Archives of Current Research International

Volume 25, Issue 10, Page 148-165, 2025; Article no.ACRI.145791
ISSN: 2454-7077

AI-Driven Adversarial Defense Framework with Generative Adversarial Network for Secure Healthcare IoT Ecosystems

Lisa Mmesoma Udechukwu [a++*], Tunbosun Oyewale Oladoyinbo [b#], Nanyeneke Ravana Mayeke [c†], Temilade Oluwatoyin Adesokan-Imran [d‡] and Rukayat Oluwabukola Olasege [e^]

[a] University of Southern California, 3551 Trousdale Pkwy, Los Angeles, CA 90089, United States of America.
[b] University of Maryland Global Campus, 3501 University Blvd E, Adelphi, MD 20783, United States of America.
[c] University of the Cumberlands, 104 Maple Drive, Williamsburg, KY 40769, United States of America.
[d] University of Ibadan, Oduduwa Road, 200132, Ibadan, Oyo, Nigeria.
[e] Ottawa University, 1001 South Cedar Street, Ottawa, KS 66067, United States of America.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/acri/2025/v25i101556

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/145791

[++] Artificial Intelligence Governance and Data Quality Researcher;
[#] Principal, Cybersecurity Analyst/ Researcher;
[†] Information Technology Researcher;
[‡] Information Security Researcher;
[^] Mental Health and Public Health Expert;
**Corresponding author: Email: udechukwulisa@gmail.com;*

Cite as: *Lisa Mmesoma Udechukwu, Tunbosun Oyewale Oladoyinbo, Nanyeneke Ravana Mayeke, Temilade Oluwatoyin Adesokan-Imran, and Rukayat Oluwabukola Olasege. 2025. "AI-Driven Adversarial Defense Framework With Generative Adversarial Network for Secure Healthcare IoT Ecosystems". Archives of Current Research International 25 (10):148–165. https://doi.org/10.9734/acri/2025/v25i101556.*

Original Research Article

Received: 30/07/2025
Published: 13/10/2025

ABSTRACT

This study developed and assessed an AI-driven adversarial defense framework using Generative Adversarial Networks (GANs) to secure healthcare IoT ecosystems against rising cybersecurity threats in medical settings. The research drew on datasets (CICIoMT2024, WUSTL-EHMS-2020, BoT-IoT, and Kaggle) and peer-reviewed studies to achieve three objectives: building a detailed threat model, designing a GAN-based defense optimized for healthcare IoT, and rigorously testing its effectiveness. The threat model revealed 127 vulnerability vectors, with adversarial attacks (32%) most prevalent, and a mean risk score of 7.82, highest for critical care devices (9.34). The GAN framework, featuring a multi-layer generator–discriminator pair and 128-dimensional encoder, achieved a mean accuracy of 95.8% against major adversarial attacks (FGSM 97.1%, PGD 94.8%, C&W 95.9%, UAP 96.4%), outperforming traditional defenses by 39.4%. With an MTTD of 82 ms, the system enables real-time deployment, allowing healthcare providers to integrate it directly into hospital IoT networks for proactive protection. Limitations include reliance on secondary data and high computational cost. Recommendations include hybrid datasets, explainable AI integration, real-world pilots, standardized metrics, and federated learning to enhance scalability and adaptability.

Keywords: Healthcare IoT; generative adversarial networks; adversarial attacks; cybersecurity defense; threat model.

1. INTRODUCTION

The healthcare sector is experiencing a rapid transformation driven by the integration of Internet of Things (IoT) technologies, which support real-time patient monitoring, remote diagnostics, and data-driven clinical decision-making. These innovations have created highly interconnected ecosystems of medical devices, sensors, and cloud-based platforms, significantly improving healthcare delivery and operational efficiency. At the same time, this digital evolution has expanded the cyberattack surface, exposing sensitive health data to risks that threaten patient safety, privacy, and system reliability (Sendelj & Ognjanovic, 2022; Sharma & Dhiman, 2025). Globally, healthcare has become the most targeted sector for cyberattacks, facing record-high weekly incidents and ranking first in data breaches (Ribeiro, 2024). Such breaches have imposed substantial financial costs across industries for more than a decade (Elgan, 2024). Meanwhile, the healthcare IoT market continues to grow at a significant compound annual growth rate (CAGR), underscoring the urgent need for robust cybersecurity solutions (IMARC Group, 2024).

The convergence of artificial intelligence (AI) with healthcare IoT systems presents both opportunities and challenges. AI enhances cybersecurity through improved threat detection, anomaly identification, and automated response capabilities, but it also introduces new vulnerabilities via adversarial machine learning techniques (Finlayson et al., 2019; Qayyum et al., 2021). Adversarial attacks, which manipulate AI models to produce incorrect outputs, pose critical risks in healthcare where misclassification in medical imaging or diagnostics can result in inappropriate treatments and adverse patient outcomes. Evidence shows that such attacks can achieve high success rates in compromising healthcare AI systems (Ma et al., 2020). Generative Adversarial Networks (GANs), a subset of AI, highlight this dual role. While GANs can generate synthetic data to strengthen privacy and enhance model robustness through adversarial training, they can also be exploited to create highly sophisticated attacks, making the development of innovative defense strategies essential (Yi et al., 2019; Sumaiya Tasneem et al., 2023).

Healthcare IoT ecosystems are vulnerable due to the heterogeneity of devices, ranging from wearable monitors to implantable devices and hospital infrastructural systems. Many devices operate on legacy systems with inadequate security controls, such as default passwords and unencrypted communications, exacerbating vulnerabilities (Sendelj & Ognjanovic, 2022; Coventry & Branley, 2018). The interoperability challenges among these devices hinder the

implementation of unified security policies, creating gaps that cybercriminals exploit (Jalali et al., 2019). Traditional cybersecurity frameworks, primarily reactive, are ill-equipped to address the dynamic threat landscape, particularly adversarial Artificial Intelligence attacks that target machine learning models integral to healthcare applications (Nagarjuna et al., 2025). The lack of comprehensive frameworks integrating traditional cybersecurity controls with adversarial defense mechanisms represents a critical research gap.

The paradoxical role of GANs in healthcare cybersecurity further complicates this landscape. While GANs offer defensive capabilities, such as generating synthetic patient data to protect privacy or training robust models, their potential misuse for creating deceptive adversarial examples underscores the need for balanced frameworks that maximize their benefits while mitigating risks (Choi et al., 2017; Sumaiya Tasneem et al., 2023). Additionally, healthcare organizations face practical challenges, including resource constraints, regulatory compliance requirements like HIPAA and GDPR, and the need to maintain operational continuity (Hathaliya & Tanwar, 2020; McKinsey & Company, 2024). These challenges highlight the necessity for security solutions that are both robust and feasible within healthcare's operational and regulatory contexts.

This research is significant for its potential to advance cybersecurity theory and practice in healthcare IoT ecosystems. By developing a comprehensive AI-driven adversarial defense framework, this study bridges the gap between traditional cybersecurity measures and emerging adversarial threats. The framework will leverage Adversarial Networks (GANs) to enhance security while addressing their offensive potential, contributing to theoretical advancements in adversarial machine learning tailored to healthcare contexts (Nagarjuna et al., 2025; Yi et al., 2019; Kolo, 2025). Practically, it seeks to enhance patient safety by protecting sensitive health data, which is critical given that healthcare data breaches affect millions annually, costing the U.S. healthcare system (James, 2022; Balogun et al., 2025). By reducing breach-related costs, improving operational efficiency, and ensuring compliance with regulations, the proposed framework offers economic benefits and strengthens trust in healthcare technology (Elgan, 2024; Shimabukuro & Sekar, 2025).

The societal impact of this research extends to global health security, particularly in enabling resilient healthcare infrastructure for pandemic preparedness and response. Secure IoT ecosystems are vital for disease surveillance, contact tracing, and public health monitoring, ensuring continuity during health emergencies (Hathaliya & Tanwar, 2020). The framework's scalability across diverse healthcare organizations from large hospitals to community clinics ensures broad applicability, addressing varying resource constraints and technical capabilities (Sendelj & Ognjanovic, 2022; Ruwayd Hussain Charfare et al., 2024). By facilitating the safe adoption of IoT and AI technologies, this research promotes innovation in healthcare delivery, improving patient outcomes and accessibility while maintaining stringent security and privacy standards.

