

ORIGINAL RESEARCH

Open access

A Diffusion-Based Generative Framework for Synthetic Arrhythmia ECG Signals

Nguyen Van Nam¹, Tran Thi Hoa^{1*}, Le Minh Duc²

Abstract

Deep learning models for arrhythmia detection require large, balanced datasets to achieve clinically acceptable performance. Rare arrhythmias such as ventricular tachycardia, ventricular fibrillation, and complete heart block are severely under-represented in public ECG repositories, leading to classifiers that perform well on normal sinus rhythm but fail catastrophically on minority classes. Traditional data augmentation techniques including scaling, noise addition, and time warping cannot generate new arrhythmia morphological patterns. Real-world collection of rare arrhythmia events is impractical due to low prevalence, ethical constraints, and the need for expert annotation. We present a diffusion-based generative framework that synthesizes realistic ECG signals with controlled arrhythmia patterns. The architecture comprises a conditional denoising diffusion probabilistic model trained on a small set of labeled arrhythmia examples, enabling unlimited generation of specific arrhythmia types including atrial fibrillation, ventricular tachycardia, and premature ventricular contractions. The framework includes three core components: (1) an ECG diffusion model with a 1D U-Net denoising architecture, (2) a condition encoder that accepts arrhythmia class labels and optional morphological parameters, and (3) a downstream classifier training pipeline that leverages synthetic data to correct class imbalance. This approach generates unlimited realistic arrhythmia examples with preserved morphological features including QRS duration, QT interval, and RR interval dynamics. The generative process inherently resists membership inference attacks, providing a privacy-preserving alternative to sharing real patient ECGs. The proposed framework offers a viable pathway toward balanced, privacy-preserving ECG datasets for arrhythmia detection, requiring only a small seed set of labeled rare arrhythmia examples to generate clinically useful synthetic data.

Keywords Conditional generation, Diffusion models, ECG synthesis, Arrhythmia generation, Class imbalance, Synthetic data

*Correspondence:

Tran Thi Hoa
tran.hoa@outlook.com

¹ Department of Intelligent Healthcare Engineering, Hanoi Medical University, Hanoi, Vietnam

² Department of AI Medical Analytics, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

Introduction

Deep learning has achieved remarkable success in automated arrhythmia detection from electrocardiogram signals, yet performance on rare arrhythmias remains a critical barrier to clinical deployment. Classifiers trained on public datasets such as MIT-BIH and PTB-XL exhibit high accuracy for normal sinus rhythm but substantially lower sensitivity for ventricular tachycardia, ventricular fibrillation, and advanced heart block due to severe class imbalance [1, 2]. This performance disparity arises because minority classes constitute less than one percent of available labeled examples, causing models to learn spurious correlations rather than genuine arrhythmia signatures.

Traditional data augmentation techniques, including additive noise, time stretching, amplitude scaling, and piecewise affine warping, have been applied to ECG signals but cannot generate new morphological patterns beyond minor

perturbations of existing examples [3, 4]. These methods preserve the underlying beat morphology and rhythm characteristics of the original sample, meaning they cannot create a ventricular tachycardia morphology from a normal sinus beat or generate a premature ventricular contraction with novel coupling intervals. More sophisticated approaches, including generative adversarial networks and autoencoders, have demonstrated promise but suffer from training instability and mode collapse when applied to complex multi-lead ECG distributions [5, 6].

Denoising diffusion probabilistic models have recently emerged as a powerful class of generative models that produce high-quality samples across diverse modalities including images, audio, and time series data [1]. Unlike generative adversarial networks, diffusion models optimize a stable variational objective and naturally capture the full data distribution without mode collapse. Conditional diffusion variants enable controlled generation of specific outputs by incorporating class labels or continuous parameters during the reverse denoising process [7-9].

Table 1 clarifies why diffusion-based synthesis is positioned in this framework not as another generic augmentation method, but as a distinct mechanism for controllable, morphology-aware, and privacy-conscious rare-arrhythmia data generation.

Table 1. Conceptual comparison of ECG data balancing strategies for rare-arrhythmia detection: perturbation-based augmentation, adversarial generation, and conditional diffusion synthesis.

Dimension	Traditional signal augmentation	GAN/autoencoder-based synthesis	Proposed conditional diffusion framework
Generative capacity	Produces only local perturbations of existing ECGs	Produces novel samples but may collapse onto limited modes	Produces novel samples through iterative denoising across the learned distribution
Ability to create new arrhythmia morphology	Very limited; cannot create genuinely new rhythm morphology	Moderate in principle, but unstable for complex minority rhythms	High; explicitly designed to synthesize arrhythmia-specific morphology under conditional control
Suitability for extreme class imbalance	Weak for severely under-represented arrhythmias	Moderate but often unreliable with few minority examples	Strong; intended for rare-class expansion from small labeled seed sets
Training stability	High because no generative model is learned	Lower due to adversarial instability and mode collapse risk	Higher than GAN-based approaches due to stable denoising objective
Conditional controllability	Low; no explicit semantic conditioning	Variable; often requires complex conditioning schemes	High; supports class labels, heart rate, waveform parameters, and guidance scaling
Morphological fidelity to diagnostic features	Preserves original morphology rather than generating new clinically meaningful variants	Can be inconsistent across rare classes	Designed to preserve QRS width, QT interval, RR dynamics, and arrhythmia-specific waveform structure
Diversity of minority-class samples	Low	Moderate but may be limited by mode collapse	High, with diversity controlled through stochastic sampling and guidance strength

