International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 10 Issue 1, Jan-Feb 2026 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

A Comparative Evaluation of GANs, Vaes, and Diffusion Models for Early Anomaly Detection in Medical Device Performance Data

Muhammad Faheem[1], Aqib Iqbal[2]

[1]IT Management, Cumberland University, USA
[2]Project Management, University of Law, USA

ABSTRACT

Reliability, safety, and sustained performance of the medical devices are highly critical issues in the contemporary healthcare system where any failure in the medical devices may have serious clinical outcomes and patient injuries. With the growing complexity and data being of a medical system, there has been the emergence of early anomaly detection in device performance data as a crucial process of preventing failure, facilitating proactive maintenance, and improving patient safety. Nevertheless, the current solutions mostly depend on conventional statistical or supervised learning schemes, which are sensitive to data imbalance, changing faults modes, and small amounts of labeled anomaly samples.

Recent developments in generative deep learning, especially Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Models, provide encouraging opportunities to support unsupervised and semi-supervised anomaly detection through learning the latent data distribution and detecting anomalies with respect to normal operation. Although these models continue to be adopted, they have not been systematically and comparatively reviewed in med device performance monitoring, and it is still unclear how they relate in terms of effectiveness, strength, and their ability to be deployed in safety-critical settings.

This paper is a comparative critical analysis of GAN-, VAE-, and Diffusion-based architectures of early anomaly detection in medical devices performance data. To evaluate the performance of a model, we would evaluate it based on several quantitative measures such as detection accuracy, precision-recall balance, false alarm rates, detection latency and computational efficiency. The study also looks at the robustness issues in the noisy environment, scalability to high-dimensional telemetry data and concept drift resiliency, which informs about real-world applicability.

The findings show that, although GANs are more efficient in restoring intricate normal samples, VAEs can offer stable latent features that can be used to score anomalies consistently, and Diffusion Models have better resilience to detect subtle and changing abnormalities. The results of these studies point to trade-offs associated with models and provide practical recommendations in the process of selecting an adequate generative frameworks in the medical safety context. Finally, this piece of work helps to develop reliable, evidence-based monitoring systems that facilitate the detection of faults early, adherence to regulations, and enhancing better clinical safety outcomes.

How to cite this paper: Muhammad Faheem | Aqib Iqbal "A Comparative Evaluation of GANs, Vaes, and Diffusion Models for Early Anomaly Detection in Medical Device Performance Data" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-10 | Issue-1, February 2026, pp.437-447, URL: www.ijtsrd.com/papers/ijtsrd100089.pdf

IJTSRD100089

KEYWORDS: *Generative Models for Anomaly Detection in Medical Devices, GAN-Based Anomaly Detection in Healthcare Systems, Variational Autoencoders for Medical Time-Series Monitoring, Diffusion Models for Unsupervised Anomaly Detection, Early Fault Detection in Medical Device Performance Data, Unsupervised Learning for Safety-Critical Healthcare Systems, Reliability and Predictive Maintenance of Medical Cyber-Physical Systems.*

1. INTRODUCTION

Statistical and communication, smart sensors are being integrated into medical devices and are turning the present healthcare into a data-based, cyber-physical ecosystem. These systems produce unceasing flows of high-dimensional operational and physiological information, which is vital in promoting trustworthiness, patient safety, and regulatory correctness. Yet, performance anomalies which are difficult to distinguish, such as sensor drift or calibration, latent hardware degradation, and similar ones, might result in the wrong clinical decision, system malfunction, or a negative patient outcome. Due to the increase towards automation and interconnectivity in healthcare systems, early and dependable detection of anomalies has become a foundational need of the safety and resiliency of medical devices (Bose et al., 2025; Desai et al., 2025; Kadiyala et al., 2025).

Conventional rule-based monitoring systems and threshold based alerting systems find it difficult to handle the complexity, and nonlinearity, and variability of medical device performance data. These constraints are even exacerbated by edge-enabled and cloud-integrated healthcare systems where devices do not only have heterogeneous workloads but also dynamic working conditions (Jadhav and Kulkarni, 2025; Loganathan et al., 2025). Therefore there has been increased popularity of data-driven and unsupervised learning methods, which can directly model normal system behavior using data without the exhaustive fault labeling that would otherwise be a very impractical requirement in a safety-critical medical setting.

Out of these methods, generative models have shown an outstanding potential of anomaly detection, by training on the underlying data distribution of normal operational states and detecting deviations via reconstruction errors, likelihood estimation, or latent space anomalies. The GANs have seen extensive use in the industrial and medical anomaly detection because they can produce plausible samples and emphasize the slightest normal deviation (Park et al., 2022; Striuk and Kondratenko, 2023). Newer augmentations of GAN-based systems have been demonstrated to be useful in IoT security in healthcare and real-time body monitoring especially when combined with hybrid deep learning systems (Loganathan et al., 2025; Utsha and Morshed, 2025).

