

## PAPER

# Manifold-Aware Diffusion-Augmented Contrastive Learning for Noise-Robust Biosignal Representation

Rami Zewail  Egypt-Japan University  
of Science & Technology,  
Alexandria, Egypt[rami.zewail@ejust.edu.eg](mailto:rami.zewail@ejust.edu.eg)**ABSTRACT**

Learning robust representations for physiological time-series signals continues to pose a substantial challenge in developing efficient few-shot learning applications. This is largely due to the complex pathological variations in bio signals. In this context, this paper introduces a manifold-aware Diffusion-Augmented Contrastive Learning (DACL) framework, which efficiently leverages the generative structure of latent diffusion models (LDMs) with the discriminative power of supervised contrastive learning. The proposed framework operates within a contextualized scattering latent space derived from Scattering Transformer (ST) features. Within a contrastive learning framework, we employ a forward diffusion process in the scattering latent space as a structured manifold-aware feature augmentation technique. We assessed the proposed framework using the PhysioNet 2017 Electrocardiogram (ECG) benchmark dataset. The proposed method achieved a competitive AUROC of 0.9741 in the task of detecting atrial fibrillation (AF) from a single-lead ECG signal. The proposed framework achieved performance on par with relevant state-of-the-art related works. In-depth evaluation findings suggest that early-stage diffusion serves as an ideal “local manifold explorer,” producing embeddings with greater precision than typical augmentation methods while preserving inference efficiency.

**KEYWORDS**

contrastive learning, diffusion models, deep clustering, data augmentation, representation learning, noise-robustness, bio signal processing, scattering transformer

## 1 INTRODUCTION

Within the domain of biomedical artificial intelligence, the limited availability of expert-annotated data constitutes a significant impediment. In contrast to natural image domains, physiological signals, such as Electrocardiogram (ECG) and Electroencephalogram (EEG), are costly to annotate and frequently exhibit significant class imbalances. This scarcity of data and expertise constrains the generalization potential of deep learning models. Consequently, self-supervised learning frameworks

Zewail, R. (2026). Manifold-Aware Diffusion-Augmented Contrastive Learning for Noise-Robust Biosignal Representation. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(4), pp. 62–75. <https://doi.org/10.3991/ijoe.v22i04.59841>

Article submitted 2025-12-03. Revision uploaded 2026-01-07. Final acceptance 2026-01-09

© 2026 by the authors of this article. Published under CC-BY.

are particularly attractive in this field. The effectiveness of contrastive learning frameworks [1–3] is fundamentally dependent on the quality of data augmentation employed to produce distinct yet semantically consistent “views” of a data sample. While the domain of computer vision possesses a well-established set of heuristic augmentations, the formulation of meaning-preserving transformations for physiological time-series data remains a significant and unresolved challenge [4–6]. Simple transformations, such as jittering or permutation, often fail to respect the underlying physiological manifold, potentially compromising critical and pathological characteristics. Therefore, there is a growing need for data-driven augmentation strategies rather than heuristic ones. Parallel to these developments, diffusion models [7, 8] have emerged as powerful generative tools, exhibiting remarkable capabilities in modeling complex data distributions through progressive noising and denoising processes. While most research has focused on the *reverse* process for high-fidelity generation, the forward noising process offers an overlooked but powerful property: it provides a mechanism for creating increasingly noisy iterations of a data point that remain mathematically grounded in the manifold structure of the data [9]. Motivated by these concurrent advancements, we introduce Diffusion-Augmented Contrastive Learning (DACL), a novel framework that substitutes heuristic augmentations with a structured learned strategy derived from the forward diffusion process. Rather than synthesizing new samples through costly reverse sampling, DACL generates training views by sampling embeddings at varying noise levels along the forward diffusion trajectory. This methodology enables the encoder to learn representations that are invariant to manifold-consistent perturbations. To ensure that the input features are robust and structured, our framework operates within a contextualized latent space derived from the Scattering Transformer (ST) [10], adopting the computational efficiency paradigm of LDMs [11]. The specific contributions of this paper are as follows:

- We propose DACL, an innovative hybrid framework that effectively integrates diffusion probabilistic models (DPMs) with supervised contrastive learning to derive noise-robust representations from biosignals. This framework operates within a context-aware scattering latent space derived using the ST [10].
- DACL introduces the application of the forward diffusion process as a principled, learned augmentation strategy. This approach explicitly differentiates our method from generative synthesis techniques by circumventing the high computational demands of the reverse process while achieving superior manifold exploration.
- We establish that DACL serves as an effective “local manifold explorer” through its use of early-stage diffusion. Our framework achieves a patient-level AUROC of 0.9741 on the PhysioNet 2017 benchmark, substantially surpassing both standard contrastive baselines and reconstruction-based methods.

The remainder of the paper is organized as follows: Section 2 provides a review of the related work in the literature. Section 3 presents the details of the proposed methodology. The experiments and results are presented in Section 4. A complete discussion is presented in Section 5. Finally, Section 6 concludes the paper.

## 2 RELATED WORK

This section provides a comprehensive overview of the pertinent literature, emphasizing recent diffusion-based frameworks. We categorize diffusion-based methodologies into three distinct paradigms: *Generative Data Augmentation*, *Reconstruction-based Detection*, and *Diffusion-Guided Representation Learning*.

