Continuous-Time Neural Ordinary Differential Equations Improve Latent Space Structure in Electrocardiogram Representation Learning

Tyme Chatupanyachotikul[1] , Lukas P.A. Arts[2] , Samuel Ruiperez-Campillo[3,4] , Wouter Huberts[1,5] , Fleur V.Y. Tjong[2] , Sean Benson[1,2]
[1] University of Amsterdam, Amsterdam, the Netherlands
[2] Amsterdam University Medical Center, Amsterdam, the Netherlands
[3] Department of Computer Science, ETH Zurich, Zurich, Switzerland
[4] Institute of Medical Engineering and Sci., Massachusetts Institute of Technology, Cambridge, USA
[5] Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

Abstract

Existing Electrocardiogram (ECG) representation learning approaches treat the signal as an arbitrary time series, ignoring that ECG morphology is the surface projection of continuous cardiac electrophysiological dynamics. We propose training separate Modulated Neural Ordinary Differential Equation Variational Autoencoders (MoNODE-VAE) on segmented P-waves and QRST complexes, imposing a continuous-time inductive bias that mirrors the generative structure of cardiac electrophysiology. We argue that the Neural Ordinary Differential Equation vector field constrains the latent space during training, pushing clinically distinct classes toward well-separated regions. Evaluated on MedalCare-XL via linear probing, K-means clustering, and latent trajectory visualisation, MoNODE-VAE consistently outperforms a Gated Recurrent Unit-VAE baseline, yielding more structured and clinically coherent latent representations.

1. Introduction

Electrocardiograms (ECGs) provide non-invasive and cost-effective assessment of cardiac function, encoding rich information on the underlying electrophysiology of the heart [1]. Recent advances in deep learning have enabled data-driven extraction of these cardiac features, with representation learning approaches demonstrating strong performance across a range of downstream cardiovascular tasks [2]. However, existing approaches [3, 4] largely treat ECGs as an arbitrary time series to be compressed into a fixed latent vector. This framing ignores a fundamental property of the signal: the ECG is the surface projection of an underlying continuous dynamical system governed by cardiac electrophysiology. A model with no inductive bias toward this structure has no principled reason to learn representations that follow it. This view is consistent with recent work showing that matching the inductive bias of a model to the underlying structure of cardiac data yields more meaningful representations [5].

Neural Ordinary Differential Equations (NODE) have previously been applied to ECG data in limited contexts. Prior work has explored their use for ECG synthesis [6] and for evaluating latent representation robustness across different sampling frequencies [7], leveraging the continuous-time formulation to handle irregular time series. However, to our knowledge, no prior work has examined the geometric structure of the latent space induced by the NODE constraint, nor whether this structure corresponds to clinically meaningful organisation.

In this work, we propose a representation learning framework that addresses this limitations. We decompose the ECG into its anatomical segments and train separate Modulated Neural Ordinary Differential Equation Variational Autoencoders (MoNODE-VAE) [8] on the P-wave and QRST complex independently. The NODE component imposes a continuous-time inductive bias: rather than compressing a segment into an unstructured fixed vector, the model is required to find an initial latent state z_0 and a learned vector field $f_\theta(z)$ such that integrating forward from z_0 reproduces the observed waveform. A key geometric consequence is that the vector field creates pressure for clinically distinct classes to separate: if samples with different pathologies occupy nearby latent regions, their reconstructions would converge, incurring a large reconstruction loss, forcing well-separated cluster structure to emerge without explicit supervision. We evaluate this property directly via linear probing on pathology classification, K-means clustering with Adjusted Rand Index (ARI) and Silhouette Score on Principal Component Analysis (PCA) reduced latent representations, and latent trajectory visualisation.

Our results demonstrate two key findings: MoNODE-VAE representations outperform Gated Recurrent Unit (GRU)-VAE representations on downstream classification, and MoNODE latent spaces exhibit substantially better-separated, clinically meaningful cluster structure, validating that the continuous-time inductive bias produces discriminative representations.