Variational Autoencoders (VAEs), however, provide probabilistic latent representations, which are easy to use to model uncertainty and use to score anomalies based on reconstruction. They are particularly appropriate to the medical time-series and multimodal data where interpretability and stability are paramount because of the structure of their latent spaces (Sadanandan and Behzadan, 2025). The broader medical use of AI has also been applied using VAEs, such as decision support systems and precision healthcare modeling, which notes that it is versatile to clinical settings (Sadanandan et al., 2025; Udoy and Hassan, 2025).

Most recently, diffusion models have come to be seen as a formidable alternative to GANs and VAEs, reaching the state-of-the-art in unsupervised anomaly detection in medical imaging and multimodal healthcare data. Diffusion models can learn to reverse a noise-corruption process sequentially to produce highly accurate reconstruction as well as localize anomalies at fine-grained scales. The enhanced diffusion-based models such as conditioned, patched and adaptive diffusion models have been shown to be more robust to detect minor abnormalities in brain MRI and preoperative imaging data (Behrendt et al., 2023; Behrendt et al., 2025; Wang et al., 2025; Yao et al., 2025). Such attributes render diffusion models especially appealing to early anomaly detection in medical equipment performance data, in which anomalies can be insidious, low-frequency, and conditional.

Although the literature on generating models to detect healthcare anomalies has been increasing, a systematic and comparative review of the GANs, VAEs, and diffusion models have not been broadly applied in terms of detecting medical device malfunctions in their initial stages. Current literature is seen to be dedicated to a single model category, a particular data modalities problem, or a wider healthcare monitoring setup without considering device-typical reliability and failure antecedents (Bose et al., 2025; Desai et al., 2025; Jadhav and Kulkarni, 2025). Although recent comparative studies have started to investigate the use of generative models to detect early failures, there is still the need to gain more analytical understanding of model behavior, sensitivity to detection, computational trade-off, and their use in safety-critical deployment settings (Sadanandan et al., 2025).

This gap is filled in this paper, which offers a detailed comparative analysis of GANs, VAEs, and diffusion models in the context of early anomaly detection of medical devices based on the performance data. It is a systematic study of the accuracy of detection, its resistance to noise, interpretability and requirements of deployment across generative paradigms, focusing on safety important applications in healthcare. This paper will offer instructions on the implementation of generative models in future medical device monitoring systems by integrating the knowledge of the latest findings in diffusion-based anomaly detection, hybrid deep learning systems, and healthcare cybersecurity systems (Behrendt et al., 2025; Loganathan et al., 2025; Sadanandan et al., 2025).

2. Related Work

2.1. The traditional anomaly detection of medical devices

Anomaly detection methods in early stages of medical devices were mostly based on rule based systems, statistical thresholds and signal processing heuristics based on domain experience. Although those techniques are computationally efficient and simple to implement, they do not easily apply to different operating conditions and device types. Non-stationary, high-dimensional data streams produced by medical devices in complex healthcare settings change over time as a result of variations in patients and the environment and hardware degradation. These properties are a big impediment to the use of static thresholds and custom features, they may either lead to large false-positive rate or slow fault detection (Bose et al., 2025; Desai et al., 2025).

Moreover, the conventional methods are not flexible enough to be used in contemporary cloud-connected and edge-enabled healthcare systems, in which the devices communicate in real-time with the rest of the cyber-physical elements. Polls on edge and healthcare network anomaly detection underscore the fact that the traditional approaches are inadequate in emulating latent failure patterns, which precondition the occurrence of critical device failures (Jadhav and Kulkarni, 2025). Such constraints have spurred the transition to data-driven deep learning algorithms that are able to learn complicated operating patterns directly on raw inputs.

2.2. Generative Approaches and Deep Learning.

Deep learning has become a leading trend of modeling nonlinear dependencies and high-dimensional feature space, making it a dominant approach in detecting anomalies in healthcare. Managed deep learning models though demand huge amount of labeled anomaly data which are in scarcity in medical device since failure is infrequent and morally difficult to cause. Because of this, unsupervised and semi-supervised generative models have become popular, since they are trained on normal behavior distributions and identify deviations without any fault labels (Sadanandan & Behzadan, 2025).

Medical device performance monitoring This is a task that generative models are especially good at since generative models allow reconstruction-based, likelihood-based, or latent-space anomaly scoring. These properties enable to identify small deviations at an early stage, which can refer to some incipient faults instead of actual failures. New medical imaging, physiological signal analysis, and healthcare IoT systems tend to implement generative frameworks more and more frequently, as recent medical-related research points to the role of generative frameworks in healthcare-related problems (Udoy & Hassan, 2025; Bose et al., 2025).

2.3. GAN-Based Anomaly Detection

Generative Adversarial Networks (GANs) have been actively studied with regards to detection of anomalies because it has the capacity to produce realistic images of normal data distributions. The GAN-based methods tend to detect anomalies with the help of reconstruction error or the level of confidence of the discriminators in the situation when they observe unfamiliar patterns. The GAN variants like AnoGAN and its variations have proved to be useful in identifying anomalies in the complex data of signals and processes in healthcare and industrial monitoring (Park et al., 2022).