## 2.1 Generative data augmentation (offline synthesis)

Diffusion Probabilistic Models have been used for the generation of high-quality artificial samples to mitigate data scarcity or class imbalances. For example, Yin et al. [12] introduced the “DiCL” framework for bearing fault diagnosis, utilizing a fault-controllable DPM to generate diverse vibration signals for minority fault classes. In a similar physiological domain, Tan et al. introduced an LDM for Heart Sound Synthesis [13] to address challenges arising from highly imbalanced databases, where normal heart sounds are more prevalent than abnormal ones. While effective, these methods function as offline augmentation strategies, incurring substantial computational costs due to the need to solve the reverse differential equation over numerous time steps ( $T \approx 1000$  steps) to produce each new training sample. In addition, the downstream classifier is trained on synthetic data, which may introduce generative artifacts. In contrast, our proposed framework, DACL, employs a diffusion *forward* process as an online feature augmentation strategy. Rather than synthesizing samples, we train the encoder to be invariant to manifold-consistent noise in a robust latent space, thereby achieving robustness without the inference overhead associated with generation.

## 2.2 Reconstruction-based anomaly detection

Recent studies have utilized the generative prior of DPMs for anomaly detection through reconstruction. The underlying hypothesis is that a diffusion model, when trained solely on non-anomalous data, is incapable of accurately reconstructing anomalous regions. Diffusion-based methodologies have emerged as promising alternatives to conventional generative adversarial networks (GANs) and Autoencoders. Wu et al. [14] introduced a conditional diffusion model designed to reduce false alarms in intensive care unit (ICU) arrhythmia monitoring. Anomalies are detected by calculating the mean squared error (MSE) between the predicted “healthy” waveform and the actual observed signal. Similarly, Zhang et al. [15] proposed “ASDDiffusion” for the detection of industrial acoustic anomalies, effectively utilizing the diffusion model as a high-capacity denoising autoencoder.

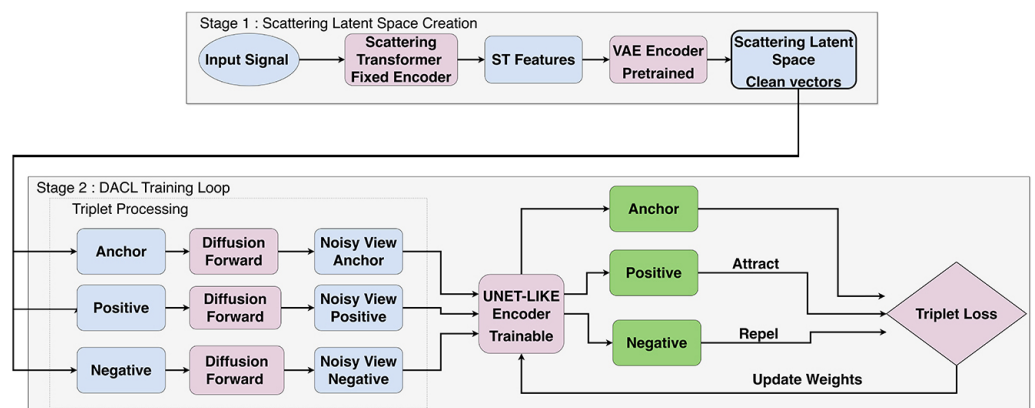


Fig. 1. Conceptual diagram of the DACL framework

Stage I involves training a variational autoencoder (VAE) on all ST features to create a general-purpose latent space. In Stage II, this latent space is used as the main training loop for the model. A triplet of latent vectors is sampled, and noisy views

are generated via the forward diffusion process. These views are passed through a shared U-Net encoder, normalized to the unit hypersphere, and trained via a supervised contrastive objective to produce noise-robust, class-discriminative embeddings.

These methodologies are fundamentally *reconstruction-based*, relying on sample-wise signal fidelity, which may be vulnerable to non-semantic variations (e.g., baseline wander or sensor noise) that are mathematically distinct yet clinically irrelevant to the task at hand. Furthermore, inference requires iterative denoising (with large diffusion time steps ( $T$ ) for reconstruction. In contrast, the DACL framework is *discriminative*; it identifies anomalies based on distance within a learned metric space, using only a single encoder forward pass ( $T = 1$ ) during inference.

### 2.3 Diffusion-guided representation learning

Our research falls within the emerging paradigm of employing diffusion processes for representation learning. Recent investigations have demonstrated that the denoising objective inherently promotes the learning of semantic structures [16, 17]. For instance, Li et al. [18] introduced Contrast DM, which integrates contrastive learning with diffusion for hyperspectral classification, while Zeng et al. [19] developed CLDF, which involves training a pixel decoder on features extracted from a frozen, pre-trained diffusion model. Our proposed framework, DACL, advances this paradigm by incorporating the diffusion forward process directly into the contrastive objective within an efficient, scattering-based latent space. In contrast to CLDF [19], which passively extracts features from a frozen network, DACL trains the encoder *from scratch* against the diffusion trajectory, compelling the encoder to actively learn manifold-invariant representations. Moreover, unlike deep clustering methods that rely solely on reconstruction (e.g., VAE-GAN [20]), our proposed framework enforces both geometric compactness and discriminative separation in a signal processing-guided latent space.

## 3 METHODOLOGY

### 3.1 Framework overview

This section presents the details of the proposed DACL framework, a hybrid framework designed to learn noise-robust representations by leveraging principles from probabilistic diffusion models and supervised contrastive learning in a contextualized scattering latent space. The central tenet of our approach is the replacement of heuristic, handcrafted data augmentations with a principled stochastic transformation derived from the forward diffusion process, operating in an efficient signal processing-guided latent space. As illustrated in Figure 1, the framework operates in two distinct stages: (1) Manifold compression via a VAE on ST features, and (2) Contrastive metric learning driven by diffusion-induced perturbations. The complete procedure for both training and inference is detailed in Algorithm 1.