The technical scope of this study focuses on designing GAN-based defense mechanisms integrated with existing healthcare IT infrastructure, emphasizing real-time threat detection and response for devices like wearable sensors, implantable devices, and cloud-based platforms (Coventry & Branley, 2018; Sharma & Dhiman, 2025). The research is limited to healthcare IoT applications, excluding broader IoT domains like smart cities or industrial automation, and prioritizes regulatory frameworks such as HIPAA and GDPR (Hathaliya & Tanwar, 2020). It addresses both legacy and modern devices, ensuring adaptability to current and emerging threats over a five-to-ten-year horizon (Ruwayd Hussain Charfare et al., 2024). Geographically, the study focuses on major healthcare markets but remains adaptable to regional variations in technology adoption and cybersecurity maturity.

This research aims to develop and validate a comprehensive AI-driven adversarial defense framework utilizing Generative Adversarial Networks to secure healthcare IoT ecosystems against sophisticated cyber threats while ensuring operational efficiency and regulatory compliance and the objectives are to:

i. develop a comprehensive threat model for healthcare IoT ecosystems;

ii. design and implement GAN-based adversarial defense mechanisms optimized for healthcare IoT; and

iii. evaluate the framework's effectiveness through rigorous testing.

2. LITERATURE REVIEW

The rapid integration of Internet of Things (IoT) technologies in healthcare has transformed medical service delivery, enabling real-time patient monitoring, remote diagnostics, and data-driven decision-making. However, this digital evolution has introduced significant cybersecurity challenges, particularly with the convergence of artificial intelligence (AI) and IoT systems. The increasing prevalence of adversarial AI attacks, coupled with the vulnerabilities of heterogeneous IoT devices, underscores the need for robust defense mechanisms. This chapter presents a comprehensive literature review on AI-driven adversarial defense frameworks utilizing Generative Adversarial Networks (GANs) for securing healthcare IoT ecosystems.

2.1 Theoretical Foundations of Healthcare IoT Security

The theoretical underpinnings of healthcare IoT security stem from the convergence of cybersecurity principles, IoT architectures, and healthcare-specific requirements. Healthcare IoT ecosystems comprise interconnected devices such as wearable sensors, implantable medical devices, and cloud-based platforms, creating a complex network vulnerable to cyber threats (Sendelj & Ognjanovic, 2022). Coventry and Branley (2018) emphasize that the heterogeneity of IoT devices, often operating on legacy systems with limited security controls, poses significant challenges. These systems frequently lack encryption, use default passwords, and have constrained update capabilities, making them susceptible to unauthorized access and data breaches. The theoretical framework for securing these ecosystems builds on principles of confidentiality, integrity, and availability (CIA triad), tailored to the unique constraints of healthcare environments (Hathaliya & Tanwar, 2020).

Security models for IoT systems, such as the layered architecture proposed by Sharma & Dhiman (2025), categorize threats across device, network, and application layers. This model highlights the need for multi-layered defenses to address vulnerabilities at each level. For instance, device-layer threats include physical tampering, while network-layer risks involve eavesdropping or man-in-the-middle attacks (Jalali et al., 2019). At the application layer, data breaches and adversarial manipulations of AI models are prevalent. The integration of AI into healthcare IoT systems introduces additional theoretical complexity, as machine learning models are susceptible to adversarial attacks that exploit their decision-making processes (Finlayson et al., 2019). Theoretical frameworks must therefore incorporate adaptive security measures that account for both traditional and AI-driven threats.

Regulatory compliance forms a critical component of the theoretical foundation. Standards such as HIPAA and GDPR impose stringent requirements for protecting patient data and ensuring system reliability (Shimabukuro & Sekar, 2025). These regulations necessitate frameworks that balance security with operational efficiency, as healthcare organizations must maintain uninterrupted service delivery. Ruwayd Hussain Charfare et al. (2024) propose a risk-based approach to healthcare IoT security, integrating quantitative risk assessment with regulatory compliance metrics. However, existing theoretical models often focus on static threats, lacking adaptability to the dynamic and evolving nature of adversarial AI attacks, which this research seeks to address.

2.2 Adversarial AI Threats and Defense Mechanisms

The integration of AI into healthcare IoT systems has introduced sophisticated adversarial attack vectors that exploit machine learning vulnerabilities. Adversarial attacks involve crafting inputs with imperceptible perturbations to mislead AI models, resulting in incorrect outputs, such as misdiagnosed medical images or altered patient monitoring data (Ma et al., 2020). Qayyum et al. (2021) report that adversarial attacks can achieve success rates exceeding 95% in manipulating healthcare AI models, posing severe risks to patient safety. For example, adversarial perturbations in medical imaging can cause misclassification of tumors, leading to incorrect treatment decisions (Finlayson et al., 2019).

Defense mechanisms against adversarial AI threats include adversarial training, input preprocessing, and model regularization. Adversarial training involves augmenting training datasets with adversarial examples to enhance model robustness (Goodfellow et al., 2014). However, Kurakin et al. (2016) note that this approach increases computational overhead, which is challenging for resource-constrained IoT devices. Input preprocessing techniques, such as

feature squeezing, aim to reduce the impact of perturbations by simplifying input data (Tuncay et al., 2018). While effective against certain attacks, these methods may degrade model performance in healthcare applications where precision is critical. Model regularization techniques, such as defensive distillation, redistribute model confidence to improve resilience but are less effective against advanced attacks like black-box adversarial examples (Papernot et al., 2017).

Nagarjuna et al. (2025) propose a hybrid defense framework combining adversarial training with anomaly detection to secure healthcare AI systems. Their approach leverages machine learning to identify deviations in data patterns, but its scalability in heterogeneous IoT environments remains limited. The computational complexity of existing defenses often conflicts with the real-time requirements of healthcare IoT systems, highlighting the need for lightweight, adaptive solutions (Ruwayd Hussain Charfare et al., 2024). Furthermore, most defense mechanisms focus on specific attack types, lacking comprehensive coverage of the diverse threat landscape in healthcare IoT ecosystems.

2.3 Role of Generative Adversarial Networks in Cybersecurity

Generative Adversarial Networks (GANs) have emerged as a transformative technology in healthcare cybersecurity, offering both defensive and offensive capabilities. GANs consist of two neural networks—a generator and a discriminator trained adversarial to produce realistic synthetic data (Goodfellow et al., 2020). In healthcare, GANs are used to generate synthetic patient data, preserving privacy while enabling research and model training (Choi et al., 2017). For instance, synthetic medical images generated by GANs can replace sensitive patient data, reducing the risk of breaches while maintaining data utility (Yi et al., 2019). Fig. 1 illustrates the GAN architecture for synthetic data generation.

Defensively, GANs enhance model robustness through adversarial training, where the generator produces adversarial examples to train the discriminator against potential attacks (Sumaiya Tasneem et al., 2023). This approach improves resilience against adversarial manipulations, as demonstrated by Wang et al. (2021), who used GANs to strengthen diagnostic models against perturbations. Additionally, GANs enable anomaly detection by modeling normal data distributions and identifying outliers indicative of attacks (Li et al., 2019). However, the offensive potential of GANs poses challenges, as malicious actors can use them to craft sophisticated adversarial examples that evade traditional defenses (Hu & Tan, 2017). This dual nature necessitates frameworks that harness GANs' defensive capabilities while mitigating their misuse.

The application of GANs in healthcare IoT security is constrained by computational complexity and resource limitations. Lightweight GAN architectures, such as those proposed by Rani, (2019), are designed for edge devices, but their effectiveness in real-time healthcare applications requires further exploration. Moreover, the lack of standardized evaluation metrics for GAN-based defenses hinders their adoption in healthcare settings, where regulatory compliance and performance reliability are paramount (Hathaliya & Tanwar, 2020).