The recent work has applied GAN-based anomaly detection to cross-domain and reconfigurable environments and enhanced their ability to be robust and adaptable with hybrid architecture and fuzzy elements (Striuk and Kondratenko, 2023). GANs have been combined with CNNs and LSTMs in the medical systems and healthcare IoT to increase detection accuracy and live responsiveness (Loganathan et al., 2025; Utsha and Morshed, 2025). Nevertheless, GANs are noted to be unstable when training and subject to mode collapse, as well as low interpretability, which are problematic in medical device safety-critical applications.

2.4. VAE-Based Anomaly Detection

Variational Autoencoders (VAEs) are a probabilistic variant of GAN, which trains to learn learned structured latent representations of normal data distributions. VAEs provide principled uncertainty estimation and hence are appealing in healthcare settings, where confidence-minded decision-making is necessary. The anomalies are normally detected by reconstruction loss or deviation of probability distributions in latent spaces.

In the medical and healthcare-oriented context, VAEs have been depicted to provide consistent training behavior and sound performance in noisy and heterogeneous data (Sadanandan and Behzadan, 2025). They can be used not only in anomaly detection but also in more general decision support systems in the domains of precision medicine and medical devices monitoring (Sadanandan et al., 2025). Nevertheless, these benefits come at a cost: VAEs can have unnaturally smooth reconstructions, and thus, they will blur fine-grained anomalies, especially those that occur in the initial stages of device degradation.

2.5. Time-Series and Healthcare Diffusion Models.

Diffusion models are a more recent category of generative methods that have recently obtained state of the art in unsupervised anomaly detection. Diffusion models can model a noisy injection and denoising procedure to achieve excellent outcomes in capturing complex distributions and high-fidelity signal generation. Diffusion-based methods have been shown to be better in terms of localizing anomalies and robustness than GANs and VAEs in medical imaging (Behrendt et al., 2023; Behrendt et al., 2025).

More enhanced diffusion models like adaptive, global-local and perception-aware diffusion models have additionally enhanced the reconstruction accuracy and anomaly sensitivity on healthcare data (Wang et al., 2025; Yao et al., 2025). Although the majority of diffusion-based works target imaging modalities, more recent research proposes that they have a high potential to be applied in time-series as well as multimodal medical data, and hence are viable alternatives to detect early anomalies in monitoring the performance of medical devices.

2.6. Comparative Gap Analysis

Despite the strengths that GANs, VAEs, and diffusion models have demonstrated in healthcare anomaly detection, the literature on this topic presents a large amount of analysis of these models either alone or in a domain-specific context, such as medical imaging or IoT security. The comparative analysis of various generative paradigms, in particular, early anomaly detection in the context of medical devices performance data is underrepresented (Sadanandan et al., 2025). The evaluation includes a dearth of systematic evaluation that tackles sensibility of detection, ability to withstand noise, comprehensibility and the ability to be deployed under healthcare constraints.

To close this gap, this paper gives a combined comparative analysis of GANs, VAEs and diffusion models, highlighting their applicability in detection of anomalies at early stages of operation in safety-critical medical device settings.

Table 1: Summary of Generative Model-Based Anomaly Detection Approaches in Healthcare

Model Type	Key Strengths	Limitations	Representative Studies
GANs	High-fidelity generation; strong anomaly discrimination	Training instability; limited interpretability	Park et al. (2022); Striuk & Kondratenko (2023); Loganathan et al. (2025)
VAEs	Stable training; probabilistic uncertainty modeling	Over-smoothing of subtle anomalies	Sadanandan & Behzadan (2025); Sadanandan et al. (2025)
Diffusion Models	Superior reconstruction quality; robust anomaly localization	Higher computational cost; limited time-series studies	Behrendt et al. (2023, 2025); Wang et al. (2025); Yao et al. (2025)

3. Description of the data and statement of the problem.

3.1. The descriptions below are based on the characteristics of the data of the medical device performance.

The performance data of medical devices is non-stationary multivariate and temporal continuous measures of sensors, control units, communication interfaces, and embedded software parts. These stream of data records physiological interactions and device focused operation states like signal integrity, latency, power usage, calibration stability and fault indicators. As mentioned in the cloud-integrated and edge-enabled healthcare systems, these data are highly variated as a result of patient heterogeneity, environmental factors, and aging of devices (Bose et al., 2025; Kadiyala et al., 2025).

In contrast to clinical outcome data, a performance data does not necessarily have clear failure labels, especially when it comes to early anomalies, which are precursors of severe malfunctions. Unsupervised and self-supervised learning paradigm is of particular interest in detecting the presence of anomalies in medical devices (Sadanandan and Behzadan, 2025; Jadhav and Kulkarni, 2025).

3.2. Data Sources: Real, Simulated and Hybrid Data.

Since the probability of failure events is very low and due to regulations in healthcare settings, the study of anomaly detection often uses hybrid datasets of real-life operating data and artificially created anomalies. The real data can be provided by healthcare IoT systems, edge-based monitoring systems, or medical infrastructures that are connected to the cloud (Bose et al., 2025; Desai et al., 2025). To simulate sensor faults, communication failures, or a slow decline in performance, simulated anomalies are usually injected and allow the evaluation of early detection limits to be controlled.

No matter the healthcare anomaly detection framework, it is demonstrated that hybrid data management can enhance robustness and overallizability when analyzing generative models that need enough coverage of normal behavior distributions (Loganathan et al., 2025; Sadanandan et al., 2025).