### 3.2 Latent space construction

To create a seamless and compressed manifold conducive to diffusion operations, we initially extract fixed-size feature vectors utilizing the ST [10]. This step yields contextualized, self-attentive, translation-invariant, and deformation-stable

representations of raw biosignals. Subsequently, a lightweight VAE is trained on these features to map the ST space  $\mathcal{X}$  into a further compressed latent space  $\mathcal{Z}$ . This step is crucial because it ensures that the subsequent diffusion process operates on semantically meaningful latent variables, rather than raw signal noise. For the downstream contrastive phase, we employ the mean of the VAE's learned posterior distribution,  $z_0 = \mu(x)$ , as the “clean” anchor representation.

### 3.3 Diffusion-based latent augmentation

Unlike generative data augmentation methods that require computationally expensive reverse sampling to synthesize new inputs [12], DACL leverages the forward diffusion process as an efficient on-the-fly augmentation strategy in a scattering latent space. For a given clean latent vector  $z_0$ , we generate a “noisy view”  $z_t$  by sampling a time step  $t$  and applying the closed-form Gaussian transition kernel:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \quad (1)$$

Here,  $\bar{\alpha}_t$  defines the noise schedule. This process injects structured, manifold-consistent noise into the representation, forcing the encoder to learn features that are invariant to the diffusion-induced perturbations. The time step  $t$  acts as a control parameter for the augmentation intensity, allowing the model to explore the local neighborhood of the data manifold ( $t \rightarrow 0$ ) or its global structure ( $t \rightarrow T$ ).

### 3.4 UNet-like encoder architecture

The core discriminative model is a UNet-like encoder, denoted as  $\text{Enc}_\theta$ , which is adapted for vector-based input. While this architecture is structurally a conditional multilayer perceptron (MLP) composed of fully connected layers, it inherits the functional inputs of the standard diffusion noise predictor, processing the noisy latent vector  $z_t$  along with the corresponding time step  $t$ . This conditional structure enables the encoder to learn the time-dependent representations of the manifold. To ensure geometric stability within the metric space and mandate features that lie on the unit hypersphere, we strictly enforce L2 Normalization on the output embeddings. For an input view  $z_t$ , the normalized embedding  $e$  is computed as:

$$\mathbf{e}_{\text{raw}} = \text{Enc}_\theta(\mathbf{z}_t, t), \mathbf{e} = \frac{\mathbf{e}_{\text{raw}}}{\|\mathbf{e}_{\text{raw}}\|_2 + \xi} \quad (2)$$

where  $\xi$  is a small constant that ensures numerical stability. This projection ensures that all the embeddings lie on the unit hypersphere.

### 3.5 Combined contrastive objective

The network is optimized using a loss function designed to simultaneously maximize inter-class separability and intra-class compactness. The total loss  $L_{\text{Total}}$  is defined as:

$$L_{\text{Total}} = L_{\text{Triplet}} + \lambda_{\text{Clustering}} L_{\text{KL-Clustering}} \quad (3)$$

The primary component, Triplet Margin Loss, enforces discriminative distance constraints on the normalized embeddings as follows:

$$L_{\text{Triplet}}(\theta) = \mathbb{E} \left[ \max \left( 0, \|\mathbf{e}_A - \mathbf{e}_P\|_2^2 - \|\mathbf{e}_A - \mathbf{e}_N\|_2^2 + m \right) \right] \quad (4)$$

where  $(\mathbf{e}_A, \mathbf{e}_P, \mathbf{e}_N)$  represent the Anchor, Positive, and Negative embeddings. To further regularize the latent manifold, we incorporate the KL clustering loss. This term aligns the clean anchor embedding  $\mathbf{e}_{\text{clean}}$  with a set of  $K$  learnable cluster prototypes  $\mathbf{C} = \{\mathbf{c}_k\}_{k=1}^K$ , enforcing a tighter grouping for the dominant normal class and penalizing outliers.

### 3.6 Inference strategy

The proposed DACL framework offers a notable advantage in terms of inference efficiency. Unlike reconstruction-based anomaly detection methods, which require iterative reverse diffusion during inference ( $T \approx 50 - 1000$  steps) [14], our approach requires only a single forward pass. For a given test sample  $\mathbf{x}_{\text{test}}$ , it is mapped into the latent space  $\mathbf{z}_{0,\text{test}}$  and processed through the encoder in a single forward pass ( $t = 1$ ). The anomaly score is then computed efficiently as the Euclidean distance to the precomputed normal prototype  $\boldsymbol{\mu}_{\text{normal}}$  within the normalized embedding space, as follows:

$$\text{Score} = \|\mathbf{e}_{\text{test}} - \boldsymbol{\mu}_{\text{normal}}\|_2 \quad (5)$$

This strategy effectively leverages the model's learned sensitivity to manifold perturbations while maintaining the speed characteristic of a standard discriminative classifier.