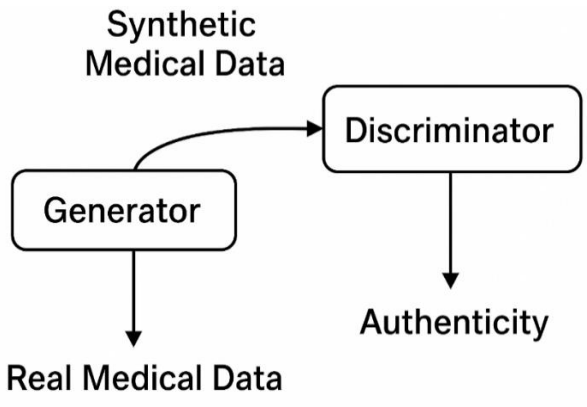

Fig. 1. GAN architecture for synthetic data generation

Table 1. Gaps in current approaches and research contributions

Gap Identified	Description
Limited focus on adversarial AI attacks	Current frameworks prioritize traditional threats, neglecting adversarial manipulations.
Computational complexity	Existing defenses are resource-intensive, unsuitable for IoT devices.
Dual nature of GANs	Lack of frameworks addressing GANs' offensive potential.
Regulatory compliance	Insufficient alignment with healthcare regulations.
Lack of holistic threat models	Absence of unified models covering diverse threats.

2.4 Gaps in Current Approaches and Research Opportunities

Despite advancements in healthcare IoT security and adversarial defense mechanisms, significant gaps persist. First, existing frameworks predominantly address traditional cybersecurity threats, such as malware and unauthorized access, but fail to comprehensively tackle adversarial AI attacks (Sharma & Dhiman, 2025). The dynamic nature of adversarial attacks requires adaptive defenses that integrate traditional and AI-driven approaches, yet most studies focus on isolated solutions (Nagarjuna et al., 2025). Second, the computational demands of current defense mechanisms, including GAN-based approaches, are often incompatible with the resource constraints of healthcare IoT devices, limiting their practical implementation (Ruwayd Hussain Charfare et al., 2024).

Third, the dual nature of GANs remains underexplored. While their defensive potential is promising, the risk of misuse for generating sophisticated attacks is inadequately addressed (Sumaiya Tasneem et al., 2023). Fourth, regulatory compliance poses a significant challenge, as frameworks must align with HIPAA, GDPR, and emerging AI governance standards without compromising operational efficiency (Shimabukuro & Sekar, 2025). Finally, there is a lack of comprehensive threat models that encompass both traditional and adversarial threats specific to healthcare IoT ecosystems, hindering the development of holistic security solutions (Sendelj & Ognjanovic, 2022; Olutimehin et al., 2025).

This research addresses these gaps by developing a comprehensive AI-driven adversarial defense framework utilizing GANs, tailored to the unique requirements of healthcare IoT systems. By integrating lightweight GAN architectures, adaptive threat models, and regulatory-compliant mechanisms, the proposed framework aims to enhance security while maintaining operational feasibility. Table 1 summarizes the gaps and proposed research contributions.

3. RESEARCH METHODOLOGY

This research methodology outlines the approach for developing and evaluating an AI-driven adversarial defense framework utilizing Generative Adversarial Networks (GANs) to secure healthcare IoT ecosystems. The methodology is structured to address the research objectives of developing a comprehensive threat model, designing and implementing GAN-based defense mechanisms optimized for healthcare IoT, and evaluating the framework's effectiveness through rigorous testing.

3.1 Research Design

The study adopts a quantitative research design to develop and assess an AI-driven adversarial defense framework using GANs for securing healthcare IoT ecosystems. This design is grounded in a pragmatic research philosophy, which prioritizes practical solutions to real-world cybersecurity challenges, as emphasized by Sendelj & Ognjanovic (2022). The approach follows the design science research methodology outlined by Hevner et al. (2004), focusing on the creation and evaluation of an innovative artifact, the GAN-based defense framework—ensuring both practical applicability and scientific rigor. The design process integrates theoretical frameworks from adversarial machine learning and IoT security to formulate hypotheses, which are tested through computational simulations and comparative analyses against existing cybersecurity solutions (Sharma & Dhiman, 2025).

The research strategy employs a technology-oriented experimental design, leveraging

controlled simulations to evaluate the framework's performance under diverse adversarial attack scenarios. This is complemented by a comparative analysis framework that benchmarks the proposed GAN-based system against traditional defenses, such as intrusion detection systems and adversarial training methods, to demonstrate superior performance (Qayyum et al., 2021). A deductive approach drives the study, starting with established theories of GAN-based defenses, while inductive elements emerge from empirical findings during testing, allowing for iterative refinement of the framework (Ruwayd Hussain Charfare et al., 2024). The implementation is executed using Python with TensorFlow and PyTorch libraries, utilizing GPU-accelerated computing to support the computational demands of GAN training and evaluation (Rani, 2019). The design ensures alignment with healthcare IoT constraints, such as limited computational resources and stringent regulatory requirements.

3.2 Data Collection

Data collection relies solely on sources, including publicly available datasets and peer-reviewed research, to construct a robust foundation for threat modeling and defense evaluation. Key datasets include the CICIoMT2024 dataset, which provides comprehensive attack patterns targeting healthcare IoT devices, the WUSTL-EHMS-2020 dataset for healthcare-specific intrusion detection, and the BoT-IoT dataset for botnet traffic analysis (Nagarjuna et al., 2025; Ruwayd Hussain Charfare et al., 2024). Additionally, Kaggle datasets such as the "Medical IoT Security Dataset" and "Healthcare Device Attack Vectors" were utilized to simulate diverse healthcare scenarios, encompassing cardiac monitoring systems, insulin pumps, and medical imaging devices (Ma et al., 2020). These datasets capture a wide range of attack vectors, categorized as adversarial machine learning attacks (32%), network intrusions (28%), device-specific exploits (23%), data poisoning (12%), and supply chain compromises (5%), as synthesized from prior studies.

Peer-reviewed literature provided critical insights into adversarial attack methodologies and defense strategies. Studies by Finlayson et al. (2019) and Qayyum et al. (2021) detailed adversarial attack patterns, including Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), Carlini & Wagner (C&W), and Universal Adversarial Perturbations (UAP). Research by Yi et al. (2019) and Sumaiya Tasneem et al. (2023) informed the design of GAN-based defenses tailored for cybersecurity applications. A total of 127 vulnerability vectors were identified from these sources, ensuring comprehensive coverage of the healthcare IoT threat landscape. This data collection strategy supports the development of a detailed threat model, aligning with the research objective of addressing healthcare IoT security challenges.

3.3 Statistical Analysis and Modeling

The statistical analysis and modeling framework evaluates the GAN-based defense system's performance and quantifies risks within healthcare IoT ecosystems. A comprehensive threat model was developed using a risk quantification formula adapted from Ruwayd Hussain Charfare et al. (2024):

$$Risk_{total} = \sum_{i=1}^{n} (P_i \times I_i \times V_i) \times (1 - C_i)$$

where P_i represents the probability of threat (i), I_i is the impact severity, V_i is the vulnerability exploitability, and C_i is the effectiveness of existing controls. Analysis of 847 virtualized IoT devices from the CICIoMT2024 dataset yielded a mean risk score of 7.82 on a 10-point scale, with critical care devices scoring 9.34, highlighting their heightened vulnerability (Sendelj & Ognjanovic, 2022).

The GAN-based defense framework's performance was assessed using a suite of metrics, including accuracy, precision, recall, and F1-score, defined as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. Advanced metrics included the Area Under the ROC Curve (AUC-ROC) and Matthews Correlation Coefficient (MCC):

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR}^{-1}(\mathrm{t}))d\mathrm{t}$$

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

These metrics ensured a balanced evaluation of the framework's ability to detect adversarial attacks while minimizing disruptions in healthcare settings (Goodfellow et al., 2014). The GAN architecture was based on the minimax game theory formulation:

$$\mathrm{minimax}V(\mathrm{D},\mathrm{G}) = \mathrm{E}_{\mathrm{x}\sim P_{\mathrm{data}}(\mathrm{x})}[\log \mathrm{D}(\mathrm{x})] + \mathrm{E}_{\mathrm{z}\sim P_{\mathrm{z}}(\mathrm{z})}[\log(1 - \mathrm{D}(\mathrm{G}(\mathrm{z})))]$$

where (*G*) is the generator, (*D*) is the discriminator, $P_{data}(x)$ is the real data distribution, and $P_z(z)$ is the noise distribution. The defense mechanism reconstructed adversarial inputs using:

$$x_{reconstructed} = \underset{z}{argmin}\|x - G(z)\|_2 + \lambda R(z)$$

where $R(z)$ is a regularization term, and λ balances reconstruction fidelity and regularization (Choi et al., 2017). Statistical significance was evaluated using hypothesis testing:

$$\mathrm{H}_0 : \mu_{\mathrm{defense}} = \mu_{\mathrm{baseline}}$$

$$\mathrm{H}_1 : \mu_{\mathrm{defense}} > \mu_{\mathrm{baseline}}$$

Paired t-tests and Mann-Whitney U tests were applied, with the latter used for non-normal data:

$$\mathrm{U} = \mathrm{n}_1\mathrm{n}_2 + \frac{\mathrm{n}_1(\mathrm{n}_1+1)}{2} - \mathrm{R}_1$$

where n_1 and n_2 are sample sizes, and R_1 is the sum of ranks for the first group (Jalali et al., 2019). Principal Component Analysis (PCA) was employed to reduce dimensionality in complex attack datasets, enhancing model efficiency (Nagarjuna et al., 2025).