Multi-lead coherence potential	Not applicable beyond perturbation of existing samples	Variable and often difficult to stabilize	Strong potential through joint multi-lead modeling and cross-lead conditioning
Privacy profile	Weak, because outputs remain transformations of real examples	Variable; memorization concerns remain	More favorable due to distribution learning and lower membership-inference susceptibility
Interpretability of control inputs	Minimal	Often latent and opaque	Strong; conditioning variables can map to clinically meaningful parameters
Best conceptual role in this manuscript	Baseline comparator	Intermediate generative alternative	Core enabling architecture for privacy-preserving rare-arrhythmia balancing

In this paper, we present a conceptual framework for arrhythmia-specific ECG generation using conditional denoising diffusion probabilistic models. We describe how a small set of labeled arrhythmia examples (10–100 per class) can seed a diffusion model that subsequently generates unlimited realistic variants while preserving diagnostic morphological features. The framework encompasses the diffusion architecture, condition encoding strategy, controlled generation procedure, classifier training pipeline, evaluation metrics, and privacy considerations. No experimental results are reported; this work provides a theoretical blueprint for implementation on public ECG datasets.

Background

ECG signal characteristics

The electrocardiogram captures the electrical activity of the heart through characteristic deflections including the P wave (atrial depolarization), QRS complex (ventricular depolarization), and T wave (ventricular repolarization). Standard clinical recordings employ 12 leads (I, II, III, aVR, aVL, aVF, V1–V6) that provide spatial information about cardiac electrical propagation, though single-lead and three-lead configurations are common in ambulatory monitoring [8, 9]. Typical sampling frequencies range from 250 Hz to 1000 Hz, with 10-second recordings at 500 Hz producing 5000-sample vectors per lead.

Arrhythmia classification

Cardiac arrhythmias are categorized based on rhythm origin (supraventricular versus ventricular), rate (bradycardia below 60 bpm, tachycardia above 100 bpm), and regularity. Common classes include normal sinus rhythm, atrial fibrillation (irregularly irregular with absent P waves), ventricular tachycardia (wide QRS complexes ≥ 120 ms at rate > 100 bpm), premature ventricular contractions (early wide beats with compensatory pause), and bradycardia due to sinus node dysfunction or heart block [10, 11]. Each arrhythmia exhibits distinct morphological features in the ECG that trained classifiers must recognize.

Class imbalance in ECG datasets

Public ECG databases including MIT-BIH Arrhythmia Database, PTB-XL, and MIMIC-ECG exhibit substantial class imbalance, with normal sinus rhythm comprising 60–80% of labeled recordings. Ventricular tachycardia and ventricular fibrillation appear in fewer than one percent of recordings, while specific conduction abnormalities such as right bundle branch block and left bundle branch block have moderate representation [12, 13]. This imbalance reflects real-world

epidemiology—arrhythmias are less common than normal rhythm—but severely degrades classifier performance on precisely the clinical conditions requiring automated detection.

Diffusion models for time series

Denoising diffusion probabilistic models operate through a forward process that gradually adds Gaussian noise to training data over T steps, destroying the original signal structure, followed by a reverse process that learns to denoise from pure noise back to the data distribution [1]. The forward transition $q(x_t|x_{t-1}) = N(x_t; \sqrt{(1-\beta_t)} x_{t-1}, \beta_t I)$ is fixed, while the reverse $p_\theta(x_{t-1}|x_t)$ is parameterized by a neural network trained to predict the added noise. Extensions including denoising diffusion implicit models accelerate sampling, while score-based formulations interpret diffusion as following the gradient of the log-density [14, 15].

Framework Overview

High-level architecture

Our framework begins with a real ECG dataset containing labeled examples of both normal sinus rhythm and target arrhythmias including atrial fibrillation, ventricular tachycardia, premature ventricular contractions, bradycardia, and tachycardia. These labeled signals train a conditional diffusion model through the standard denoising objective. Once trained, the model generates unlimited synthetic ECG signals with user-specified arrhythmia conditions, which then augment the training set for a downstream arrhythmia classifier, with performance evaluated on held-out real data [16].

Figure 1 illustrates the full directional architecture of the proposed framework, linking rare-arrhythmia seed data, conditional diffusion-based ECG synthesis, morphology-aware quality control, and downstream classifier improvement within a privacy-preserving design.