3.3. Preprocessing of Data and Normalization.

Preprocessing is also important in the creation of consistent training and comparability of generative models. Raw medical device data usually have missing values, noise, scale differences between sensor channels. Coulderary pipelines Standard pipelines used in preprocessing can be signal denoising, temporal alignment, normalization, and windowing to conserve temporal dependencies at the expense of computational complexity.

The common normalization methods utilized to overcome scale sensitivity among multivariate inputs include z-score scaling, or min-max normalization (Udoy & Hassan, 2025). The steps considered especially crucial to diffusion models that involve noise perturbation and reconstruction processes that are sensitive to input distributions (Behrendt et al., 2023; Yao et al., 2025).

3.4. The structure of the time series can be analyzed using Temporal and Multivariate Structure.

The best data representation of medical device performance is a multivariate time series, with one time step comprising many correlated variables. Temporal dependencies are used to store both short-term variations in characteristics, as well as long-term degradation patterns, which tourist spots typically require to identify anomalies early. Generation models should hence be able to learn shared spatial-temporal representations instead of observations being viewed as independent samples.

The current research indicates that failing to take into account the temporal structure can introduce prominent losses to the detection performance of anomalies in healthcare systems, especially at the initial deviation stage (Utsha and Morshed, 2025; Loganathan et al., 2025). This prompts sequence-aware architecture and time reconstruction goals on GANs, VAEs, and diffusion-based designs.

3.5. The formal definition of anomalies is given below.

An anomaly in medical device context is a variation in the learned distribution of normal operational behavior that could be indicative of incipient faults, degradation and abnormal interaction with the clinical environment. In contrast to catastrophic failures, initial anomalies are not always very noticeable but gradually develop with time.

Formally, let $X = \{x_t \in \mathbb{R}^d\}_{t=1}^{T}$ denote a multivariate time series of device performance measurements. An

observation x_t is considered anomalous if it lies outside the high-probability region of the learned normal data distribution p(x) as estimated by a generative model (Park et al., 2022; Behrendt et al., 2025).

3.6. Problem Formulation

The anomaly detection problem is formulated as an unsupervised learning task where a generative model G_θ is trained on normal operational data to approximate the underlying data distribution $p_{normal}^{(X)}$. During inference, anomaly scores are computed using reconstruction error, likelihood estimation, or latent-space deviations, depending on the model class.

For GANs and VAEs, anomalies are typically identified via reconstruction loss:

$$\mathcal{A}(x_t) - \| x_t - \hat{x}_t \|$$

Where x_t is the reconstructed signal For diffusion models, anomaly scores may be derived from denoising residuals or reconstruction likelihood across multiple diffusion steps (Behrendt et al., 2023; Wang et al., 2025).

The objective of this study is to comparatively evaluate GANs, VAEs, and diffusion models under a unified problem formulation, assessing their effectiveness for early anomaly detection in medical device performance data.

4. Generative Model Architectures and Methodology

4.1. Overall Experimental Framework

The experimental system is meant to facilitate a comparative study (controlled and fair) of GANs, VAEs and diffusion models in early anomaly detection of medical devices performance data. Modeling is done only on normal operation data to get a feel of the underlying distribution of normal device behavior. During inference, the anomalies are identified by measuring differences between this learned distributions by the use of model-specific anomaly scoring.

The pipeline is a unified one that comprises of data preprocessing, model training, scoring of anomalies, and evaluation. This single design is such that variation of performance is due to modeling abilities and not differences in data handling. Frankly speaking, comparable experimental isolation procedures have been highlighted in the recent comparative research on generative model approaches to medical device failure detection (Sadanandan et al., 2025).

4.2. Architecture and Training The architecture and training of GANs.

Anomaly detecting models based on GAN include a generator-discriminator network that is trained in adversarial mode. The generator is trained to recreate real samples of typical functioning data of medical devices, whereas the discriminator differentiates between real and generated samples. Inference and dissimilarities during the process of inference are determined on the basis of reconstruction error or discriminator-based confidence scores (Park et al., 2022).

A generator with convolutional and recurrent layers is used in order to extract both the spatial correlations and the temporal dependencies that allow dealing with the time-varying and multivariate characteristics of medical device data. This architecture is inspired by architectures in biosignal based systems, including EEG based deep learning pipelines, where subtle temporal changes are vital to downstream system behavior (Kachhia et al., 2020; Kachhia and George, 2021). Hybrid GAN models that include convolutional and sequential modeling have also been shown to be effective in health care IoT anomaly detection scenarios (Loganathan et al., 2025).

Although they are powerful since expressive, GANs are prone to training instabilities and mode collapse. These restrictions may inhibit their capability to spot anomalies at low stages that are near the regular data procession, which has been raised a variety of times in the research on cross-domain anomaly detection (Striuk & Kondratenko, 2023).

4.3. VAE Architecture and Latent Modeling.

VAEs represent medical device performance data by the encoder-decoder architecture which transforms inputs into a probabilistic latent space. The encoder becomes trained on the parameters of a latent distribution and the decoder is trained on the sample latent variables. Reconstruction loss or latent likelihood deviation is used to detect anomalies.