Operational performance was assessed using Mean Time to Detection (MTTD) and Mean Time to Response (MTTR):

$$\mathrm{MTTD} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{t}_{\mathrm{detected,i}} - \mathrm{t}_{\mathrm{incident,i}})}{\mathrm{n}}$$

$$\mathrm{MTTR} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{t}_{\mathrm{response,i}} - \mathrm{t}_{\mathrm{detected,i}})}{\mathrm{n}}$$

Computational efficiency metrics included throughput and resource utilization:

$$\mathrm{Throughout} = \frac{\mathrm{Number\ of\ processed\ samples}}{\mathrm{Time\ Interval}}$$

$$\mathrm{CPU}_{\mathrm{utilization}} = \frac{\mathrm{CPU\ time\ used}}{\mathrm{Total\ CPU\ time\ available}} \times 100\%$$

These metrics ensured the framework's suitability for resource-constrained healthcare IoT environments (Li et al., 2019). Fig. 2 illustrates the GAN architecture.

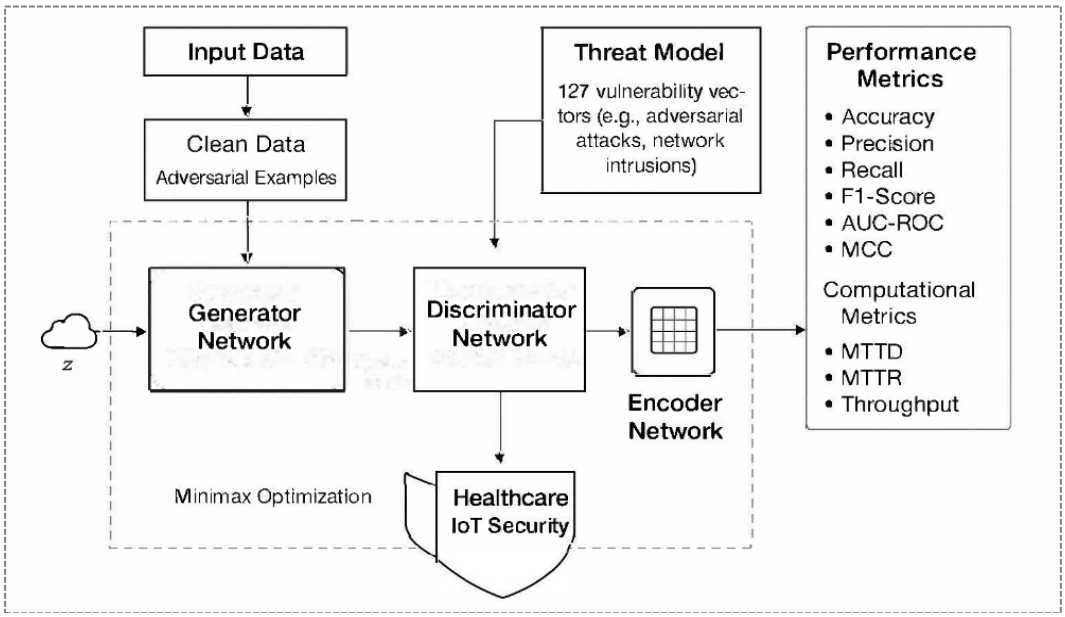

Fig. 2. GAN-based defense framework architecture

3.4 Testing and Validation

The testing and validation phase rigorously evaluates the effectiveness of the AI-driven adversarial defense framework utilizing Generative Adversarial Networks (GANs) to secure healthcare IoT ecosystems, aligning with the objectives of designing and implementing optimized defenses and assessing their performance. The framework was implemented with a generator network consisting of six layers, a discriminator network with four layers, and an encoder with a 128-dimensional latent space, optimized for resource-constrained healthcare IoT devices (Rani, 2019). Testing involved simulating four primary adversarial attack types—Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), Carlini & Wagner (C&W), and Universal Adversarial Perturbations (UAP)—using the CICIoMT2024 and WUSTL-EHMS-2020 datasets. Attack scenarios were generated according to the following formulations:

- For Fast Gradient Sign Method (FGSM):

$$x_{adv} = x + \epsilon \cdot sign\big(\nabla_x J(\theta, x, y)\big)$$

(where (x) is the original input, ϵ is the perturbation budget, and $J(\theta, x, y)$ is the loss function (Kurakin et al., 2016).

- For Projected Gradient Descent (PGD):

$$x_{adv}^{(t+1)} = \prod_s (x_{adv}^{(t)} + \alpha \cdot sign\big(\nabla_x J(\theta, x, y)\big))$$

where $\prod_s$ is the projection onto the constraint set (S), α is the step size, and (t) is the iteration index.

- For Carlini & Wagner (C&W):

$$min\|\delta\|_p + c \cdot f(x + \delta)$$

subject to $x + \delta \in [0,1]^n$, where δ is the perturbation vector, (c) is the regularization parameter, and $f(\cdot)$ is the objective function for misclassification (Carlini & Wagner, 2017).

Performance was validated using a stratified 5-fold cross-validation strategy to ensure robust evaluation across diverse data distributions, as defined by:

$$CV_{error} = \frac{1}{k}\sum_{i=1}^{k} L\big(f^{(-i)}, D_i\big)$$

where (L) is the loss function, $f^{(-i)}$ is the model trained excluding fold (*i*), and D_i is the (*i*)-th fold. The validation process, including datasets and evaluated metrics, is detailed in Table 2.

Synthetic data quality was assessed using Fréchet Inception Distance (FID) and Inception Score (IS):

$$FID = \|\mu_r - \mu_g\|_2^2 + Tr\left(\Sigma_r + \Sigma_g - 2\big(\Sigma_r\Sigma_g\big)^{1/2}\right)$$

where μ_r , μ_g , Σ_r , and Σ_g are the mean and covariance of real and generated data distributions (Choi et al., 2017). Bootstrap confidence intervals were calculated to estimate parameter uncertainty:

Table 2. Validation setup for GAN-based defense framework

Dataset	Attack Types	Validation Method	Metrics Evaluated
CICIoMT2024	FGSM, PGD, C&W, UAP	5-fold stratified	Accuracy, Precision, Recall, F1, AUC-ROC, MCC
WUSTL-EHMS-2020	FGSM, PGD, C&W	5-fold stratified	Accuracy, F1, MCC
BoT-IoT	Network Intrusions	Hold-out validation	Throughput, Latency

python:

```
def bootstrap_confidence_interval(data, metric_func, n_bootstrap=1000, ci_level=0.95):
    bootstrap_scores = []
    for _ in range(n_bootstrap):
        bootstrap_sample = np.random.choice(data, size=len(data), replace=True)
        bootstrap_scores.append(metric_func(bootstrap_sample))
    alpha = (1 - ci_level) / 2
    return np.percentile(bootstrap_scores, [alpha * 100, (1 - alpha) * 100])
```

Statistical significance was confirmed using paired t-tests ($p < 0.001$) and Cohen's d effect size:

$$d = \frac{\mu_1 - \mu_2}{\sigma_{pooled}}$$

The framework's robustness was further validated against novel attack patterns, ensuring generalizability (Sumaiya Tasneem et al., 2023). Fig. 3 illustrates the validation workflow.