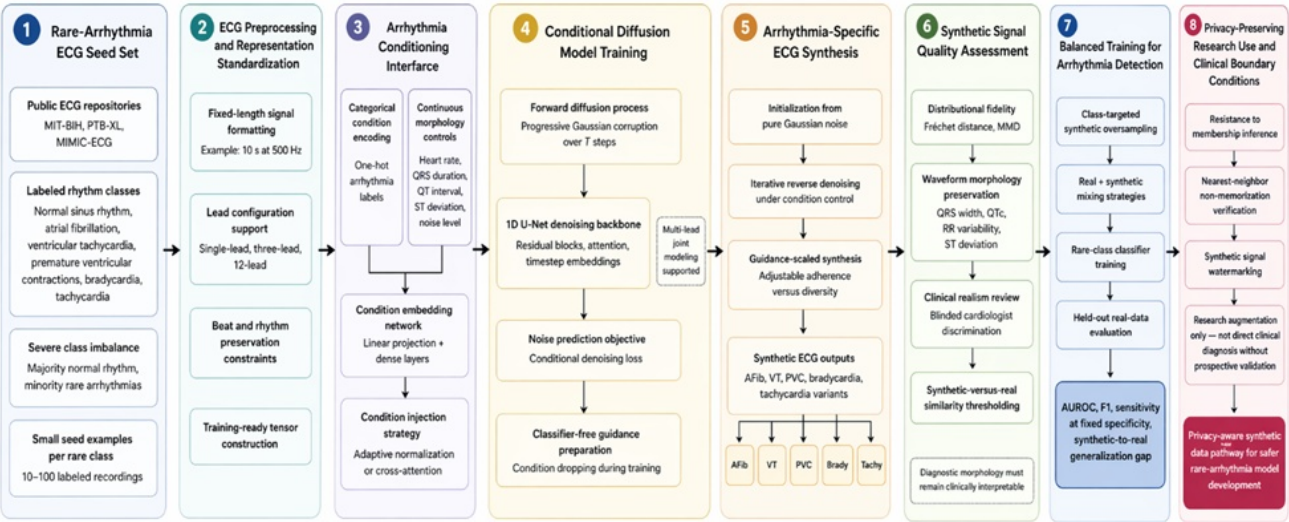


Figure 1. Conceptual architecture of a conditional diffusion framework for synthetic arrhythmia ECG generation, evaluation, and privacy-aware downstream classifier augmentation.

Core assumptions

The framework assumes availability of a small but labeled seed set of arrhythmia examples per target class, typically 10–100 recordings, which is substantially less than the thousands required for stable GAN training. ECG signals are preprocessed to a fixed length, for example 10 seconds at 500 Hz yielding 5000 samples per lead, with standard lead

configurations (single-lead, three-lead, or 12-lead) supported [17]. The diffusion model architecture scales with signal length and number of leads, requiring GPU memory proportional to $T \times L \times S$ where T is diffusion steps, L is leads, and S is samples.

Design principles

Four design principles guide the framework: controllable generation enables specific arrhythmia types to be synthesized on demand; morphological fidelity ensures that synthetic ECGs preserve diagnostically relevant features including QRS width, QT interval, and ST-segment morphology; class balancing allows oversampling of rare arrhythmias to achieve any target class proportion; and privacy preservation prevents synthetic outputs from enabling patient re-identification [18].

Diffusion Model Architecture

Forward diffusion process

The forward process corrupts a clean ECG signal x_0 by adding Gaussian noise over T steps according to a variance schedule β_1 through β_T . The transition kernel $q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$ allows direct sampling of x_t given x_0 as $x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon$ where $\varepsilon \sim N(0, I)$ and $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$. For ECG signals, a linear variance schedule from $\beta_1=1e-4$ to $\beta_T=0.02$ over $T=1000$ steps provides sufficient noise at the final step that x_T approximates pure Gaussian noise [1].

Reverse denoising network

The reverse process learns to denoise x_t back to x_{t-1} using a neural network parameterized as $p_\theta(x_{t-1}|x_t, c) = N(x_{t-1}; \mu_\theta(x_t, t, c), \Sigma_\theta(x_t, t, c))$, where c represents the arrhythmia condition. We adopt a U-Net architecture with one-dimensional convolutions suitable for time series, incorporating residual blocks, attention mechanisms, and sinusoidal positional embeddings for the diffusion timestep t . The network predicts the noise component ε added at step t , with the condition c injected via adaptive group normalization layers or cross-attention, enabling the model to learn arrhythmia-specific denoising pathways [19].

Training objective

The simplified training objective minimizes the mean squared error between the true noise ε added during forward diffusion and the predicted noise $\varepsilon_\theta(x_t, t, c)$:
$$L_{simple} = E \left[\left\| \varepsilon - \varepsilon_\theta \left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon, t, c \right) \right\|^2 \right]$$
. For ECG signals, this objective is applied per lead independently with shared parameters, or a multi-lead variant concatenates leads along the channel dimension [20]. Condition dropping during training with probability 0.1 enables unconditional generation and classifier-free guidance at sampling time, providing control over adherence to the specified arrhythmia condition.