It is important to note that the probabilistic formulation of VAEs includes implicit uncertainty estimation, which is especially helpful in the domain of safety-important medical systems. This property helps to detect early as the ambiguous or low-confidence reconstructions are called to attention and out-of-band failures are not yet made in the case of an overt failure (Sadanandan and Behzadan, 2025). VAEs have found extensive applications in the medical analytics and precision medicine workflows, where generative fidelity is not emphasized on the stability or interpretability (Udoy & Hassan, 2025).

Nonetheless, VAEs tend to generate over-smooth reconstructions, which can be used to obscure features of anomalies present in the initial stages of degradation. This is a major drawback in the stability versus sensitivity trade-off between their comparative assessment with GANs and diffusion models (Sadanandan et al., 2025).

4.4. Conditional Sampling Process and Formulation of the Diffusion Model.

Diffusion models use a process of adding noise to their training data and learning a denoising denoising process. In contrast to GANs and VAEs, diffusion models do not use adversarial training or explicit inference of the latent variable, leading to the fact that they have better training stability and reconstruction quality.

The recent developments on the diffusion-based anomaly detection have proven to be better performers in the healthcare environment, especially in the recognition of small yet significant structural aberrations in medical imaging and multimodal data (Behrendt et al., 2023; Behrendt et al., 2025). The formulations of adaptive and global-local diffusion further make the model more sensitive by reflecting on the fine-grained and the global pattern of anomalies (Wang et al., 2025; Yao et al., 2025).

Even though the majority of diffusion-based analyses are dedicated to the image, their iterative reconstruction processes are applicable to multivariate medical device time-series data, where a gradual deviation is typical. This renders the diffusion models good options in detecting anomalies in early stages in device performance monitoring.

4.5. Anomaly Scoring Mechanisms

Model-specific criteria are used to develop the anomaly scores. In the case of GANs and VAEs, the reconstruction error is the most popular:

$$\mathcal{A}(x) = \| x - \hat{x} \|$$

Where $\hat{x}$ denotes the reconstructed input. Additional discriminator-based scores may be used for GANs to enhance sensitivity (Park et al., 2022).

For diffusion models, anomaly scores are computed using denoising residuals or reconstruction likelihood aggregated across diffusion steps. These multi-step scores provide robust detection of subtle deviations that may not be apparent in single-pass reconstructions (Behrendt et al., 2025; Wang et al., 2025).

Early anomaly detection is particularly important in safety-critical pipelines, as it enables downstream mitigation strategies such as automated alerts or self-healing mechanisms. Prior work on proactive system reliability and self-healing architectures underscores the value of accurate early anomaly signals for triggering corrective actions (Tewari et al., 2025a; Tewari et al., 2025b)

4.6. Computational Complexity Considerations

Computational efficiency is a critical factor for deploying anomaly detection models in real-world medical device ecosystems. GANs and VAEs generally offer lower inference latency, making them suitable for edge-based or resource-constrained environments (Bose et al., 2025). Diffusion models, while computationally more demanding due to iterative sampling, provide higher detection fidelity and robustness.

In large-scale healthcare infrastructures, where reliability and safety outweigh strict latency constraints, diffusion models may be preferable. Conversely, hybrid architectures that combine fast detection with downstream self-healing or mitigation pipelines can balance performance and efficiency (Tewari et al., 2025).

Table 2: Comparison of Generative Model Architectures for Early Anomaly Detection in Medical Device Data

Model	Strengths	Limitations	Computational Cost	Key Supporting Studies
GAN	High-fidelity generation; strong discrimination	Training instability; mode collapse	Moderate	Park et al. (2022); Loganathan et al. (2025); Striuk & Kondratenko (2023)
VAE	Stable training; uncertainty modeling	Over-smoothed reconstructions	Low–Moderate	Sadanandan & Behzadan (2025); Udoy & Hassan (2025)
Diffusion	Superior reconstruction; robust early anomaly detection	High inference cost	High	Behrendt et al. (2023, 2025); Wang et al. (2025); Yao et al. (2025)

5. Experimental Implementation and Assessment of the results.

5.1. Modeling Environment and Hardware.

Experiments were carried out in a controlled environment of the cloud-based research to achieve reproducibility and scalability. High-performance computer system training and evaluation Model training and evaluation of high-performance computer systems based on multi-core CPUs and NVIDIA GPU accelerators were used to meet the computational requirements of deep generative models, especially diffusion-based models. To run the experimental pipeline the popular deep learning frameworks were used allowing the training, logging and evaluation of the models to be standard.

In order to provide fairness, the same data splits, preprocessing pipelines and optimization settings were used across all the evaluated models. Hyperparameter selection was done on a validation subset and ultimate performance measures were reported on a held out test dataset containing normal operation behaviour and injected anomaly patterns of actual medical device failures.

5.2. Evaluation Metrics

Finding anomalies was evaluated using a set of metrics that are widely used in machine learning work related to safety-critical and healthcare:

➢ **Area Under the Receiver Operating Characteristic Curve (AUC-ROC):** Indicates the total discriminative power of each of the models at different levels of anomalies.