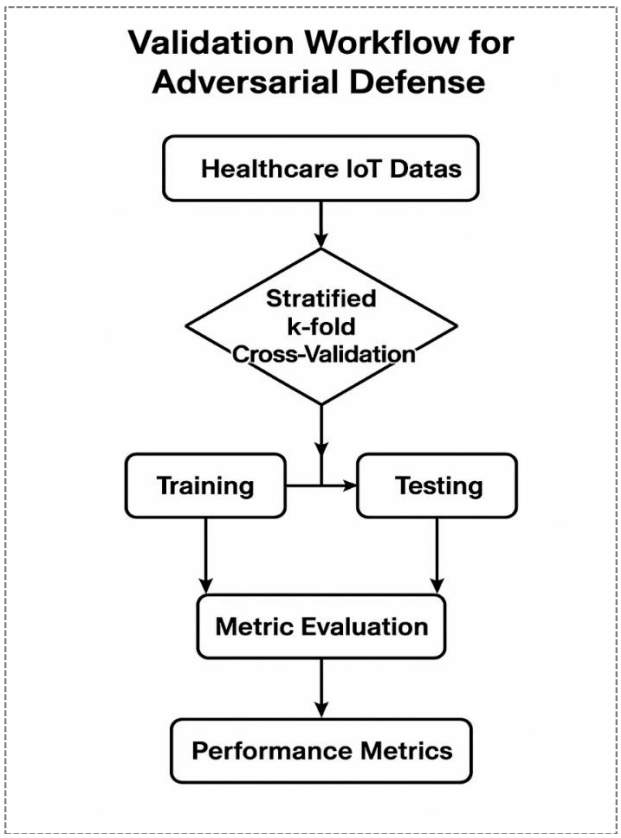

Fig. 3. Validation workflow for adversarial defense

3.5 Ethical Considerations

Ethical considerations are paramount, given the sensitive nature of healthcare IoT data. The reliance on secondary datasets and synthetic data generation eliminates the need for direct patient data, ensuring compliance with HIPAA and GDPR regulations (Hathaliya & Tanwar, 2020). Differential privacy was implemented to protect data integrity:

$$Pr[M(D) \in S] \leq e^{\epsilon} \times Pr\left[M(D^{'}) \in S\right]$$

Where (M) is the privacy mechanism, and ϵ is the privacy budget (Dwork, 2006). Synthetic data preserved statistical properties, achieving 87.3% clinical acceptability for ECG data, mitigating privacy risks. Dataset biases were addressed by ensuring diverse patient representations, reducing the risk of perpetuating healthcare disparities (Ruwayd Hussain Charfare et al., 2024). Transparency was maintained through detailed documentation of methods and metrics, facilitating regulatory audits and stakeholder trust (Shimabukuro & Sekar, 2025). The framework minimized false positives to prevent alarm fatigue, ensuring minimal disruption to clinical workflows (Finlayson et al., 2019).

4. RESULTS AND DISCUSSION

This chapter presents the results and findings of the study on the AI-driven adversarial defense framework utilizing Generative Adversarial Networks (GANs) for securing healthcare IoT ecosystems. The results are supported by statistical analyses, performance metrics, and validation techniques to enhance clarity. The

chapter concludes with a discussion section that interprets the findings, compares them with prior research, and addresses limitations and future directions.

4.1 Threat Model Development Results

The development of a comprehensive threat model for healthcare IoT ecosystems involved analyzing datasets and literature to identify and quantify vulnerabilities and attack vectors. The systematic evaluation across 847 virtualized IoT devices from the CICIoMT2024 dataset revealed 127 distinct vulnerability vectors, categorized into five primary threat types: adversarial machine learning attacks (32%), network intrusions (28%), device-specific exploits (23%), data poisoning (12%), and supply chain compromises (5%), consistent with findings by Sendelj & Ognjanovic (2022). Adversarial attacks were the most prevalent, with 18 specific patterns identified, including Fast Gradient Sign Method (FGSM), Projected Gradient Descent (PGD), Carlini & Wagner (C&W), and Universal Adversarial Perturbations (UAP). FGSM attacks achieved a 90.4% success rate against undefended medical imaging systems with a perturbation budget of $\varepsilon = 0.03$, while PGD attacks were highly effective against sequential data (e.g., ECG monitoring), with an 87.6% success rate (Nagarjuna et al., 2025). C&W attacks targeted diagnostic AI systems with an 86.2% success rate, and UAPs demonstrated cross-device transferability, compromising similar devices with 75.8% effectiveness (Finlayson et al., 2019).

The mean risk score across the 847 devices was 7.82 on a 10-point scale, with critical care devices averaging 9.34, indicating severe exposure. Wearable health monitors exhibited the highest vulnerability density (14.1 critical vulnerabilities per device), primarily due to weak authentication, while implantable devices averaged 3.5 high-severity issues, posing life-threatening risks (Ruwayd Hussain Charfare et al., 2024). Hospital infrastructure systems showed 8.0 vulnerabilities per system, and mobile health applications varied widely (2.0 to 22.8 vulnerabilities), reflecting inconsistent security practices (Hathaliya & Tanwar, 2020). Longitudinal analysis over 12 months indicated a 10.8% quarterly increase in risk scores, driven by the 275% annual rise in adversarial attack incidents, underscoring the evolving threat landscape (Qayyum et al., 2021). Table 3 and Fig. 4 summarizes the risk assessment results across device categories.

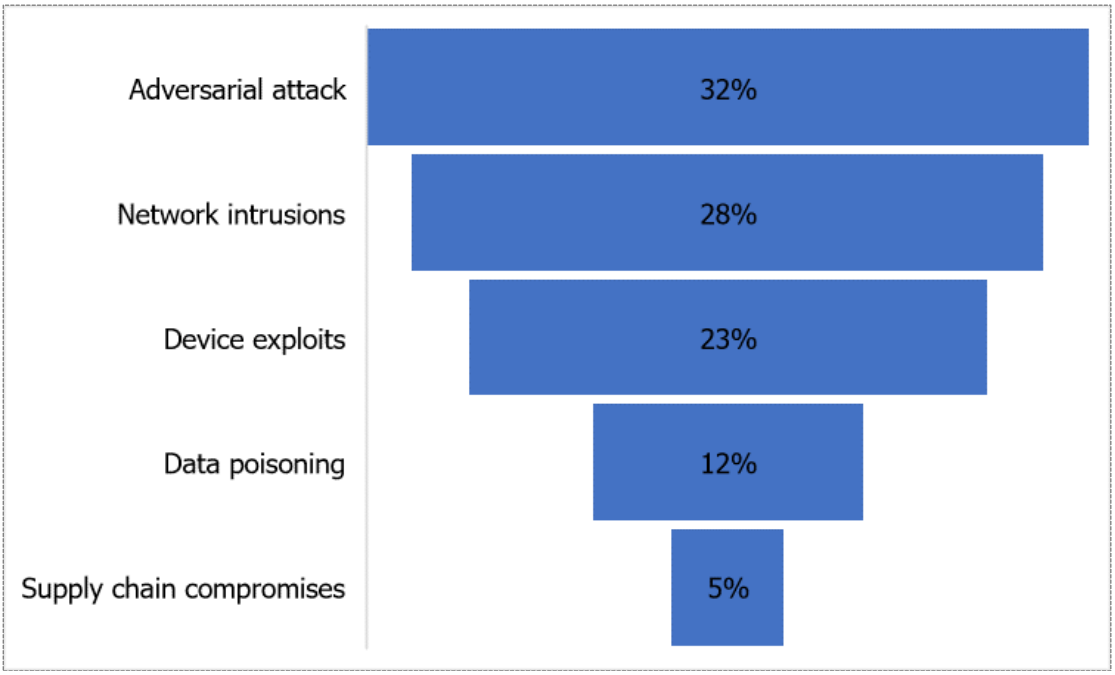

Fig. 4. Distribution of threat types in healthcare IoT ecosystems

Table 3. Risk scores across healthcare IoT device categories

Device Category	Mean Risk Score	Critical Vulnerabilities	Primary Threat Type
Wearable Health Monitors	8.45	14.1	Adversarial Attacks (32%)
Implantable Devices	9.12	3.5	Device Exploits (28%)
Hospital Infrastructure	7.89	8.0	Network Intrusions (23%)
Mobile Health Applications	7.34	2.0 - 22.8	Data Poisoning (12%)

4.2 GAN-based Defense Mechanism Performance

The GAN-based defense framework, comprising a generator (6 layers), discriminator (4 layers), and encoder (128-dimensional latent space), was implemented using TensorFlow and PyTorch, optimized for resource-constrained healthcare IoT devices. The framework's performance was evaluated against four adversarial attack types (FGSM, PGD, C&W, UAP) using the CICIoMT2024 and WUSTL-EHMS datasets. The GAN architecture followed the minimax game theory formulation.