#### Algorithm 1: DACL Training and Inference Procedure

- 1: **Require:** Training features  $\mathcal{X}$ , Labels  $\mathcal{Y}$ ,  $\xi = 1e-6$ . Stability constant
- 2: **Phase 1: Manifold Compression**
- 3: Train VAE on  $\mathcal{X}$ ;
- 4: Extract latent set  $\mathcal{Z}_0 \leftarrow E_{\text{VAE}}(\mathcal{X})$
- 5: Initialize cluster prototypes  $\mathbf{C} = \{\mathbf{c}_k\}_{k=1}^K$  and Encoder  $\theta$ .
- 6: **Phase 2: Diffusion-Augmented Training**
- 7: for each training batch do
- 8:   Sample triplet  $(\mathbf{z}_A, \mathbf{z}_P, \mathbf{z}_N)$  from  $\mathcal{Z}_0$ .
- 9:   Sample timesteps  $t_A, t_P, t_N \sim \mathcal{U}(1, T_{\text{train}})$
- 10:   for  $v \in \{A, P, N\}$  do
- 11:     Generate noisy view  $\mathbf{z}_{t_v}$  via Eq. (1).
- 12:      $\mathbf{e}_{\text{raw},v} \leftarrow \text{Enc}_{\theta}(\mathbf{z}_{t_v}, t_v)$
- 13:      $\mathbf{e}_v \leftarrow \mathbf{e}_{\text{raw},v} / (\|\mathbf{e}_{\text{raw},v}\|_2 + \xi)$
- 14:   end for
- 15: Update  $\theta, \mathbf{C}$  minimizing  $L_{\text{total}}$  (Eq. 3).
- 16: end for
- 17: **Phase 3: Inference (Anomaly Detection)**
- 18: Given test  $\mathbf{x}_{\text{test}}$  project  $\mathbf{z}_{0,\text{test}} \leftarrow E_{\text{VAE}}(\mathbf{x}_{\text{test}})$
- 19:  $\mathbf{e}_{\text{test}} \leftarrow \text{Enc}_{\theta}(\mathbf{z}_{0,\text{test}}, t=1)$
- 20: **return** Score  $\|\mathbf{e}_{\text{test}} - \boldsymbol{\mu}_{\text{normal}}\|_2$

## 4 EXPERIMENTS AND RESULTS

### 4.1 Dataset and preprocessing

The proposed framework was evaluated using the PhysioNet 2017 Challenge Dataset [21], a widely recognized benchmark for ECG rhythm classification. The dataset consists of single-lead ECG recordings categorized into *normal*, *AF*, and *other* rhythms. To frame the problem as a rigorous anomaly detection task, we designated the *normal* recordings as the in-distribution (healthy) class, while the *AF* and *other* rhythms were combined to form a single *anomaly* class. To prevent data leakage and ensure subject independence, the dataset was partitioned by patient ID, with 70% of the patients allocated for training and 30% for testing.

### 4.2 Evaluation protocol and baselines

Performance was reported using the patient-level Area Under the Receiver Operating Characteristic (AUROC), alongside Precision, Recall, and F1-Score. To isolate the contribution of diffusion-based augmentation in the scattering latent space, we compared DACL with three rigorous baselines. To ensure a fair comparison, all contrastive baselines were implemented using the unit hypersphere projection mechanism.

- Contrastive + Gaussian Aug. (Baseline 1): This represents the standard heuristic approach. It is a supervised contrastive model in which “views” are generated by adding isotropic Gaussian noise ( $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ ) to the latent vectors. This directly tests whether *structured* diffusion noise is superior to *random* noise.
- Denoising Autoencoder (DAE) (Baseline 2): This represents the reconstruction-based paradigm. A standard DAE was trained to minimize the MSE between the noisy and clean inputs. This tests whether *discriminative distance* is superior to *sample reconstruction*.
- DACL (Triplet Only): An ablation of our framework that utilizes diffusion augmentation but removes the  $\lambda_{\text{Clustering}}$  term. This specifically isolates the effect of the deep-clustering objective.

### 4.3 Internal benchmarking results

All models were optimized using the Adam optimizer with a learning rate of  $1e-4$  and a batch size of 64. A critical component of the DACL framework is the 32-dimensional scattering latent space. As detailed in Table 1 and illustrated in Figure 2, DACL (Triplet + Clustering) achieved the highest patient-level AUROC of 0.9741 among all comparison baselines.

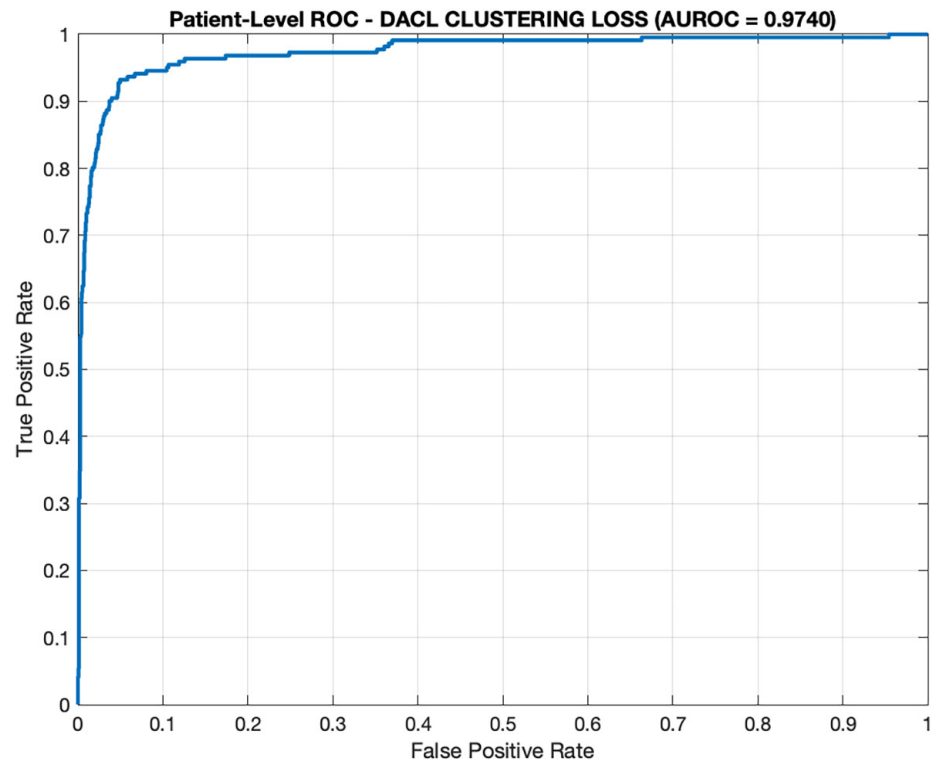
The comparison with the Gaussian baseline is particularly revealing. Whereas Gaussian noise yields a high recall (0.9005), it suffers from significantly lower precision (0.5408). This indicates that the Gaussian model exhibits a bias toward hypersensitivity, erroneously flagging many healthy samples as anomalous. In contrast, DACL achieves a superior balance, boosting the precision to 0.7464 while maintaining high recall. This confirms that structured diffusion-induced perturbations assist the encoder in defining a tighter and more precise decision boundary around the normal manifold. Furthermore, comparing the full DACL framework