2. Methods

2.1. Datasets

MedalCare-XL [9] consists of simulated 12-lead ECGs generated from electrophysiological models, where P-waves and QRST complexes are synthesised independently from separate atrial and ventricular anatomical models and combined via statistical sampling across 13 torso models. Seven pathological classes are generated per torso model, each belonging to either atrial or ventricular pathology; crucially, ventricular pathology cases use healthy atrial simulation parameters for the P-wave, and vice versa. The atrial pathological classes include Atrioventricular block (AVBlock), Fibrotic Atrial Cardiomyopathy (FAM), Left Atrial Enlargement (LAE), Complete Interatrial Conduction Block (IAB). The ventricular pathological classes consists of Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB), and Myocardial Infarction (MI), where MI is further divided into 6 subclasses pertaining to the three predominant arteries of Right-Anterior Descending (RAD), Left Anterior Descending (LAD), and Left Circumflex (LCX).

2.2. Pre-processing

P-wave and QRST segments were extracted from the median beat via delineation and averaging, minimising intra-beat noise relative to individual beat sampling. Anomalous segmentations were discarded based on a duration threshold ($p < 0.05$). Each lead is preprocessed by applying a 4th-order Butterworth low-pass filter (cutoff 40 Hz), followed by baseline wander correction via successive median filters of 200 ms and 600 ms and per-lead normalisation. For MedalCare-XL, healthy samples were resampled to at most twice the size of the largest pathological class to prevent class imbalance. The MedalCare-XL split for the P-wave consists of 7424/1036/1022, and 8704/1987/1979 for the QRST complex.

2.3. Baseline

The baseline is a GRU-VAE, in which a GRU encoder maps the input segment to the parameters of a variational posterior over the initial latent state, $z_0 \sim q(z_0 \mid x_0, \ldots, x_N)$. A GRU decoder then evolves this latent state sequentially, with a Multilayer Perceptron (MLP) projecting each hidden state to the reconstructed ECG at each timestep.

2.4. Model Architecture

The proposed model is a MoNODE-VAE [8], in which the latent dynamics are governed by a continuous-time ODE rather than discrete GRU transitions. The GRU encoder produces the initial latent state z_0, which represents the initial hidden cardiac state. A second encoder produces a static modulator variable d, capturing patient-specific physiological parameters that condition but do not evolve with the dynamics. The latent trajectory is then obtained by integrating the learned vector field $f_\theta(z; d)$ forward in time:

$$z_i = z_0 + \int_{t_0}^{t_i} f_\theta(z(\tau);\, d)\, d\tau \tag{1}$$

where f_θ is parameterised by a MLP. Each point along the trajectory is projected to the reconstructed ECG via a shared MLP decoder. This formulation mirrors the generative structure of cardiac electrophysiology: z_0 corresponds to the initial cardiac state, f_θ approximates the governing electrophysiological transition rules, and the decoder acts as a projection from latent cardiac dynamics to surface ECG voltage. The decoupling of the time-varying trajectory $z(t)$ from the static modulator d further allows the model to separate dynamic depolarisation/repolarisation morphology from fixed patient-specific parameters.

2.5. Training

Both models are trained by minimising the evidence lower bound (ELBO) [10], comprising a reconstruction term and a Kullback–Leibler divergence penalty that regularises the posterior toward a standard normal prior:

$$\begin{aligned}\mathcal{L}_{\text{VAE}}(\theta, \phi; \mathbf{x}) = &- \mathbf{E}_{\mathbf{q}_\phi(\mathbf{z}|\mathbf{x})}[\log \mathbf{p}_\theta(\mathbf{x}|\mathbf{z})] \\ &+ D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel \mathbf{p}(\mathbf{z}))\end{aligned} \tag{2}$$

The Adam optimiser was used with a learning rate of 0.002 across all runs. The NODE is solved using the Runge-Kutta method. Variable-length segments are zero-padded to the maximum length within each batch. Models were trained for 50 epochs (batch size 64) on MedalCare-XL, with the best checkpoint selected by validation loss. To ensure a fair comparison, the total latent dimensionality is matched across both models at 64 dimensions: the GRU-VAE uses a single 64-dimensional initial state z_0, while MoNODE-VAE allocates 32 dimensions to z_0 and 32 to d.

2.6. Evaluation

Representation quality is assessed through three complementary evaluations. First, a linear classifier is trained

Table 1: Pathology classification results on MedalCare-XL for atrial and ventricular models respectively.

Model	Accuracy	F1	ROC-AUC
Atrial			
GRU-VAE	0.773	0.776	0.959
MoNODE-VAE	**0.804**	**0.808**	**0.967**
Ventricular			
GRU-VAE	0.660	0.657	0.956
MoNODE-VAE	**0.706**	**0.705**	**0.972**

on frozen latents for pathology classification. The use of linear probes directly measures the degree to which clinically meaningful information is linearly accessible in the latent space, without confounding from nonlinear decoder capacity. Classification performance was evaluated using accuracy, F1, and Receiver Operating Characteristic Area Under the Curve (ROC-AUC).