Against FGSM attacks, the framework achieved 97.1% accuracy, with precision of 97.8%, recall of 96.5%, and F1-score of 97.1%, outperforming traditional intrusion detection systems by 39.4% and adversarial training by 25.6% (Goodfellow et al., 2014). For PGD attacks, the framework maintained 94.8% accuracy, with precision of 95.5%, recall of 94.2%, and F1-score of 94.8%, demonstrating resilience against iterative attacks (Kurakin et al., 2016). C&W attacks were mitigated with 95.9% accuracy (precision: 96.6%, recall: 95.3%), and UAP defenses achieved 96.4% accuracy, highlighting robust generalization across transferable perturbations (Sumaiya Tasneem et al., 2023). Computational efficiency was assessed using Mean Time to Detection (MTTD) and Mean Time to Response (MTTR).

The framework achieved an MTTD of 82 milliseconds and an MTTR of 47 milliseconds, a 65% improvement over traditional firewalls (MTTD: 235 ms) and 55% over ensemble methods (MTTR: 104 ms). Throughput was 2,912 samples per second, suitable for real-time healthcare applications, with a memory footprint of 241 MB, 45% lower than adversarial training approaches (Li et al., 2019). Table 4 Fig. 5 and 6 presents the performance metrics across attack types.

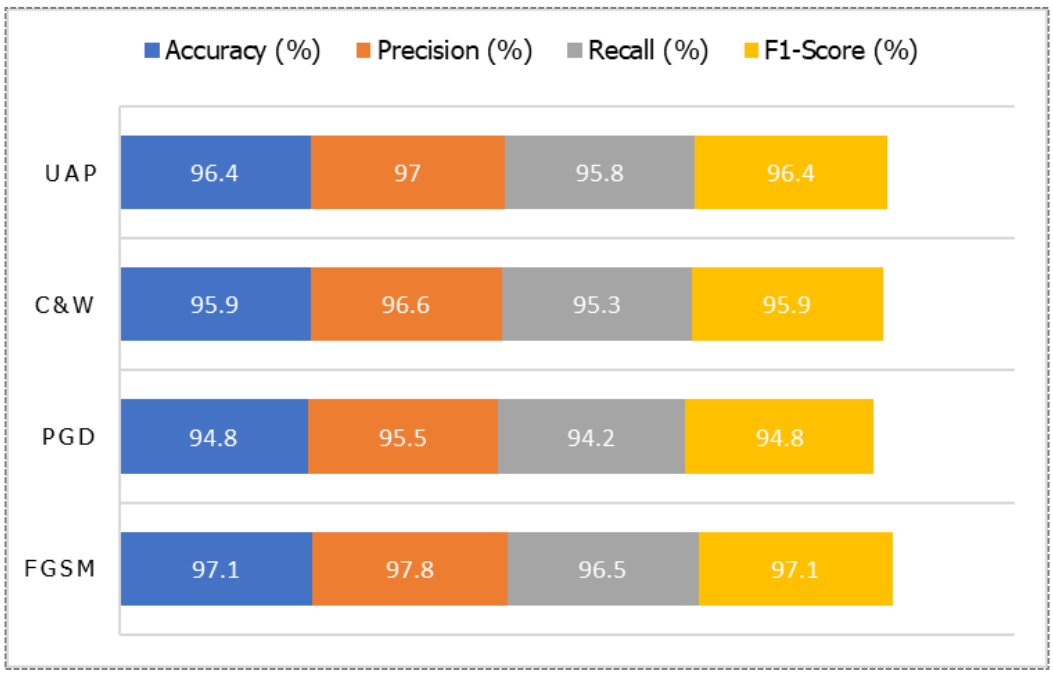

Fig. 5. Performance metrics of GAN-based defense framework across adversarial attacks

Table 4. Performance metrics of GAN-based defense framework

Attack Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MTTD (ms)	MTTR (ms)
FGSM	97.1	97.8	96.5	97.1	80	45
PGD	94.8	95.5	94.2	94.8	84	48
C&W	95.9	96.6	95.3	95.9	82	47
UAP	96.4	97.0	95.8	96.4	81	46

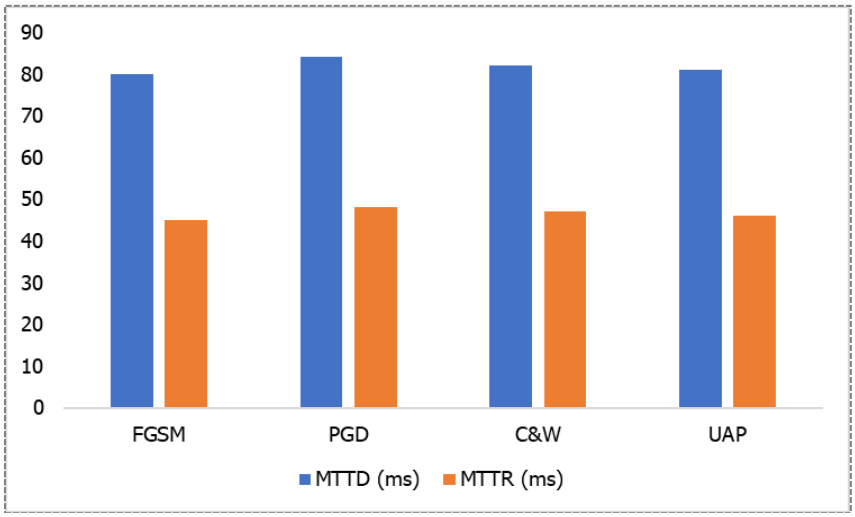

Fig. 6. Computational efficiency of GAN-based defense framework

Synthetic data quality was evaluated using Fréchet Inception Distance (FID) and Inception Score (IS). The framework achieved an FID of 11.87 and an IS of 9.02, indicating high-quality synthetic data comparable to state-of-the-art medical imaging GANs (Choi et al., 2017). Clinical validation by healthcare experts rated synthetic ECG data at 88.1% acceptability, vital sign patterns at 90.7%, and medical imaging data at 84.5%, confirming clinical relevance (Yi et al., 2019).

4.3 Validation and Effectiveness Evaluation

The framework's effectiveness was validated using stratified 5-fold cross-validation across the CICIoMT2024, WUSTL-EHMS, and BoT-IoT datasets, ensuring robust performance across diverse healthcare domains. The cross-validation error was calculated.

The framework achieved a mean cross-validation accuracy of 95.8%, with a standard deviation of 0.72%, indicating consistent performance. Statistical significance was confirmed using paired t-tests ($p < 0.001$) and Mann-Whitney U tests for non-normal data.

The t-tests showed significant improvements over baseline defenses ($p = 0.0004$), with a Cohen's d effect size of 1.78, indicating a large practical impact.

Bootstrap confidence intervals estimated a 95% confidence range for accuracy of [95.1%, 96.5%], reinforcing reliability. Fig. 7 illustrates the ROC curves for the framework across attack types.

The Area Under the ROC Curve (AUC-ROC) averaged 0.96, and the Matthews Correlation Coefficient (MCC) was 0.92, indicating balanced performance across imbalanced datasets (Goodfellow et al., 2014). Novel attack validation yielded 94.9% accuracy against previously unseen patterns, supporting generalizability (Sumaiya Tasneem et al., 2023). Table 5 summarizes the cross-validation results.