with the triplet-only variant highlights the necessity of the deep clustering objective. Although the triplet-only configuration achieved a competitive AUROC (0.9508), its low precision (0.4821) and consequently reduced F1-score (0.6313) indicate an insufficient consolidation of the normal class manifold. Therefore, the integration of the KL-Clustering term is essential for enforcing geometric compactness, serving as the decisive factor in achieving an optimal balance between sensitivity and specificity. The resulting class separation in the latent space is visually corroborated by the t-SNE embedding shown in Figure 3.

**Table 1.** Performance comparison on PhysioNet 2017

Model	AUROC	Precision	Recall	F1-Score	Accuracy
DACL (Triplet + Clustering)	0.9741	0.7464	0.9321	0.8290	0.9544
DACL (Triplet Only)	0.9508	0.4821	0.9140	0.6313	0.8638
Gaussian Aug.	0.9513	0.5408	0.9005	0.6757	0.8898
Denosing AE	0.7675	0.2767	0.6561	0.3893	0.7374

*Notes:* DACL achieves the optimal trade-off between sensitivity and specificity, significantly outperforming the reconstruction-based DAE and the heuristic Gaussian baseline.



**Fig. 2.** Patient-Level ROC Curve for the DACL model utilizing the Combined Contrastive Loss on PhysioNet 2017

The curve demonstrates robust performance with an AUROC of 0.9741, confirming high sensitivity at low false-positive rates.

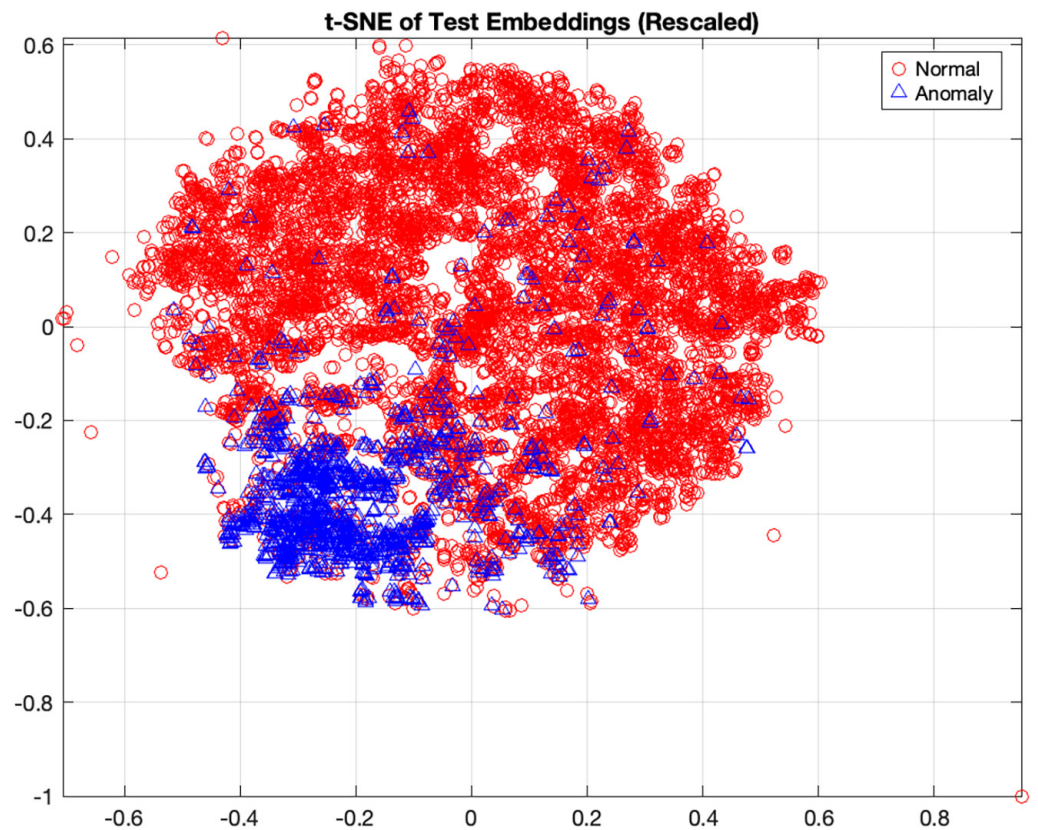


Fig. 3. t-SNE visualization of the test embeddings generated by the DACL encoder

The clusters show markedly improved separation compared to baselines, with the Normal class (Red) forming a compact, well-defined manifold distinct from anomalies.

#### 4.4 Comparison with state-of-the-art frameworks

To thoroughly assess the proposed framework (DACL), we benchmarked it against established external baselines, categorizing them according to their learning paradigms (refer to Table 2).