- **F1-Score:** The F1-Score balances both precision and recall with the added benefit that it is especially important in imbalanced medical data where anomalies are uncommon.
- **Precision-Recall(PR) Curve:** Gives information on the quality of detection on high class imbalance scenarios.
- **Detection Delay:** This is a measure of the delay between when an anomaly first arises, and its successful detection, an extremely important warning metric in medical equipment.

These metrics combined are the measures of classification accuracy and real-time responsiveness, which are vital in the insurance of clinical safety.

5.3. Quantitative Results between Models.

The comparative analysis showed that the GANs, VAEs and diffusion models had different performance features. GAN-based methods were revealed to have high capability to distinguish anomalies with AUC scores reaching high levels in the case where the deviation patterns were clear. Nevertheless, their performance was not consistent in terms of the performances across the runs because of training instability and sensitivity to hyperparameter decisions.

Models based on VAE had more predictable behavior, especially in the reconstruction of normal operational patterns and placing larger reconstruction errors on anomalous samples. They were slightly less accurate at detection, however, in certain instances, VAEs were much more robust and interpretable with latent-space representations.

Diffusion models always performed better in most evaluation measures than GANs and VAEs. Their iterative denoising formulation allowed modeling of fine-grained complex temporal relationships, leading to better early detection of anomalies and reduced false-positive. Table 3 provides a summary of the comparison of quantitative performance of all the models.

5.4. Robustness and Sensitivity Analysis.

Robustness tests were done to assess the sensitivity of the models to noise, missing values, and the levels of anomalies. GAN-based models exhibited a level of degradation with higher levels of noise, whereas VAEs did not deteriorate significantly because of their probabilistic latent space.

Diffusion models were the most robust, and their detection capabilities remained high even in the case of extreme injection of noise and partial corruption of data. Sensitivity analysis also revealed that the diffusion-based anomaly scores had a smooth scaling with the severity of anomaly, and thus end up being well applicable in graded risk detection in the context of medical equipment monitoring.

5.5. Tables and Figures of Performance Comparison.

In order to present the quantitative results, several comparison tables and visualization figures were created, such as ROC curves, precision-recall plots, and distribution of detection delays. These visual studies indicate trade-offs among accuracy, stability, and computational overhead among generative model families.

On the whole, the experimental findings indicate that GANs and VAEs can still be used in the context of anomaly detection of medical equipment data, but diffusion models offer a more efficient and solid solution to the early detection of anomalies in the safety-critical clinical setup.

6. Discussion and Practical Implications.

6.1. Interpretation of Results

The comparative analysis shows that generative deep learning models offer a plausible basis of early anomaly detection in multi-dimensional, complex data on medical devices performance. In experiments, the models all had different behavioral features that indicate the learning paradigm. The results are supportive of the increasingly influential role of deep learning systems in safety-critical settings where minute deviations can act as an indicator of imminent system failures, a problem also apparent in the field of brain-computer interface and biomedical signal processing (Kachhia et al., 2020; Kachhia and George, 2021).

Finding The findings of the analysis show that a model with the ability to learn rich latent or probabilistic representations is more appropriate in capturing the more subtle temporal correlations in medical telemetry streams. This is in accordance with the previous findings that deep networks are better than shallow models in non-linear biomedical and cyber-physical signals (Kachhia & George, 2021).

6.2. The Strengths and Weaknesses of Each Model.

The use of GAN-based methods showed high levels of anomaly discriminating in cases where the abnormal patterns were distinctly separated in the abnormal behavior and the nominal behavior. Nonetheless, their unstable nature on training and hyperparameter sensitivity is also a problem to deploy them on a regular basis in medical settings. Such restrictions prove to be crucial in controlled systems where predictability and reproducibility are the most important.

VAE-based models were more stable and understandable, which provided structured latent representations. They were applicable to normal day operations because they were formulated probabilistically and thus the normal conditions of operation could be modeled uniformly. However, as they were also based on reconstruction-based anomaly scoring, they could only detect abrupt and transient failures, which is also a concern of the complex signal classification problems in biomedical systems (Kachhia et al., 2020).

Diffusion models had better resilience and flexibility to different working conditions. Their trial and error approach allowed them to distinguish nominal and anomalous behavior slowly and accurately, which justifies their applicability in the context of early-warning systems. It is resilient and reflects the concepts used in self-healing and adaptive security architecture applied to mission-critical systems (Tewari et al., 2025; Tewari et al., 2025).

6.3. Deployment Usability to the Medical setting.

Operationally, in terms of deployments, two important factors include computational overhead and operation reliability. Although diffusion models were shown to be the most effective in the context of detection, their computational requirements might make them impossible to apply in real-time to resource-constrained medical devices. VAE-based models, in their turn, provide a reasonable tradeoff between performance and efficiency, which is why they are more applicable to edge or embedded medical equipment.

These limitations could be overcome by the use of hybrid deployment plans where lightweight models can be used at the device level and more advanced generative models can be deployed in centralized monitoring systems. These types of layered architecture are similar to adaptive and AI-based defense systems employed in contemporary DevSecOps and system monitoring pipelines (Tewari et al., 2025).