Second, to assess the geometric structure of the learned latent space, K-means clustering is applied to PCA-reduced latent representations (10 components), with the number of clusters set to the number of ground-truth pathological classes. Cluster quality is measured using the Adjusted Rand Index (ARI), which quantifies agreement between the inferred clusters and true clinical labels, and the Silhouette Score, which measures intra-cluster cohesion relative to inter-cluster separation. Together these metrics assess whether the model has organised its latent space into well-separated, clinically coherent regions without access to label supervision during training

Third, latent trajectory visualisation is performed by projecting the learned latent trajectories onto the first two principal components and plotting the mean trajectory per class. This provides a qualitative assessment of the smoothness and separability of class-specific dynamics in the latent space, and directly illustrates the geometric consequence of the NODE's vector field constraint.

3. Results and Discussion

3.1. Pathology Classification

The results of training a linear classifier on the MedalCare-XL pathology classification task are shown in Table 1. Note that for the atrial models, P-waves from ventricular pathology cases are labelled as healthy, and for the ventricular models, QRST complexes from atrial pathology cases are labelled as healthy, consistent with how the simulated data were generated. The improved performance of MoNODE-VAE over the GRU-VAE baseline across both atrial and ventricular models indicates that the ODE-structured latents encode more discriminative morphological information.

Table 2: Clustering metrics (ARI and Silhouette Score) for K-means applied to PCA-reduced (10 components) latent representations, for atrial and ventricular models.

Model	ARI	Silhouette
Atrial		
GRU-VAE	0.028	0.100
MoNODE-VAE	**0.038**	**0.158**
Ventricular		
GRU-VAE	0.186	0.102
MoNODE-VAE	**0.563**	**0.176**

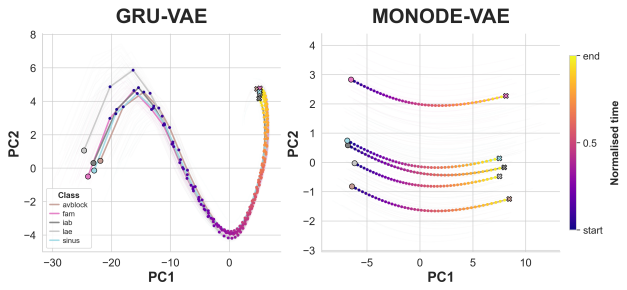
3.2. Latent Space Structure

Table 2 reports ARI and Silhouette Scores from K-means clustering on PCA-reduced latent representations (10 components). For ventricular models, MoNODE-VAE achieves substantially higher ARI and Silhouette Score than GRU-VAE, indicating that the latent space is organised into well-separated clusters that correspond closely to ground-truth clinical labels. Gains for atrial models are more modest, potentially due to the smaller morphological differences of ECG between atrial pathology classes.

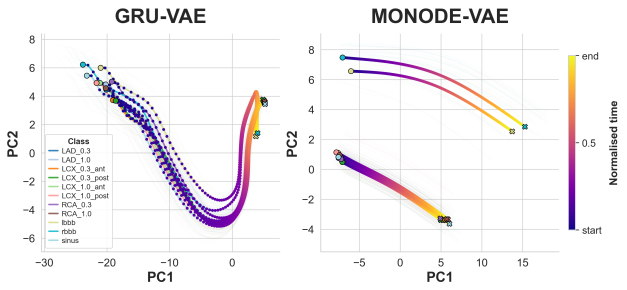
Figure 1 shows the mean PCA projected latent trajectories per class. For ventricular MoNODE-VAE, MI subclasses are grouped together and separated from LBBB and RBBB, though the sinus class lies in close proximity to MI. For atrial MoNODE-VAE, classes follow visually distinct trajectories with clearer separation throughout. In both cases, MoNODE trajectories are smoother than GRU-VAE, consistent with the vector field constraint: the ODE must generate a globally consistent flow, so latent trajectories cannot vary drastically between timesteps. By contrast, GRU-VAE trajectories follow similarly shaped paths regardless of class, indicating that the discrete hidden state transitions impose no such organising pressure on the latent space.