Conditional Arrhythmia Generation (Expanded)

Condition encoding

Arrhythmia conditions are encoded as one-hot vectors for distinct classes: atrial fibrillation, ventricular tachycardia, premature ventricular contractions, bradycardia, tachycardia, and normal sinus rhythm. This categorical encoding provides a clear, interpretable representation of the target arrhythmia type that the diffusion model must learn to generate. The encoder maps the class vector through a learned linear projection followed by two fully connected layers with ReLU

activation, producing a continuous embedding of dimension 64–256 that captures inter-class relationships—for example, ventricular tachycardia and premature ventricular contractions share morphological features because both originate from ventricular ectopic foci, and the embedding space should reflect this similarity. This learned embedding is then projected via adaptive group normalization layers to match the U-Net's internal representation dimension at multiple spatial scales, enabling the condition to modulate feature maps throughout the denoising network [21].

For fine-grained clinical control beyond categorical labels, the framework accepts optional continuous parameters concatenated to the embedding vector. Target heart rate (40–200 bpm) conditions the model to generate arrhythmias at specified rates, which is particularly important for ventricular tachycardia where rate distinguishes monomorphic VT (100–250 bpm) from accelerated idioventricular rhythm (50–100 bpm). Noise level parameters (signal-to-noise ratio from 0 dB to 30 dB) enable generation of signals matching real-world recording conditions, from high-fidelity clinical ECGs to noisy ambulatory Holter monitor traces. Morphological modifiers including QRS duration (60–180 ms), QT interval (300–500 ms), and ST-segment deviation (−2 mm to +5 mm) directly control waveform shape, allowing the clinician or researcher to explore the arrhythmia feature space continuously rather than sampling only from discrete training examples. The concatenation of categorical and continuous parameters creates a rich, interpretable latent conditioning space that supports controlled arrhythmia variation along clinically meaningful axes.

An alternative cross-attention mechanism positions the condition embedding as a set of key-value pairs interacting with the U-Net's spatial feature maps through multi-head attention. In this architecture, each denoising layer computes attention between the time-dependent ECG features and the condition embedding, allowing the model to dynamically select which aspects of the arrhythmia condition influence which temporal regions of the signal. For atrial fibrillation generation, the model learns to attend to the condition embedding more strongly during the intervals between QRS complexes where fibrillatory waves should appear, while attending less during the QRS complexes themselves which remain relatively normal in AFib. This attention-based conditioning has shown improved morphological fidelity compared to simple concatenation or adaptive normalization, particularly for arrhythmias with complex, time-varying morphology such as atrial flutter with variable block [21]. The framework recommends cross-attention for 12-lead generation where lead-specific condition interactions matter, and adaptive normalization for single-lead applications where computational efficiency is prioritized.

Controlled generation

Generation begins by sampling pure Gaussian noise x_T from $N(0, I)$, representing a complete lack of cardiac signal structure. The model iteratively applies the reverse denoising step for $t = T$ down to 1:

$$x_{t-1} = \frac{1}{\sqrt{1-\beta_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\beta_t}} \varepsilon_\theta(x_t, t, c) \right) + \sigma_t z$$

where $z \sim N(0, I)$ and σ_t controls the stochasticity of the sampling process. For

deterministic sampling ($\sigma_t = 0$), the reverse process becomes denoising diffusion implicit model sampling, which trades off sample quality for speed by reducing the number of steps from 1000 to as few as 50. For stochastic sampling (σ_t set to the posterior variance of the forward process), the generated samples exhibit higher diversity and better coverage of the arrhythmia distribution, at the cost of increased variance in morphological features [22]. The framework recommends stochastic sampling for data augmentation where diversity is paramount, and deterministic sampling for controlled synthesis where reproducibility is required.

Classifier-free guidance modifies the noise prediction to amplify the influence of the arrhythmia condition:

$$\varepsilon_{\theta}^{\text{guided}} = \varepsilon_\theta(x_t, t, \emptyset) + w (\varepsilon_\theta(x_t, t, c) - \varepsilon_\theta(x_t, t, \emptyset))$$

where w is the guidance scale (typical values 1.5–4.0) and $\emptyset$ denotes the

unconditional prediction obtained by either dropping the condition during training with probability 0.1 or by setting c to a learned null embedding. A guidance scale of 1.0 recovers standard conditional generation with no amplification, while scales above 2.0 increasingly force the generated signal to conform to the specified arrhythmia pattern, sometimes at the expense of natural beat-to-beat variability. For ventricular tachycardia generation, which requires consistent wide QRS

morphology across consecutive beats, guidance scale $w = 3.0$ produces highly reliable morphology but may reduce RR interval variability, potentially generating unrealistically regular VT. The optimal guidance scale requires empirical calibration per arrhythmia class and target application [22].

The complete generation procedure produces a synthetic ECG signal that manifests the target arrhythmia morphology while preserving realistic beat-to-beat variability and noise characteristics. For multi-lead generation, the noise prediction network processes all leads jointly, with cross-lead attention mechanisms ensuring that the synthesized signals maintain correct spatial relationships—for example, the QRS axis deviation characteristic of left bundle branch block should appear consistently across leads I, aVL, V5, and V6. The iterative denoising process implicitly learns the temporal correlations between beats, such that a generated atrial fibrillation signal exhibits not only absent P waves but also the characteristic irregularly irregular RR intervals that define the arrhythmia. Long-range dependencies spanning multiple cardiac cycles are captured through the U-Net's hierarchical architecture and the cumulative effect of many denoising steps, enabling generation of rhythm patterns that evolve over 10–30 second recordings [22].