6.4. Regulatory and Clinical-Considerations.

Policymaking is a significant barrier to AI-based anomaly detection implementation in the clinical practice. Models have to be transparent, traceable and predictable in various environments of operation. Interpretability of the latent representations in VAE and diffusion models could enable the use of the models in post-hoc analysis and clinical validation, necessary in both regulatory approval and clinician trust.

Moreover, automated detection systems should be modeled as decision-support tools and not decision-makers so that the clinical control is maintained at the center. Other forms of restrictions have been highlighted in previous biomedical AI uses, with a human-in-the-loop validation being required to reduce risk and guarantee patient safety (Kachhia & George, 2021).

6.5. Limitations and Threats to validity.

A number of limitations may be seen on this study. First, the analysis has been done on controlled datasets, and these might not be representative of the operational heterogeneity that is experienced in actual medical settings. Second, although simulated anomalies are beneficial using benchmarking, they might not resemble rare but critical failure modes in deployed devices.

Also, not all computational scalability and long-term model drift were also thoroughly investigated. With the development of medical systems, anomaly detection models cannot afford to lose safety, which is also a problem that is also found in the study of adaptive security and self-healing systems (Tewari et al., 2025). These drawbacks will be tackled by longitudinal studies, practical implementations, and increased cooperation among the AI researchers, clinicians, and regulation authorities.

Conclusion and Future Research Direction

This paper has given a thorough comparative analysis of generative deep learning systems, Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models, in medical device performance data to detect early anomalies. The study revealed that using their patterns in sensitive, multivariate, and time-dependent cues, the current generative models can detect fine changes that are indicative of the devices breaking down, thus the models help to enhance reliability and patient safety.

The results show that the performance trade-offs of the considered models are different. GAN-based methods demonstrated a high level of discriminative power in the case of clearly defined anomaly patterns but were restricted due to the unstable training. The VAEs delivered consistent and interpretable latent representations, which was appropriate to use in continuous monitoring of working medical systems. Diffusion-based models were always the most robust and most accurate detections, especially on low-amplitude and early-stage anomalies, but incurred greater computational cost. Combined, these findings emphasize the fact that there is no single model that can be considered the best, and the choice of a model has to be dictated by the limitations of deployment and safety needs.

Contributionwise, the work contributes to the medical AI research by providing one of the limited number of structured, comparative studies on the generation models specifically applied to the medical device performance monitoring. It fills an urgent need between hypothetical developments in the field of generative modeling and realistic safety implementation in the healthcare domain and offers a methodological basis on how to incorporate anomaly detection into intelligent medical systems of the next generation.

To practitioners, the findings indicate that they should implement hybrid deployment, which would provide good detection performance and at the same time be computationally viable. Simple generative models can be deployed at the edge or at the device level to monitor continuously and at the centre to do an in-depth analysis and scale up to early warning. Notably, anomaly detecting systems are supposed to be created as decision-support systems, which should be transparent and do not impact clinical supervision.

Future studies ought to involve high scale practical validation with longitudinal medical device data sets, study on adaptive learning algorithms to address concept drift and study on training methods that preserve privacy and can be applied to controlled healthcare settings. Miscellaneous efforts are also required to provide better explainability, lessen computational burden, and provide standard evaluation benchmarks that meet clinical and regulatory requirements. Overcoming these issues will be critical towards transferring generative anomaly detection models in the laboratory to trusted medical safety infrastructure elements.

References

[1] Behrendt, F., Bhattacharya, D., Mieling, R., Maack, L., Krüger, J., Opfer, R., & Schlaefer, A. (2025). Guided reconstruction with conditioned diffusion models for unsupervised anomaly detection in brain MRIs. Computers in Biology and Medicine, 186. https://doi.org/10.1016/j.compbiomed.2025.109660

[2] Behrendt, F., Bhattacharya, D., Krüger, J., Opfer, R., & Schlaefer, A. (2023). Patched Diffusion Models for Unsupervised Anomaly Detection in Brain MRI. In Proceedings of Machine Learning Research (Vol. 227, pp. 1019–1032). ML Research Press.

[3] Bose, A. S. C., Eliazer, M., Maheswari, B. U., Sivaneshkumar, A., Sumithra, M., & Qamar, S. (2025). Securing healthcare data in mobile edge computing: a hybrid deep learning framework for privacy and anomaly detection. Knowledge and Information Systems, 67(12), 12079–12117. https://doi.org/10.1007/s10115-025-02586-0

[4] Desai, M., Rumale, A., & Asadinia, M. (2025). SHIELD: Securing Healthcare IoT with Efficient Machine Learning Techniques for Anomaly Detection. In 2025 IEEE 6th Annual World AI IoT Congress, AIIoT 2025 (pp. 521–528). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/AIIoT65859.2025.11105287

[5] Dombrowski, M., & Kainz, B. (2023). Quantifying Sample Anonymity in Score-Based Generative Models with Adversarial Fingerprinting. ArXiv Preprint ArXiv:2306.01363.

