Synthetic EGM Generation with Variational Autoencoders

Miriam G Fernández[1,3], Karen L Linares[3,4], Carlos F Santos[2],
María S Guillem[2], Andreu M Climent[2], Óscar B Pérez[1]

[1]Universidad Rey Juan Carlos, [2]ITACA Institute, Universitat Politècnica de València, Spain
[3]Vicomtech, Basque Research and Technology Alliance, [4]eHealth Group, Biogipuzkoa Institute

Abstract

Atrial fibrillation (AF) is the most common sustained arrhythmia, and its management requires accurate characterization of atrial electrical activity. Electrocardiographic imaging (ECGI) and deep learning (DL) methods aiming at estimating electrograms (EGMs) noninvasively from body surface potentials (BSPMs) are promising, but progress is limited by the scarcity of paired BSPM–EGM datasets. Thus, we explore variational autoencoders (VAE) to generate synthetic EGMs training two models: a sinus-only VAE (VAE-S) and a class-conditioned VAE (VAE-C) trained on sinus rhythm and AF. Both models were evaluated through morphological, spectral, and distributional similarity metrics. VAE-S achieved higher fidelity with real signals, while VAE-C enabled rhythm-specific generation at the cost of sinus quality. As a proof of concept, we also assessed data augmentation in a downstream task: noninvasive EGM estimation, where moderate inclusion of synthetic signals improved performance. These results demonstrate that VAEs can generate physiologically plausible atrial EGMs, offering a promising tool to alleviate data scarcity in noninvasive EGM estimation.

1. Introduction

Atrial fibrillation (AF), the most common sustained arrhythmia, affects over 33 million people worldwide and increases risks of stroke, heart failure, and mortality [1, 2]. Current treatment uses invasive mapping to localize arrhythmogenic substrates, usually near pulmonary vein ostia or regions of slowed conduction and reentry [3–5], but remains limited by sparse sampling, long procedures, and poor capture of transient or epicardial activity [6].

These challenges have motivated noninvasive strategies such as electrocardiographic imaging (ECGI), which reconstructs epicardial electrograms (EGMs) from body surface potentials (BSPMs) and anatomical models by solving an ill-posed inverse problem [7]. More recently, deep learning (DL) has been proposed to map BSPMs directly to EGMs without explicit inverse formulations, though progress remains limited by scarce paired BSPM–EGM datasets [8].

In this study, we propose a self-supervised framework for synthetic EGM generation using variational autoencoders (VAEs) for the first time, motivated by the use of these models for other type of signal synthesis [9, 10]. Our contributions are: (i) learning morphology-preserving latent representations using two convolutional β-VAE with perceptual loss and annealed regularization, and (ii) generating synthetic EGMs for data augmentation in downstream noninvasive DL-based EGM estimation tasks under sinus and AF conditions. The manuscript is organized as follows: Section 2 presents methods, Section 3 reports results, and Section 4 provides conclusions.

2. Methods

This section outlines the datasets, VAE architectures, training procedures, and evaluation strategies used to generate synthetic EGMs and assess their impact on noninvasive reconstruction.

2.1. Dataset

Two datasets of intracardiac EGMs were used for this study. Throughout the manuscript, we refer to real EGMs as computational simulations of atrial activity obtained from biophysically detailed electrophysiology models [11]. All real EGMs were simulated at a sampling rate of 500 Hz from 2048 atrial sites and represent 2–4 seconds of activity. Further details on the electrophysiological and torso models are available in [11]. Synthetic EGMs, in contrast, denote signals generated by our VAE models.

Dataset A - sinus only: This dataset consists of simulated EGMs from 19 sinus patients, derived from realistic atrial electrophysiology models. It was used to train the VAE model for synthetic sinus signal generation.

Dataset B – sinus + AF: This dataset comprised EGMs from the same 19 sinus patients plus 33 AF patients. It is utilized for training a class-conditioned VAE to generate both sinus and AF EGMs.

2.2. Generative VAE for EGM Synthesis

Two different VAE architectures were explored for EGM generation: one trained only on sinus rhythm data (VAE-S) and one conditioned on rhythm class (VAE-C).

2.2.1. Generation Models and Architecture

VAE-S was designed to model the distribution of multi-channel sinus rhythm EGMs. Input signals were reshaped into 2D tensors (*time* $\times$ *channels*), with dimensions of (400, 2048). The encoder consisted of successive convolutional and pooling layers that compressed the inputs into a 50-dimensional latent distribution, parameterized by mean and log-variance. Latent vectors were sampled using the reparameterization trick, a mathematical technique used in VAEs to enable gradient-based training through random nodes [12]. The decoder reconstructed EGMs from the latent space using transposed convolutions, intermediate anti-aliasing filters, and a final *tanh* activation.

VAE-C extended the same encoder–decoder design to both sinus and AF EGMs. Conditioning was introduced by incorporating the rhythm class, which can be sinus or AF, into the latent representation, guiding the decoder to generate class-specific reconstructions.

2.2.2. Training Settings

For VAE-S and VAE-C, signals from Dataset A and Dataset B (respectively) were normalized between -1 and 1 and downsampled to 200 Hz to optimize batch size. The dataset was then divided into training (75%), validation (15%), and test (10%) subsets using a random split for the former and stratified split (class-wise) for the later. Both models were trained using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 400 for 90 epochs, with early stopping at 10 epochs to avoid overfitting. A learning rate scheduler was applied to improve convergence.