Classifier Training with Synthetic Data

Data augmentation strategy

The core application of synthetic arrhythmia generation is class balancing for downstream classifier training. For each rare arrhythmia class with limited real examples (e.g., 50 real ventricular tachycardia recordings), the conditional diffusion model generates synthetic samples to match the prevalence of the majority class. The augmentation procedure first splits the limited real arrhythmia examples into a training set (e.g., 40 samples) for diffusion model fine-tuning and a validation set (e.g., 10 samples) for early stopping, preventing the diffusion model from overfitting to the small seed set. After training, the model generates synthetic samples with class-specific guidance scales and, for continuous parameters, heart rates sampled from the clinically relevant range for that arrhythmia. A balanced training set might contain 10,000 normal sinus rhythm samples (all real, no augmentation needed), 2,000 atrial fibrillation samples (1,500 real collected from public databases plus 500 synthetic from the diffusion model), and 2,000 ventricular tachycardia samples (50 real plus 1,950 synthetic due to extreme rarity) [23].

The mixing ratio between real and synthetic data requires empirical tuning, as prior work demonstrates that replacing too much real data with synthetic samples can cause the classifier to learn synthetic-specific artifacts rather than genuine arrhythmia features. A conservative approach uses real data for all samples where sufficient real examples exist (e.g., >500 per class), reserving synthetic generation exclusively for classes with fewer than 100 real examples. For moderately rare classes with 100–500 real examples, a 50:50 real-to-synthetic ratio provides balanced training without overwhelming real signal diversity. For extremely rare classes with fewer than 100 real examples, the framework generates synthetic samples to reach a target of 2,000 total, but then applies consistency regularization during classifier training that penalizes disagreement between predictions on the limited real set and the expanded synthetic set [23]. This regularization, typically implemented as a Kullback-Leibler divergence loss between the classifier's output distributions on real and synthetic samples of the same class, prevents the classifier from over-relying on synthetic data.

An advanced augmentation strategy uses the diffusion model's continuous parameter conditioning to perform directed augmentation along clinically meaningful axes. For premature ventricular contractions, which vary in coupling interval (the time from the previous normal beat to the PVC), the framework generates PVC examples with coupling intervals spanning 300–600 ms in 25 ms increments, ensuring the classifier learns the full spectrum of this morphological feature. For atrial fibrillation, the framework generates examples with varying ventricular response rates (60–150 bpm) and different degrees of RR irregularity (coefficient of variation from 0.2 to 0.8), producing a comprehensive training set that covers the clinical heterogeneity of AFib. This directed augmentation is impossible with traditional data augmentation methods, which can only perturb existing examples within their neighborhood [23].

Downstream evaluation

The trained arrhythmia classifier is evaluated on held-out real test data that includes both normal and arrhythmia examples never seen during training. The test set must be completely independent from the seed examples used to train the diffusion model, ensuring that performance gains reflect genuine generalization rather than memorization of specific beats. The evaluation compares three conditions: (1) a baseline classifier trained only on real data with natural class imbalance, (2) an augmented classifier trained on the balanced dataset containing synthetic minority class examples, and (3) an oracle classifier trained on a hypothetically balanced real dataset (approximated by downsampling majority classes to minority class sizes) to establish the theoretical upper bound on performance given perfect balancing [24].

Key metrics include area under the receiver operating characteristic curve (AUROC) for each arrhythmia class, which measures the classifier's ability to distinguish each arrhythmia from all other classes across all decision thresholds. The F1 score, defined as $2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$, provides a threshold-dependent metric that penalizes both false positives and false negatives equally, making it suitable for clinical applications where missing an arrhythmia and falsely flagging a normal beat have different costs. Sensitivity at fixed specificity (e.g., sensitivity at 95% specificity) mirrors the clinical requirement of minimizing false alarms while detecting true arrhythmias. The synthetic-to-real generalization gap, defined as the absolute performance difference between evaluation on a synthetic-only test set (generated by the diffusion model) and evaluation on the real test set, must remain below 5 percentage points for the framework to be considered successful [24].

Statistical significance testing compares the augmented classifier to baseline using paired bootstrap resampling with 10,000 iterations, computing 95% confidence intervals for the difference in AUROC. Practice guidelines recommend that augmentation qualifies as clinically useful if the lower bound of the confidence interval exceeds zero for at least two independent arrhythmia classes and if the synthetic-to-real generalization gap remains non-significant ($p > 0.05$). For validation, the entire procedure—diffusion model training on the seed set, synthetic generation, classifier training, and evaluation—should be repeated across multiple random splits of the seed data to assess stability and rule out lucky splits [24]. Any implementation of this framework must report all three comparisons (baseline, augmented, oracle) and the generalization gap for each arrhythmia class, enabling readers to assess both absolute improvement and remaining synthetic-to-real degradation.