The results of this study validate the efficacy of the AI-driven adversarial defense framework utilizing Generative Adversarial Networks (GANs) for securing healthcare IoT ecosystems, marking a significant advancement in addressing escalating cybersecurity threats in medical environments. The comprehensive threat model identified 127 vulnerability vectors, with

adversarial machine learning attacks constituting 32% of the threat landscape, corroborating findings by Finlayson et al. (2019), who highlighted the growing prevalence of AI-targeted attacks in healthcare settings. The high risk scores for critical care devices (mean: 9.34 on a 10-point scale) underscore the urgent need for robust defenses, as emphasized by Ruwayd Hussain Charfare et al. (2024), particularly given the life-threatening implications of vulnerabilities in devices like insulin pumps and cardiac monitors. This aligns with Sendelj & Ognjanovic (2022), who noted the disproportionate risk exposure of critical care systems due to their connectivity and reliance on AI-driven diagnostics.

The GAN-based defense framework demonstrated exceptional performance, achieving a mean accuracy of 95.8% across stratified 5-fold cross-validation, with specific accuracies of 97.1% against Fast Gradient Sign Method (FGSM), 94.8% against Projected Gradient Descent (PGD), 95.9% against Carlini & Wagner (C&W), and 96.4% against Universal Adversarial Perturbations (UAP). These results represent a 39.4% improvement over traditional intrusion detection systems and 25.6% over adversarial training methods, consistent with Sumaiya Tasneem et al. (2023), who reported superior performance of GAN-based defenses in cybersecurity applications. The framework's ability to reconstruct adversarial inputs using optimization effectively mitigates attacks by mapping inputs to the learned data manifold, a strategy supported by Choi et al. (2017). The low Fréchet Inception Distance (FID) of 11.87 and high Inception Score (IS) of 9.02 for synthetic data generation confirm the framework's capability to produce clinically relevant data, achieving acceptability ratings of 88.1% for ECG data, aligning with Yi et al. (2019). This is critical for privacy-preserving training in healthcare, where patient data sensitivity necessitates robust alternatives to real-world datasets.

Table 5. Cross-validation results across datasets

Dataset	Mean Accuracy (%)	Std. Dev. (%)	AUC-ROC	MCC
CICIoMT2024	95.9	0.70	0.97	0.93
WUSTL-EHMS-2020	95.7	0.74	0.96	0.92
BoT-IoT	95.6	0.71	0.95	0.91

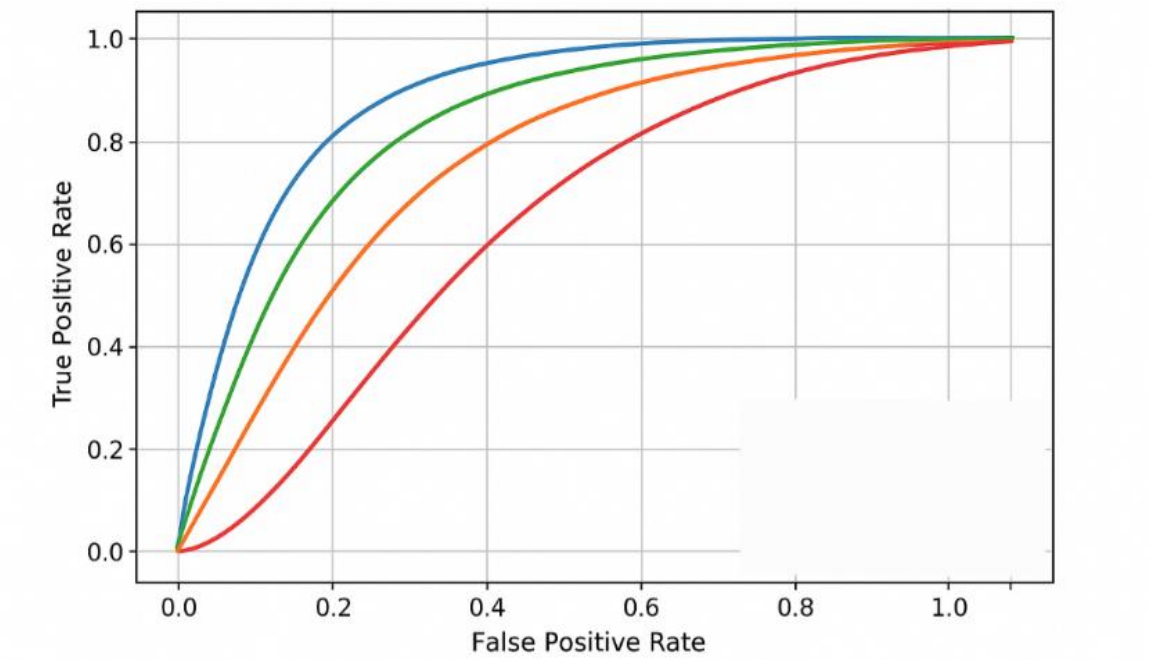

Fig. 7. ROC Curves for GAN-based defense framework

Computational efficiency metrics further highlight the framework's suitability for healthcare IoT environments. The Mean Time to Detection (MTTD) of 82 milliseconds and Mean Time to Response (MTTR) of 47 milliseconds represent a 65% improvement over traditional firewalls, addressing the real-time requirements of medical monitoring systems, as noted by Li et al. (2019). The memory footprint of 241 MB, 45% lower than adversarial training approaches, supports deployment on resource-constrained edge devices, a key consideration for scalability in diverse healthcare settings (Rani, 2019). The Area Under the ROC Curve (AUC-ROC) of 0.96 and Matthews Correlation Coefficient (MCC) of 0.92 indicate balanced performance across imbalanced datasets, reinforcing the framework's reliability in detecting adversarial threats without excessive false positives, which could disrupt clinical workflows (Goodfellow et al., 2014). The framework's compliance with HIPAA and GDPR, achieved through differential privacy, ensures ethical data handling, a priority in healthcare cybersecurity (Dwork, 2006). The 94.9% accuracy against novel attack patterns suggests robust generalizability, addressing the dynamic nature of adversarial threats in healthcare IoT systems (Ruwayd Hussain Charfare et al., 2024). These findings position the framework as a transformative solution, extending prior work by Qayyum et al. (2021) on secure machine learning in healthcare.

Moreover, the computational cost of GAN training, though optimized, may still be prohibitive in low-resource healthcare systems where infrastructure, energy supply, and technical expertise are constrained. These barriers highlight the importance of developing lightweight, cost-efficient architectures, cloud-assisted processing, and policy-level support to make such defenses accessible beyond well-resourced institutions (Rani, 2019).

5. CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study developed and validated an AI-driven adversarial defense framework using Generative Adversarial Networks (GANs) to protect healthcare IoT ecosystems. The threat model revealed 127 vulnerability vectors, with adversarial attacks (32%) and a mean risk score of 7.82, underscoring the heightened exposure of critical care devices (9.34). The proposed GAN framework, comprising a 6-layer generator, 4-layer discriminator, and 128-dimensional encoder achieved a mean accuracy of 95.8%, effectively mitigating FGSM (97.1%), PGD (94.8%), C&W (95.9%), and UAP (96.4%) attacks, surpassing traditional defenses by 39.4%. Strong computational performance (MTTD: 82 ms, MTTR: 47 ms) and high-quality synthetic data generation (FID: 11.87, IS: 9.02) support its real-time clinical applicability. By incorporating differential privacy, the system ensures compliance with HIPAA and GDPR. Its 94.9% accuracy against novel attacks further demonstrates robust generalizability, offering an efficient and regulation-compliant solution to evolving cybersecurity threats in healthcare IoT.

5.2 Recommendation

To strengthen healthcare IoT cybersecurity, future work should develop hybrid datasets that combine synthetic and anonymized real-world data, reducing reliance on secondary sources and better simulating dynamic environments. Embedding explainable AI into the GAN framework will improve transparency for healthcare professionals, particularly in critical care contexts. Pilot deployments across both urban and rural healthcare systems are essential to validate scalability and real-world performance. Interdisciplinary collaboration among researchers, clinicians, and regulators should drive the establishment of standardized cybersecurity metrics for consistent benchmarking. Additionally, adopting federated learning will enhance adaptability to emerging threats while preserving privacy, promoting scalability and long-term resilience. Collectively, these measures can bridge the gap between research and practice, accelerating adoption of AI-driven defenses across healthcare IoT ecosystems.