**Self-supervised vs. Supervised SOTA.** A fundamental objective of representation learning is to achieve parity with the performance of fully supervised models. Notably, DACL (0.9741 AUROC) substantially surpasses the attention-based supervised model MINA [22] (0.9488) and effectively aligns with the performance of the state-of-the-art supervised CWT-MB-ResNet [23] (0.9761). This finding illustrates that manifold-aware diffusion augmentation enables the encoder to capture highly discriminative pathological features, thereby competing effectively with supervised deep learning approaches.

**Efficiency vs. Diffusion Baselines.** The proposed framework DACL introduces an innovative application of the diffusion paradigm by leveraging its capabilities exclusively for representation learning within an efficient scattering latent space while preserving the simplicity of single-pass inference. Although recent diffusion-based methods, such as Wu et al. [14] (reconstruction) and DiCL [12] (synthetic data),

demonstrate potential, they are hindered by high inference latency due to the need for iterative reverse sampling ( $T \approx 200 - 1000$ ). By employing the diffusion forward process for metric learning rather than for generation or reconstruction, the DACL framework operates with a single encoder forward pass ( $T = 1$ ). This results in a speedup factor of approximately  $T$ , making it particularly suitable for real-time clinical monitoring, where low latency is essential.

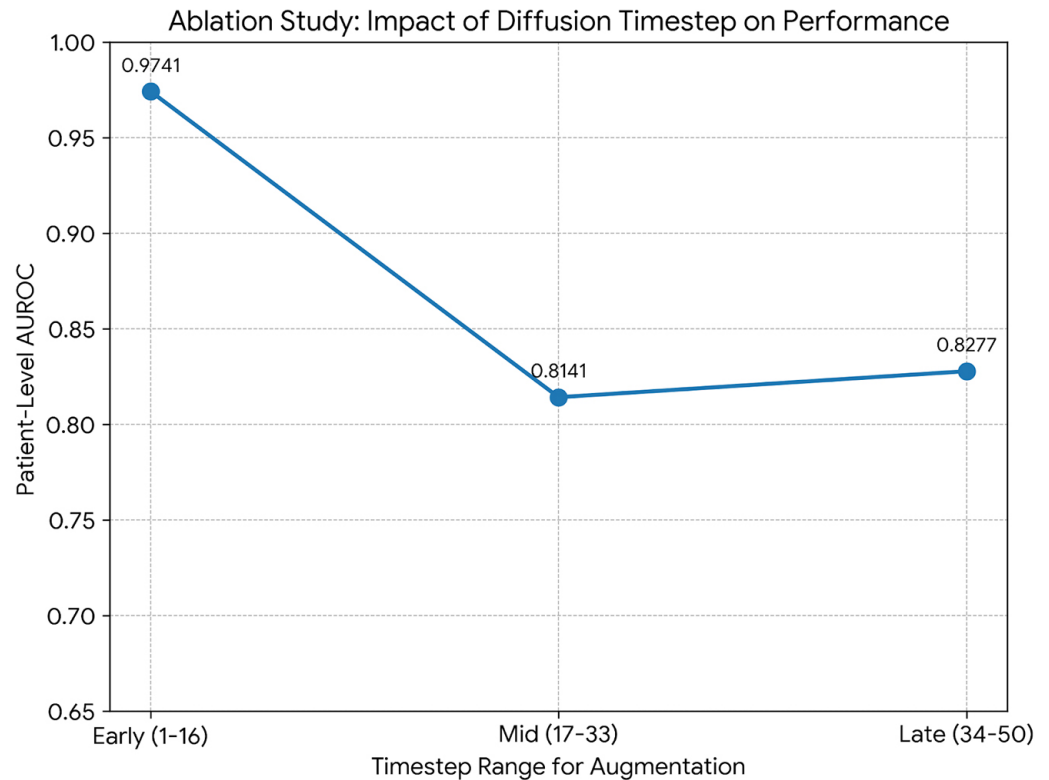
**Table 2.** Quantitative comparison

Method	Paradigm	Target Domain	Inference Mechanism	Cost ( $T$ )	AUROC	Acc/F1
<i>Independent Supervised SOTA (PhysioNet 2017)</i>						
MINA [22]	Supervised (Attn)	ECG	Attention Network	1	0.9488	–
CWT-MB-ResNet [23]	Supervised (CNN)	ECG	Multi-Branch ResNet	1	0.9761	–
<i>Direct Comparison (This Work)</i>						
Gaussian Baseline	Contrastive	ECG	Encoder Forward Pass	1	0.9513	88.98% (Acc)
Denosing AE	Reconstruction	ECG	Encoder-Decoder	1	0.7675	–
DACL (Ours)	Latent Aug.	ECG	Encoder Forward Pass	1	0.9741	95.44% (Acc)
<i>External Reference</i>						
Wu et al. [14]	Diffusion-Reconstruction	ICU Vitals	Iterative Denosing	$\approx 200$	–	91.71% (Acc)
DiCL [12]	Diffusion-Synthesis	Vibration	Reverse Sampling	$\approx 1000$	–	99.82% (Acc)
HeartSound-LDM [13]	Diffusion-Synthesis	Heart Sounds	Reverse Sampling	$\approx 1000$	–	93.4% (Acc)
TS-TCC [5]	Heuristic Contrastive	EEG	Encoder Forward Pass	1	–	83.0% (Acc)

DACL matches the performance of fully supervised state-of-the-art methods (MINA, CWT-MB-ResNet) while operating as a self-supervised anomaly detector. Furthermore, it maintains an inference cost of  $T = 1$ , offering a significant speedup over reconstruction-based diffusion methods, which require  $T \gg 1$ .

