: Springer Nature. p. 74–90.
20. Almogadwy B, Alqarafi A. Fused federated learning framework for secure and decentralized patient monitoring in healthcare 5.0 using IoMT. Sci Rep. 2025;15(1):24263. doi:10.1038/s41598-025-06574-w.
21. Baihan A, Kryvinska N, Amoon M, Jiang W, Ullah Z, Shafiq M. Cloud assisted blockchain-enabled split federated learning framework for security and privacy-preserving of IoMT in healthcare 5.0. Sci Rep. 2026;16(1):15599. doi:10.1038/s41598-026-41771-1.
22. Naik N, Surendranath N, Raju SAB. Hybrid deep learning-enabled framework for enhancing security, data integrity, and operational performance in Healthcare Internet of Things (H-IoT) environments. Sci Rep. 2025;15(1):31039. doi:10.1038/s41598-025-15292-2.

23. Rahmany M. AIDA: a health informatics-oriented adaptive cybersecurity framework for clinically aware detection and response in healthcare. Inform Health. 2026;3(1):152–9.
24. Lai Q, Hua H. Secure medical image encryption scheme for Healthcare IoT using novel hyperchaotic map and DNA cubes. Expert Syst Appl. 2025;264(1):125854. doi:10.1016/j.eswa.2024.125854.
25. Rezk NG, Alshathri S, Sayed A, Hemdan EE-D, El-Behery H. Secure hybrid deep learning for MRI-based brain tumor detection in smart medical IoT systems. Diagnostics. 2025;15(5):639. doi:10.3390/diagnostics15050639.
26. Alserhani F. Intrusion detection and real-time adaptive security in medical IoT using a cyber-physical system design. Sensors. 2025;25(15):4720. doi:10.3390/s25154720.
27. Berguiga A, Harchay A, Massaoudi A. HIDS-IoMT: a deep learning-based intelligent intrusion detection system for the internet of medical things. IEEE Access. 2025;13:32863–82.
28. Doménech J, León O, Siddiqui MS, Pegueroles J. Evaluating and enhancing intrusion detection systems in IoMT: the importance of domain-specific datasets. Inter Things. 2025;32(1):101631. doi:10.1016/j.iot.2025.101631.