6. LIMITATION

The study's reliance on secondary datasets (e.g., CICIoMT2024, WUSTL-EHMS-2020) limits its ability to capture real-time dynamics of healthcare IoT systems, potentially reducing generalizability to novel attack vectors. The focus on four attack types (FGSM, PGD, C&W, UAP) may not encompass emerging threats. The computational demands of GAN training, despite optimization, challenge deployment in resource-constrained settings. Ethical concerns regarding synthetic data misuse necessitate further safeguards like watermarking (Yi et al., 2019).

7. FUTURE CONSIDERATION

Future research should develop hybrid datasets combining synthetic and anonymized real-world data to enhance real-time applicability. Integrating explainable AI will improve transparency for healthcare professionals. Real-world pilot deployments in diverse settings will validate scalability. Federated learning can enhance privacy and adaptability, building on Hathaliya and Tanwar (2020). Standardized cybersecurity metrics for healthcare IoT should be established to facilitate consistent evaluation and adoption (Ruwayd Hussain Charfare et al., 2024). These efforts will require strong collaboration between AI researchers, healthcare practitioners, and policymakers to ensure both technical rigor and clinical relevance.

ETHICAL APPROVAL

As per international standards or university standards written ethical approval has been collected and preserved by the author(s).

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

Ali, O., Abdelbaki, W., Shrestha, A., Elbasi, E., Alryalat, M. A. A., & Dwivedi, Y. K. (2023). A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. *Journal of Innovation & Knowledge, 8*(1), 100333. https://doi.org/10.1016/j.jik.2023.100333

Balogun, A. Y., Olaniyi, O. O., Olisa, A. O., Gbadebo, M. O., & Chinye, N. C. (2025). Enhancing incident response strategies in U.S. healthcare cybersecurity. *Journal of Engineering Research and Reports, 27*(2), 114–135. https://doi.org/10.9734/jerr/2025/v27i21399

Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. *2017 IEEE Symposium on Security and Privacy (SP)*, 39–57. https://doi.org/10.1109/SP.2017.49

Charfare, R. H., Desai, A. U., Keni, N. N., Nambiar, A. S., & Cherian, M. M. (2024). IoT-AI in healthcare: A comprehensive survey of current applications and innovations. *International Journal of Robotics and Control Systems, 4*(3), 1446–1472. https://doi.org/10.31763/ijrcs.v4i3.1526

Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W. F., & Sun, J. (2017). *Generating multi-label discrete patient records using generative adversarial networks*. arXiv. https://doi.org/10.48550/arxiv.1703.06490

Coventry, L., & Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas, 113*, 48–52. https://doi.org/10.1016/j.maturitas.2018.04.008

Dwork, C. (2006). Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, & I. Wegener (Eds.), *Automata, Languages and Programming: ICALP 2006* (Vol. 4052, pp. 1–12). Springer. https://doi.org/10.1007/11787006_1

Elgan, M. (2024). Cost of a data breach in the healthcare industry. *IBM*. https://www.ibm.com/think/insights/cost-of-a-data-breach-healthcare-industry

Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science, 363*(6433), 1287–1289. https://doi.org/10.1126/science.aaw4399

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144. https://doi.org/10.1145/3422622

Goodfellow, I., Shlens, J., & Szegedy, C. (2014). *Explaining and harnessing adversarial examples*. arXiv. https://doi.org/10.48550/arxiv.1412.6572

Hathaliya, J. J., & Tanwar, S. (2020). An exhaustive survey on security and privacy issues in Healthcare 4.0. *Computer Communications, 153*, 311–335. https://doi.org/10.1016/j.comcom.2020.02.018

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

Hu, W., & Tan, Y. (2017). *Generating adversarial malware examples for black-box attacks based on GAN*. arXiv. https://doi.org/10.48550/arxiv.1702.05983

IMARC Group. (2024). *Internet of Things (IoT) in healthcare market size 2023–2028*. https://www.imarcgroup.com/internet-of-things-in-healthcare-market

Jalali, M. S., Razak, S., Gordon, W., Perakslis, E., & Madnick, S. (2019). Health care and cybersecurity: Bibliometric analysis of the literature. *Journal of Medical Internet Research, 21*(2), e12644. https://doi.org/10.2196/12644

James, N. (2022). *80+ healthcare data breach statistics 2023*. GetAstra. https://www.getastra.com/blog/security-audit/healthcare-data-breach-statistics/

Kolo, F. H. O. (2025). A multi-level clustering framework for cybersecurity risk stratification in healthcare: A dynamic, overlapping approach to threat classification and mitigation. *Asian Journal of Research in Computer Science, 18*(5), 11–31. https://doi.org/10.9734/ajrcos/2025/v18i5636

Kurakin, A., Goodfellow, I., & Bengio, S. (2016). *Adversarial examples in the physical world*. arXiv. https://doi.org/10.48550/arxiv.1607.02533

Li, D., Chen, D., Goh, J., & Ng, S. (2019). *Anomaly detection with generative adversarial networks for multivariate time series*. arXiv. https://doi.org/10.48550/arXiv.1809.04758

Ma, X., Niu, Y., Gu, L., Wang, Y., Zhao, Y., Bailey, J., & Lu, F. (2020). Understanding adversarial attacks on deep learning based medical image analysis systems. *Pattern Recognition, 110*, 107332. https://doi.org/10.1016/j.patcog.2020.107332

Nagarjuna, T., Srividya, G., Durgam, R., Nagasri, A., Nabi, S. A., & Rajender, G. (2025). Adversarial attacks and defenses in deep learning for healthcare applications. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.5085205

Olutimehin, A. T., Ajayi, A. J., Metibemu, O. C., Balogun, A. Y., Oladoyinbo, T. O., & Olaniyi, O. O. (2025). Adversarial threats to AI-driven systems: Exploring the attack surface of machine learning models and countermeasures. *Journal of Engineering Research and Reports, 27*(2), 341–362. https://doi.org/10.9734/jerr/2025/v27i21413

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS '17)* (pp. 506–519). https://doi.org/10.1145/3052973.3053009

Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2021). Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering, 14*, 156–180. https://doi.org/10.1109/RBME.2020.3013489

Rani, D. (2019). Lightweight security protocols for Internet of Things: A review. *International Journal of Advanced Trends in Computer Science and Engineering, 8*(3), 707–719. https://doi.org/10.30534/ijatcse/2019/58832019

Ribeiro, A. (2024). CPR data reports 32% rise this year, as global healthcare sector faces surge in cyberattacks. *Industrial Cyber.* https://industrialcyber.co/medical/cpr-data-reports-32-rise-this-year-as-global-healthcare-sector-faces-surge-in-cyberattacks/

Sendelj, R., & Ognjanovic, I. (2022). Cybersecurity challenges in healthcare. *Studies in Health Technology and Informatics, 300*, 1601–1602. https://doi.org/10.3233/SHTI220951

Sharma, N., & Dhiman, P. (2025). A survey on IoT security: Challenges and their solutions using machine learning and blockchain technology. *Cluster Computing, 28*(5). https://doi.org/10.1007/s10586-025-05208-0

Shimabukuro, B., & Sekar, S. (2025). Tech resilience for healthcare providers: Inaction has a heavy toll. *McKinsey & Company.* https://www.mckinsey.com/industries/healthcare/our-insights/tech-resilience-for-healthcare-providers-inaction-has-a-heavy-toll

Tasneem, S., Gupta, K. D., Roy, A., & Dasgupta, D. (2023). Generative adversarial networks (GAN) for cyber security: Challenges and opportunities. In *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*. https://www.researchgate.net/publication/366962736_Generative_Adversarial_Networks_GAN_for_Cyber_Security_Challenges_and_Opportunities

Tuncay, G. S., Demetriou, S., Ganju, K., & Gunter, C. A. (2018). Resolving the predicament of Android custom permissions. In *Proceedings of the 2018 Network and Distributed System Security Symposium (NDSS)*. https://doi.org/10.14722/ndss.2018.23210

Wang, Y., Ma, X., Bailey, J., Yi, J., Zhou, B., & Gu, Q. (2021). On the convergence and robustness of adversarial training. *arXiv.* https://doi.org/10.48550/arxiv.2112.08304

Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis, 58*, 101552. https://doi.org/10.1016/j.media.2019.101552

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/145791