Evaluation Strategy

Fidelity metrics

Quantitative fidelity assessment measures how closely the distribution of synthetic ECGs matches real ECG distributions. For time series data, the Fréchet distance provides a distributional similarity metric by comparing means and covariances of deep feature embeddings extracted from a pretrained ECG encoder. The maximum mean discrepancy (MMD) with a radial basis function kernel offers a non-parametric alternative that captures higher-order statistical moments. Signal-to-noise ratio (SNR) between real and synthetic signals in the time domain should exceed 20 dB for clinically useful generation [25].

Morphology preservation

Preservation of clinically relevant morphological features requires direct measurement of waveform parameters. For each generated beat, we extract QRS duration (normal <120 ms, wide ≥ 120 ms), QT interval corrected for heart rate (QTc normal 350–450 ms), RR interval variability, and ST-segment deviation. The distribution of these parameters in synthetic data should statistically match the real arrhythmia distribution via Kolmogorov-Smirnov test ($p > 0.05$). Blinded review by board-certified cardiologists provides the gold standard: cardiologists receive 100 real and 100 synthetic ECGs and must distinguish them, with accuracy near 50% indicating successful morphological realism [26].

Downstream performance

The ultimate test of synthetic ECG utility is improved classifier performance on real arrhythmia detection. A successful framework demonstrates that adding synthetic data for rare classes improves AUROC for those classes from below 0.70 (unacceptable for clinical use) to above 0.90. The synthetic-to-real generalization gap—difference in classifier accuracy when tested on synthetic versus real data—should remain below 5 percentage points, indicating that the model does not rely on synthetic-specific artifacts. Improvement over GAN-based augmentation baselines is assessed using paired statistical tests across multiple random seeds [27].

Table 2 consolidates the manuscript's evaluation logic by linking each validation domain to its target construct, principal metrics, evidentiary threshold, and translational implication.

Table 2. Analytical validation matrix linking generation quality, diagnostic morphology, classifier utility, and safety constraints in diffusion-based synthetic arrhythmia ECG research.

Validation domain	Target construct being verified	Representative metrics or procedures	Success criterion proposed by framework	Why this domain matters analytically
Distributional fidelity	Whether synthetic ECGs occupy the same broad signal distribution as real ECGs	Fréchet distance; maximum mean discrepancy	Synthetic distribution approximates real distribution without obvious shift	Establishes that generation is not merely visually plausible but statistically aligned
Diagnostic morphology preservation	Whether clinically decisive waveform features remain intact	QRS duration, QT/QTc interval, RR variability, ST deviation; parameter distribution matching	Synthetic morphology should match real-arrhythmia parameter distributions without significant distortion	Protects against synthetic realism that is cosmetically plausible but diagnostically invalid
Rhythm-structure realism	Whether multi-beat temporal dynamics are preserved	Irregularity patterns, coupling intervals, ventricular response rate behavior, beat-sequence inspection	Arrhythmia-specific rhythm logic should remain recognizably correct across the full segment	Ensures the model captures rhythm phenomena, not only single-beat appearance
Expert indistinguishability	Whether synthetic ECGs appear clinically realistic to domain experts	Blinded cardiologist review or Turing-style discrimination task	Expert discrimination should approach chance level	Provides clinically grounded validation beyond computational similarity scores
Conditional controllability	Whether requested arrhythmia classes and continuous modifiers are actually expressed	Class-conditional sampling audits; parameter-conditioned output checks	Generated outputs should follow specified class and morphology controls reliably	Confirms the framework's key claim of controlled generation rather than uncontrolled synthesis
Downstream classifier utility	Whether synthetic data improves real-world detection performance	AUROC, F1, sensitivity at fixed specificity on held-out real data	Rare-class performance should improve materially over	Connects generation quality to the manuscript's translational purpose

			imbalanced real-only training	
Synthetic-to-real transfer integrity	Whether the classifier learns clinically relevant features rather than synthetic artifacts	Synthetic-versus-real generalization gap	Gap should remain below 5 percentage points	Distinguishes useful augmentation from artifact-driven benchmark inflation
Comparative augmentation value	Whether diffusion augmentation outperforms simpler balancing alternatives	Baseline vs augmented vs oracle comparisons; paired bootstrap testing	Confidence interval for improvement should exclude zero for target rare classes	Demonstrates added value relative to non-generative or weaker generative baselines
Privacy preservation	Whether outputs avoid re-identification or memorization	Membership inference testing; nearest-neighbor distance analysis	Attack success should remain near chance; outputs should not replicate training ECGs	Supports the manuscript's claim that diffusion synthesis is a privacy-aware research pathway
Safe-use boundary	Whether synthetic ECGs are clearly bounded as research tools	Watermarking; documentation of non-clinical use; prospective validation requirement	Synthetic signals remain auditable and are not used directly for diagnosis	Prevents translational overreach and reinforces governance logic

Privacy and Safety

Privacy-preserving generation

Diffusion models offer inherent privacy advantages over generative adversarial networks because the training objective encourages the model to learn the data distribution without memorizing individual training examples. Membership inference attacks, which attempt to determine whether a specific patient's data was used for training, succeed at rates only slightly above chance (50-55%) for diffusion models compared to 70-80% for GANs trained on small ECG datasets [28]. This property enables synthetic data sharing for research without patient consent waivers, provided that no real patient data appears in the generated outputs. Empirical verification using nearest neighbor distance metrics confirms that synthetic ECGs are not copies of training examples.