#### 4.5 Ablation study: Manifold exploration

To examine the influence of noise intensity on the representation quality, we trained the DACL framework using restricted diffusion time step ranges (see Figure 4). The findings indicate a distinct optimal range for manifold exploration in the study. Performance reaches its peak during the early-stage diffusion ( $t \in [1,16]$ ), achieving a maximum AUROC of 0.9741. This suggests that minor-structured perturbations enable the encoder to effectively explore the local neighborhood of the scattering manifold without compromising its semantic identity. Conversely, performance deteriorates markedly during the Mid ( $t \in [17,33]$ ) and Late ( $t \in [34,50]$ ) stages. This observation confirms that excessive diffusion noise ultimately degrades the fine-grained pathological features essential for anomaly detection, thereby displacing the data from the semantic scattering manifold.



**Fig. 4.** Results from the diffusion timestep ablation study

Performance clearly peaks during early timesteps ( $t \leq 16$ ) and degrades as noise intensity increases, confirming that local manifold exploration is superior to heavy corruption for this task.

## 5 DISCUSSION

This study introduces DACL, a framework that utilizes the forward diffusion process within a ST latent space as a systematic and learnable augmentation strategy for contrastive physiological representation learning. The empirical success of this method, demonstrated by an outstanding AUROC of 0.9741, provides several critical insights into the nature of self-supervised learning for bio signals.

### 5.1 Precision and manifold topology

The pronounced performance disparity between DACL and the Gaussian baseline underscores a fundamental limitation inherent in heuristic augmentations. Notably, the Gaussian model's commendable AUROC (0.9513) confirms the inherent robustness of the underlying scattering latent space. However, its low precision (0.5408) suggests a "hypersensitive" decision boundary. The introduction of isotropic noise results in the Gaussian model uniformly expanding the rejection zone in all directions, thereby erroneously identifying valid but rare physiological variations as anomalies. In contrast, DACL achieved markedly higher precision (0.7464). This improvement is attributed to the manifold-aware characteristics

of the diffusion noise. Given that the diffusion kernel is conditioned on the data distribution, the generated “views” remain semantically valid. Consequently, the encoder can delineate a decision boundary that precisely aligns with the true physiological manifold, effectively rejecting anomalies without misclassifying healthy edge cases.

## 5.2 Impact of the KL-clustering objective

The ablation study indicated that the triplet-only loss function encountered difficulties in achieving high precision ( $\approx 0.48$ ). The incorporation of the KL-clustering loss was crucial for stabilizing the metric space. Even with the unit hypersphere constraint applied, the triplet loss separates negative samples but does not necessarily ensure compactness within the normal class. The clustering objective serves as a significant geometric regularizer, compelling normal data to converge into a high-density region centered on the prototype  $\mu_{\text{normal}}$ . This density convergence enhances the reliability of the Euclidean distance metric utilized for inference.

## 5.3 Discriminative distance vs. Sample-level reconstruction

The suboptimal performance of the Denoising Autoencoder (AUROC 0.7675) underscores a significant observation: *sample-level reconstruction is often inadequate for semantic anomaly detection in complex biosignals*. Standard DAEs prioritize low-level signal fidelity and frequently expend capacity reconstructing non-essential high-frequency noise or baseline wander. Conversely, the proposed DACL framework learns *discriminative distances* within a scattering latent space by optimizing feature invariance rather than signal reconstruction. Consequently, it is more adept at disregarding clinically irrelevant artifacts and concentrating exclusively on the structural deviations that characterize arrhythmia.

## 6 CONCLUSION

This study introduces DACL, an innovative methodology that effectively integrates generative diffusion models with discriminative representation learning. The framework utilizes the forward diffusion process as a structured, manifold-consistent augmentation technique within a context-aware scattering latent space. Our empirical evaluation of the PhysioNet 2017 benchmark indicates that DACL achieves detection performance comparable to leading supervised models (AUROC 0.9741) while functioning as a self-supervised anomaly detector. Unlike generative baselines that require computationally intensive reverse sampling, DACL operates with a single encoder forward pass, rendering it highly suitable for real-time clinical applications. Our analysis of the noise trajectory reveals that early-stage diffusion serves as an optimal local manifold explorer, enhancing model robustness without compromising semantic integrity. These findings highlight the promising potential of signal processing-guided LDMs in biomedical AI. This is particularly relevant in resource-limited applications, such as AI-powered mobile healthcare systems, which are currently receiving increased attention [24, 25].

## 7 ACKNOWLEDGEMENTS

The author(s) acknowledge the use of Paperpal for grammar correction and stylistic refinement during the preparation of this manuscript. The author(s) have reviewed and edited the content and assume full responsibility for the final publication.