[6] Jadhav, S., & Kulkarni, A. (2025). Comprehensive Survey on Detection of Anomalies in Edge Computing Network and Deep Learning Solutions (pp. 37–45). INSTICC. https://doi.org/10.5220/0013344100004646

[7] Kachhia, J., Natharani, R., & George, K. (2020, October). Deep Learning Enhanced BCI Technology for 3D Printing. In 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0125-0130). IEEE. https://doi.org/10.1109/UEMCON51285.2020.9298124

[8] Kachhia, J., & George, K. (2021, January). EEG-based Image Classification using Machine Learning Algorithms. In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0961-0966). IEEE. https://doi.org/10.1109/CCWC51732.2021.9375931

[9] Kadiyala, B., Kumar Alavilli, S., Priya Nippatla, R., Boyapati, S., & Vasamsetty, C. (2025). "Cloud-Integrated IoT-based Healthcare Monitoring and Emergency Response System with Deep Learning Cloud-Integrated IoT-based Healthcare Monitoring and Emergency Response System with Deep Learning. International Research Journal of

Advanced Engineering and Science, 10(1), 193–197.

[10] Loganathan, S., Francis Raj, J. R., Krishnan, R. S., Blessie, A. A., Kumar, V. V., & Kavitha, S. (2025). An AI-Driven Cybersecurity Framework for the Internet of Medical Things: A Hybrid LSTM-CNN-GAN Approach. In 2nd International Conference on Machine Learning and Autonomous Systems, ICMLAS 2025 - Proceedings (pp. 1883–1890). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICMLAS64557.2025.10967601

[11] Park, C., Lim, S., Cha, D., & Jeong, J. (2022). Fv-AD: F-AnoGAN Based Anomaly Detection in Chromate Process for Smart Manufacturing. Applied Sciences (Switzerland), 12(15). https://doi.org/10.3390/app12157549

[12] Sadanandan, B., & Behzadan, V. (2025). Promise of Data-Driven Modeling and Decision Support for Precision Oncology and Theranostics. arXiv preprint arXiv:2505.09899. https://doi.org/10.48550/arXiv.2505.09899

[13] Sadanandan, B., Nobar, B. A., & Behzadan, V. (2025). Comparative Study of Generative Models for Early Detection of Failures in Medical Devices. arXiv preprint arXiv:2505.04845. https://doi.org/10.48550/arXiv.2505.04845

[14] Sethupathy, U. K. A. (2018). Architecting Scalable IoT Telematics Platforms for Connected Vehicles. International Journal of Computer Technology and Electronics Communication, 1(1). https://doi.org/10.15680/IJIRSET.2016.0503278

[15] Sethupathy, U. K. A. REAL-TIME SUPPLY CHAIN PROCESS AUTOMATION AND MONITORING WITH STREAM PROCESSING. https://doi.org/10.56726/IRJMETS9871

[16] Sethupathy, U. K. (2018). Self-Healing systems and telemetry-driven automation in DevOps pipelines. International Journal of Novel Research and Development, 3(7), 148–155. https://doi.org/10.56975/ijnrd.v3i7.309065

[17] Striuk, O., & Kondratenko, Y. (2023). Cross-Domain Reconfigurable GAN with Fuzzy Components for Anomaly Detection. In 2023 13th International Conference on Dependable Systems, Services and Technologies, DESSERT 2023. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/DESSERT61349.2023.10416521

[18] Tewari, R. Shilpi, S., Tyagi, N. K., Khemka, S., & Chhibber, N., (2025, November). An integrated architecture for detecting message-level failures across SWIFT, MQ, and Treasury channels. In International Conference on Innovations and Trends in Advanced Engineering Technologies (ICIAET 2025). https://doi.org/10.2139/ssrn.5917702

[19] Tewari, R., Anugula, P., Bhardwaj, A. K., Chhibber, N., Khemka, S., & Ranjan, P. (2025). AutoGuard: A Self-Healing Proactive Security Layer for DevSecOps Pipelines Using Reinforcement Learning. arXiv preprint arXiv:2512.04368. https://doi.org/10.48550/arXiv.2512.04368

[20] Udoy, I. A., & Hassan, O. (2025, March 1). AI-Driven Technology in Heart Failure Detection and Diagnosis: A Review of the Advancement in Personalized Healthcare. Symmetry. Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/sym17030469

[21] Utsha, U. T., & Morshed, B. I. (2025). GAN-enhanced hybrid deep learning model for real-time cardiac beat detection in the CardioHelp app with edge computing. Smart Health, 38. https://doi.org/10.1016/j.smhl.2025.100601

[22] Wang, J., Liu, M., Shen, W., Ding, R., Wang, Y., & Meijering, E. (2025). EPDiff: Erasure Perception Diffusion Model for Unsupervised Anomaly Detection in Preoperative Multimodal Images. IEEE Transactions on Medical Imaging. https://doi.org/10.1109/TMI.2025.3597545

[23] Yao, H., Liu, M., Yin, Z., Yan, Z., Hong, X., & Zuo, W. (2025). GLAD: Towards Better Reconstruction with Global and Local Adaptive Diffusion Models for Unsupervised Anomaly Detection. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 15129 LNCS, pp. 1–17). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-73209-6_1