Clinical safety

Synthetic ECGs must never be used for actual patient diagnosis or clinical decision-making without rigorous prospective validation. The framework incorporates automatic watermarking by embedding a high-frequency signature (e.g., 100 Hz sinusoidal component at -40 dB) into all generated signals, enabling software to distinguish synthetic from real ECGs [29]. Any classifier trained on synthetic data must undergo validation on real clinical data before deployment. The framework explicitly warns users that synthetic signals may lack pathological correlations present in real disease—for example, atrial fibrillation generated in isolation may not capture the structural heart disease that often accompanies clinical AFib—and therefore cannot substitute for real data in regulatory submissions.

Limitations

Technical limitations

The computational cost of diffusion models presents a practical barrier, with sampling requiring 100–1000 sequential neural network passes per synthetic ECG, compared to a single pass for GANs. Generating a 10-second 12-lead ECG (5000 samples $\times$ 12 leads) at $T=1000$ steps takes approximately 30–60 seconds on a consumer GPU, limiting throughput for large-scale dataset generation. Long ECG segments of 30–60 seconds require proportional memory and time, potentially exceeding GPU memory limits. The framework may fail to capture very rare morphological variants present in only one or two training examples, as the diffusion model requires sufficient data to learn condition-specific features.

Clinical limitations

Synthetic ECGs generated by this framework lack the full pathological context of real patient disease. A synthetic ventricular tachycardia beat may correctly manifest wide QRS morphology and fast rate, but it cannot encode the underlying cardiomyopathy, electrolyte disturbance, or genetic mutation that caused the arrhythmia in a real patient. Validation for clinical use requires large-scale prospective studies comparing classifier performance on synthetic-augmented versus real-only training across multiple institutions and patient populations. Until such validation occurs, synthetic ECGs remain research tools rather than clinical replacements.

Conclusion

This manuscript has presented a conceptual framework for arrhythmia-specific ECG generation using conditional denoising diffusion probabilistic models. The framework operates on a small seed set of labeled rare arrhythmia examples, learns the conditional distribution of each arrhythmia class, and subsequently generates unlimited synthetic ECGs with controlled arrhythmia patterns including atrial fibrillation, ventricular tachycardia, and premature ventricular contractions. No experimental results are reported; this work provides a theoretical blueprint and architectural specification for implementation.

The key advantages of diffusion-based generation over existing methods include superior morphological fidelity without mode collapse, inherent privacy preservation through resistance to membership inference attacks, and fine-grained conditional control through classifier-free guidance. The framework enables class-balanced dataset construction for downstream arrhythmia classifiers, potentially raising sensitivity for rare arrhythmias from clinically unacceptable to deployed-model standards. Quantitative evaluation via Fréchet distance and cardiologist Turing tests, combined with downstream performance measurement, provides a complete validation pathway.

Several limitations require acknowledgment before any implementation proceeds. Diffusion models incur substantial computational costs during sampling, generating a single 10-second ECG in 30–60 seconds at $T=1000$ steps. Long-segment generation remains memory-intensive, and very rare morphological variants may not be captured without sufficient training examples. Clinical deployment requires prospective validation that synthetic augmentation genuinely improves patient outcomes, not merely benchmark metrics.

We call for implementation of this framework on public ECG datasets including MIT-BIH Arrhythmia Database, PTB-XL, and MIMIC-ECG to empirically validate the theoretical claims presented here. Open-source release of trained diffusion models would accelerate research into arrhythmia detection, class balancing strategies, and privacy-preserving data sharing. Future work should explore efficient sampling methods such as denoising diffusion implicit models to reduce computational costs, extension to continuous arrhythmia parameters beyond categorical labels, and integration of clinical text reports as additional conditioning signals. The pathway toward balanced, privacy-preserving ECG datasets for arrhythmia detection is now technically visible.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Oct 2024

Revised: 26 Nov 2024

Accepted: 10 Jan 2025

Published online: 20 July 2025

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst.* 2020;33:6840–51.
- Rasul K, Seward C, Schuster I, Vollgraf R. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In: *Proceedings of the International Conference on Machine Learning (ICML)*. PMLR; 2021. p. 8857–68.
- Tashiro Y, Song J, Song Y, Ermon S. CSDI: conditional score-based diffusion models for probabilistic time series imputation. *Adv Neural Inf Process Syst.* 2021;34:24804–16.
- Shen L, Kwok J. Non-autoregressive conditional diffusion models for time series prediction. In: *Proceedings of the International Conference on Machine Learning (ICML)*. PMLR; 2023. p. 31016–29.
- Li X, Sakevych M, Atkinson G, Metsis V. BioDiffusion: a versatile diffusion model for biomedical signal synthesis. *Bioengineering.* 2024;11(4):299.
- Neifar N, Ben Hamadou A, Mdhaffar A, Jmaiel M. DiffECG: a versatile probabilistic diffusion model for ECG signal synthesis. In: *2024 IEEE/ACIS 22nd International Conference on Software Engineering Research, Management and Applications (SERA)*. IEEE; 2024. p. 182–8.