## 8 REFERENCES

- [1] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 8, no. 2, p. 36, 2020. <https://doi.org/10.3390/technologies9010002>
- [2] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, no. 149, 2020, pp. 1597–1607.
- [3] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 9726–9735. <https://doi.org/10.1109/CVPR42600.2020.00975>
- [4] T. T. Um *et al.*, "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the ACM ICMI*, 2017, pp. 216–220. <https://doi.org/10.1145/3136755.3136817>
- [5] E. Eldele *et al.*, "Time-series representation learning via temporal and contextual contrasting," in *Proceedings of the IJCAI*, 2021, pp. 2352–2359. <https://doi.org/10.24963/ijcai.2021/324>
- [6] Z. Yue *et al.*, "TS2Vec: Towards universal representation of time series," *AAAI*, vol. 36, no. 8, pp. 8980–8987, 2022. <https://doi.org/10.1609/aaai.v36i8.20881>
- [7] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, no. 574, 2020, pp. 6840–6851.
- [8] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the International Conference on Machine Learning*, vol. 37, 2015, pp. 2256–2265.
- [9] P. Vincent, H. Larochelle, P.-A. Manzagol, and Y. Bengio, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the ICML*, 2008, pp. 1096–1103. <https://doi.org/10.1145/1390156.1390294>
- [10] R. Zewail, "Scattering transformer: A training-free transformer architecture for heart murmur detection," *arXiv preprint arXiv:2509.18424*, 2025.
- [11] R. Rombach, D. Lorenz, A. Blattmann, P. Esser, and B. Ommer, "High resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10674–10685. <https://doi.org/10.1109/CVPR52688.2022.01042>
- [12] K. Yin, L. Yang, Y. Zhang, F. Ou, C. Chen, and B. Mu, "Diffusion augmented contrastive learning framework for quantitative diagnosis under limited data conditions," *Advanced Engineering Informatics*, vol. 69, p. 103930, 2026. <https://doi.org/10.1016/j.aei.2025.103930>
- [13] Y. Tan, H. Zhang, Y. Chang, Q. Wu, K. Qian, and B. Hu, "A latent diffusion model for heart sound synthesis," in *Proceedings of the 4th International Symposium on Computer Applications and Information Technology (ISCAIT)*, 2025, pp. 1314–1319. <https://doi.org/10.1109/ISCAIT64916.2025.11010636>

- [14] F. Wu, G. Zhao, X. Qian, and L.-W. H. Lehman, "A diffusion model with contrastive learning for ICU false arrhythmia alarm reduction," in *Proceedings of the IJCAI*, 2023, pp. 4912–4920. <https://doi.org/10.24963/ijcai.2023/546>
- [15] F. Zhang, X. Xie, and K. Guo, "ASD-diffusion: Anomalous sound detection with diffusion models," *arXiv preprint arXiv:2409.15957*, 2024. [https://doi.org/10.1007/978-3-031-78119-3\\_24](https://doi.org/10.1007/978-3-031-78119-3_24)
- [16] G. Luo, L. Dunlap, D. H. Park, A. Holynski, and T. Darrell, "Diffusion hyper features: Searching through time and space for semantic correspondence," in *Proceedings of the 37th International Conference Neural Information Processing Systems*, no. 2057, 2023, pp. 47500–47510.
- [17] Q. Song, J. Hu, L. Xiao, B. Sun, X. Gao, and S. Li, "DiffCL: A diffusion-based contrastive learning framework with semantic alignment for multimodal recommendations," *arXiv preprint arXiv:2501.01066*, 2025. <https://doi.org/10.1109/TNNLS.2025.3583509>
- [18] J. Li, J. Wang, and Z. Cao, "Contrastdm: Combining contrastive learning and diffusion model for hyperspectral image classification," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, 2024, pp. 9966–9970. <https://doi.org/10.1109/IGARSS53475.2024.10641234>
- [19] D. Zeng, X. Hu, Y.-J. Chen, Y. Wu, X. Xu, and Y. Shi, "Contrastive learning with diffusion features for weakly supervised medical image segmentation," *arXiv preprint arXiv:2506.23460*, 2025. <https://doi.org/10.48550/arxiv.2506.23460>
- [20] Z. Xing, W. A. P. Smith, and O. Mehmood, "Unsupervised anomaly detection with a temporal continuation, confidence-aware VAE-GAN," *Pattern Recognition*, vol. 166, p. 111699, 2025. <https://doi.org/10.1016/j.patcog.2025.111699>
- [21] G. D. Clifford *et al.*, "AF classification from a short single lead ECG recording: The PhysioNet/ computing in cardiology challenge 2017," in *Proceedings of the Computing in Cardiology (CinC)*, Rennes, France, 2017, pp. 1–4. <https://doi.org/10.22489/CinC.2017.065-469>
- [22] S. Hong, H. Li, J. Sun, C. Xiao, and T. Ma, "MINA: Multilevel knowledge guided attention for modeling electrocardiography signals," in *Proceedings of the IJCAI*, 2019, pp. 5888–5894. <https://doi.org/10.24963/ijcai.2019/816>
- [23] J. Xie, B. Yao, and S. Stavrakis, "Automated identification of atrial fibrillation from single-lead ECGs using multi-branching ResNet," *Frontiers in Physiology*, vol. 15, 2024. <https://doi.org/10.3389/fphys.2024.1362185>
- [24] C. B. Chandrakala, S. Pooja, C. Pujari, S. Katavarapu, S. Awatramani, and S. Gohil, "Health-lens: A health diagnosis companion," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 19, no. 12, pp. 68–102, 2025. <https://doi.org/10.3991/ijim.v19i12.51525>
- [25] R. Jeya, H. A. Karim, and S. B. Mansor, "Artificial intelligence and mobile apps support intelligent healthcare systems for mental health services," *International Journal of Interactive Mobile Technologies (ijIM)*, vol. 18, no. 20, pp. 157–168, 2024. <https://doi.org/10.3991/ijim.v18i20.50743>

## 9 AUTHOR

**Rami Zewail** is affiliated with the Department of Computer Science and Engineering at Egypt–Japan University of Science and Technology, Alexandria, Egypt. He holds a Ph.D. degree from the University of Alberta, Canada. Research interests in the fields of resource-aware artificial intelligence, computer vision, and embedded intelligence. He has over 15 years of academic and industrial research and development experience and serves as a reviewer for several scientific journals (E-mail: [rami.zewail@ejust.edu.eg](mailto:rami.zewail@ejust.edu.eg)).