4. Discussion and Conclusion

The results demonstrate two consistent findings. First, MoNODE-VAE outperforms the GRU-VAE baseline on pathology classification across both atrial and ventricular models. Second, the MoNODE-VAE latent space exhibits substantially better geometric structure: clusters are more cohesive, better separated, and align more closely with ground-truth labels, as evidenced by both ARI and Silhouette Score. The trajectory visualisations corroborate this quantitatively: MoNODE-VAE produces smooth, class-specific paths through latent space, whereas GRU-VAE trajectories are less organised, and tend to cluster in similar regions.



(a) Atrial Latent Trajectory. GRU-VAE vs. MoNODE-VAE.



(b) Ventricular Latent Trajectory. GRU-VAE vs. MoNODE-VAE.

Figure 1: Comparison of learned latent trajectories between the baseline GRU-VAE and the proposed MoNODE-VAE. The left panel (a) highlights the atrial representations, while the right panel (b) illustrates the ventricular representations.

Both findings are attributable to the continuous-time inductive bias of the NODE. Because the latent trajectory must follow a single learned vector field, samples whose latent representations overlap would be forced toward the same reconstruction, incurring a large reconstruction loss. This creates implicit pressure during training to push clinically distinct classes into well separated regions of latent space, a consequence that is not present in the GRU-VAE, whose hidden state transitions are unconstrained. The stronger effect observed for ventricular models is consistent with the larger morphological differences between ventricular pathology classes relative to atrial ones.

Moreover, modelling ECG as trajectories of a continuous dynamical system aligns more closely with the underlying cardiac electrophysiology generative process. This physiological alignment may itself be a contributing factor to the improved latent organisation observed, as a model whose inductive bias matches the true data generating process has a more principled basis for learning representations that reflect clinically meaningful structure [5].

Several limitations should be acknowledged. NODE training is considerably more computationally expensive than standard VAE training due to iterative ODE solvers. Furthermore, all evaluations were conducted solely on the MedalCare-XL simulated dataset. Evaluation on real clinical ECGs would provide stronger evidence of generalisation and determine whether the observed benefits of the continuous-time inductive bias extend to the noise and variability inherent in real-world recordings.

Acknowledgments

This research has been supported by grants from the Dutch Heart Foundation (Dekker 03-005-2025-0129) Dutch Research Council (Rubicon 452019308, AINed XS NGF.1609.243.045) and Amsterdam Cardiovascular Sciences, Health Holland. Computational resources have been provided by the University of Amsterdam and SURF. This work has been supported by the Accelerate Programme of the UvA Data Science Centre.

References

[1] Razavipour F, Hossein Nikoo M. Numerical estimation of ejection fraction from 12-lead ECG. Journal of Bioengineering Biomedical Science 2015;06(01). ISSN 21559538.

[2] Muzammil MA, et al. Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases. Journal of Electrocardiology January 2024;83:30–40.

[3] Li J, et al. An electrocardiogram foundation model built on over 10 million recordings with external evaluation across multiple domains. NEJM AI October 2025;.

[4] McKeen K, Masood S, Toma A, Rubin B, Wang B. ECG-FM: An open electrocardiogram foundation model, 2025.

[5] Wiers J, et al. Geometry-aware cardiac MRI representation learning with equivariant neural fields. In Medical Imaging with Deep Learning. 2026; .

[6] Habiba M, et al. Ecg synthesis with neural ode and gan models. In ICECET. 2021; .

[7] Yan A, Sampson ML, Melchior P. A novel approach to classification of ecg arrhythmia types with latent odes. In NeurIPS. 2025; .

[8] Auzina IA, Çağatay Yıldız, Magliacane S, Bethge M, Gavves E. Modulated neural ODEs, 2023.

[9] Gillette K, et al. MedalCare-XL: 16,900 healthy and pathological synthetic 12 lead ECGs from electrophysiological simulations. Scientific Data August 2023;10(1):531. ISSN 2052-4463.

[10] Kingma DP, Welling M. An introduction to Variational Autoencoders. Foundations and Trends® in Machine Learning November 2019;12(4):307–392. ISSN 1935-8245.

Address for correspondence:

Lukas P.A. Arts
Department of Clinical and Experimental Cardiology, Amsterdam University Medical Center
Meibergdreef 9, 1105 AZ Amsterdam, NL
research@lukas-arts.com