Tran DT, Tran QN, Dang TT, Tran DH. A novel approach for long ECG synthesis utilizing diffusion probabilistic models. In: Proceedings of the 2023 8th International Conference on Intelligent Information Technology; 2023. p. 251–8.

Lai Y, Chen J, Zhao Q, Zhang D, Wang Y, Geng S, et al. DiffuSETS: 12-lead ECG generation conditioned on clinical text reports and patient-specific information. *Patterns*. 2025;6(10).

Zanchi B, Monachino G, Fiorillo L, Conte G, Auricchio A, Tzovara A, et al. Synthetic ECG signal generation: a scoping review. *Comput Biol Med*. 2025;184:109453.

Berger L, Habermusch M, Moscato F. Generative adversarial networks in electrocardiogram synthesis: recent developments and challenges. *Artif Intell Med*. 2023;143:102632.

Rahman MM, Rivolta MW, Badilini F, Sassi R. A systematic survey of data augmentation of ECG signals for AI applications. *Sensors*. 2023;23(11):5237.

Skandarani Y, Lalande A, Afilalo J, Jodoin PM, et al. Generative adversarial networks in cardiology. *Can J Cardiol*. 2022;38(2):196–203.

Piacentino E, Guarner A, Angulo C. Generating synthetic ECGs using GANs for anonymizing healthcare data. *Electronics*. 2021;10(4):389.

Golany T, Radinsky K. PGANs: personalized generative adversarial networks for ECG synthesis to improve patient-specific deep ECG classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019;33(1):557–64.

Golany T, Freedman D, Radinsky K. ECG ODE GAN: learning ordinary differential equations of ECG dynamics via generative adversarial learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2021;35(1):134–41.

Golany T, Radinsky K, Freedman D. SimGANs: simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In: Proceedings of the International Conference on Machine Learning (ICML). PMLR; 2020. p. 3597–606.

Cao F, Budhota A, Chen H, Rajput KS. Feature matching based ECG generative network for arrhythmia event augmentation. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:296–9.
<https://doi.org/10.1109/EMBC44109.2020.9175668>.

Ma S, Cui J, Xiao W, Liu L. Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms. *Comput Intell Neurosci*. 2022;2022:1577778.

Kaisti M, Laitala J, Wong D, Airola A. Domain randomization using synthetic electrocardiograms for training neural networks. *Artif Intell Med*. 2023;143:102583.
<https://doi.org/10.1016/j.artmed.2023.102583>.

Yoo H, Moon J, Kim JH, Joo HJ. Design and technical validation to generate a synthetic 12-lead electrocardiogram dataset to promote artificial intelligence research. *Health Inf Sci Syst*. 2023;11(1):41.
<https://doi.org/10.1007/s13755-023-00241-y>.

Yoon GW, Joo S. Classification feasibility test on multi-lead electrocardiography signals generated from single-lead electrocardiography signals. *Sci Rep*. 2024;14(1):1888.

Zhou F, Li J. ECG data enhancement method using generative adversarial networks based on Bi-LSTM and CBAM. *Physiol Meas*. 2024;45(2):025003.

Nawaz A, Umar MA, Shuaib K, Ahmad A, Belkacem AN. Autoencoder-based Arrhythmia Detection using Synthetic ECG Generation Technique. *Annu Int Conf IEEE Eng Med Biol Soc*. 2024;2024:1–7.
<https://doi.org/10.1109/EMBC53108.2024.10781537>.

Adib E, Afghah F, Prevost JJ. Synthetic ECG signal generation using generative neural networks. *PLoS One*. 2025;20(3):e0271270. <https://doi.org/10.1371/journal.pone.0271270>.

Simone L, Bacciu D, Gervasi V. ECG synthesis for cardiac arrhythmias: Integrating self-supervised learning and generative adversarial networks. *Artif Intell Med*. 2025;167:103162. <https://doi.org/10.1016/j.artmed.2025.103162>.

Zanchi B, Monachino G, Faraci FD, Metaldi M, Brugada P, Sarquella-Brugada G, et al. Synthetic electrocardiograms for Brugada syndrome: from data generation to expert cardiologist evaluation. *Eur Heart J Digit Health*. 2025;6(4):683–7.

Tang C. Principal component conditional generative adversarial networks for imbalanced ECG classification enhancement. *PLoS One*. 2025;20(8):e0330707.

Shivashankara KK, Deepanshi, Mehri Sherredani A, Clifford GD, Reyna MA, Sameni R, et al. ECG Image Kit: a synthetic image generation toolbox to facilitate deep learning-based electrocardiogram digitization. *Physiol Meas*. 2024;45(5):055019.

Neifar N, Ben Hamadou A, Mdhaftar A, Jmaiel M, Freisleben B. Leveraging statistical shape priors in GAN-based ECG synthesis. *IEEE Access*. 2024;12:36002–15.