The VAEs were trained to minimize an objective loss function that extends beyond the standard reconstruction loss. In this work, we contribute additional perceptual terms specifically designed to preserve the morphology, local dynamics, and frequency content of real EGMs. The resulting total loss objective is defined as:

$$\begin{aligned}\mathcal{L}_{\mathcal{T}} = 0.35 \cdot \mathcal{L}_{\mathcal{R}} + \beta \cdot \mathcal{L}_{\mathcal{KL}} + 0.5 \cdot \mathcal{L}_{\mathcal{C}} + 0.35 \cdot \mathcal{L}_{\mathcal{G}} \\ + 0.25 \cdot \mathcal{L}_{\mathcal{H}} + 0.10 \cdot \mathcal{L}_{\mathcal{S}}\end{aligned} \tag{1}$$

where $\mathcal{L}_{\mathcal{R}}$ is the mean squared error (MSE) between original and reconstructed EGMs, $\mathcal{L}_{\mathcal{KL}}$ is the KL divergence between the approximate posterior and the prior, where the parameter β was linearly increased from 0 to a maximum of 4.0 over the first 10 training epochs. This gradual schedule mitigated posterior collapse in early stages and progressively enforced latent space regularization. For VAE-C, the latent prior was conditioned on rhythm class using a one-hot encoding.

The perceptual terms were designed to capture complementary aspects of signal quality. The correlation loss $\mathcal{L}_{\mathcal{C}}$ was computed as one minus the average Pearson correlation coefficient between real and reconstructed signals across nodes, emphasizing morphological similarity. The gradient loss $\mathcal{L}_{\mathcal{G}}$ imposed a first-order temporal gradient penalty to enforce similarity in local dynamics, particularly during rapid transitions. The high-frequency loss $\mathcal{L}_{\mathcal{H}}$ encouraged the preservation of sharp deflections, such as atrial depolarizations, while the noise suppression loss $\mathcal{L}_{\mathcal{S}}$ penalized spurious high-frequency components absent from the original signal. All loss weights were tuned empirically to maximize performance.

2.3. EGM Generation Strategies

After training, synthetic signals were generated by sampling latent vectors $\mathbf{z}$ from the learned posterior distribution and decoding them into the time domain. For VAE-S, 200 synthetic signals were generated. RMSE was computed pairwise between synthetic and real signals, retaining for each synthetic the minimum value as its similarity score. The 25 signals with the lowest scores were selected to form the **synt-S-Dataset**. For VAE-C, generation was performed separately for each rhythm class, yielding the **synt-C-Dataset**, comprising 50 signals (25 sinus-conditioned, 25 AF-conditioned).

2.4. Evaluation

We evaluated VAE performance by (i) intrinsic fidelity metrics on test set, and (ii) testing their utility for dataset augmentation in a downstream noninvasive EGM reconstruction task.

Fidelity metrics: To assess the similarity between synthetic and real EGM signals, we employ metrics capturing both sample-level and distributional alignment. At the reconstruction level, we report MSE, where smaller values indicate closer signals; KL divergence, which increases with distributional mismatch; log-spectral distance (LSD), where lower values denote better spectral similarity; and Pearson correlation, where values close to 1 indicate strong linear similarity. To evaluate distributional alignment, we compute the maximum mean discrepancy (MMD) with smaller values reflecting closer distributions. Finally, to qualitatively assess coverage and separation in the learned latent space, we project both real and synthetic samples using t-distributed stochastic neighbor embedding (t-SNE).

Downstream Task: Noninvasive EGM estimation was

performed using the DL pipeline in [13]. First, dataset B was split into training (75%), validation (15%), and test (10%) sets using a stratified split (class-wise) ensuring that samples reserved for testing in the generation task were also included in the test set. To assess the impact of synthetic data, we defined two augmentation scenarios for the training set. In the first ($VAE - S@k$), Dataset B was augmented with k synthetic sinus rhythm signals from the synt-S-Dataset, with $k \in \{10, 14, 18, 20, 25\}$. In the second ($VAE - C@k_S + k_{AF}$), Dataset B was augmented with k_S synthetic sinus and k_{AF} synthetic AF signals from the synt-C-Dataset, where $k_S, k_{AF} \in \{10, 14, 18, 20, 25\}$. Two settings were explored, i) targeted augmentation of k_S, only augmenting sinus rhythm, ii) symmetric augmentation of k_S and k_{AF}, including both sinus and AF instances. Then, synthetic EGMs from the synt-S and synt-C datasets were projected to BSPMs via the forward model, producing paired BSPM–EGM samples that were concatenated with real data in the training set. To prevent bias, real and synthetic signals were alternated during training. To evaluate the fidelity of estimated EGM signals from BSPMs, we compute Pearson correlation and RMSE between the reconstructed and ground-truth signals.

3. Results and Discussion

Next, we present VAE performance, in terms of generation fidelity metrics and impact on a downstream task.

3.1. EGM Generation Performance

This section reports fidelity metrics from EGM generation using VAE-S and VAE-C. For synthetic EGM generation with VAE-S, the model obtained a mean LSD of 2.39 ± 0.32 and a correlation of 0.56 ± 0.09 with real signals, confirming reasonable frequency and morphology fidelity. Distributional similarity was assessed with MMD, yielding an average score of 0.27. The learned latent representation exhibited an average KL divergence of 0.45 ± 0.06, with 24 active units out of a total latent dimensionality of 50, indicating that the model effectively utilized part of the latent capacity while avoiding posterior collapse. Some examples of generation as 1d signals in Fig. 1.

For synthetic EGM generation with VAE-C, the model achieved a global LSD of 2.66 ± 0.39 and correlation of 0.13 ± 0.06, indicating moderate frequency- and morphology-level fidelity. For AF, distributional similarity was stronger (MMD 0.19) than for sinus (MMD 0.56), suggesting more consistent generative performance in the AF domain. Latent representations showed a KL divergence of 3.35 ± 1.68, with 33 active units out of 50 confirming robust class separation. The t-SNE projection of the latent space, shown in Fig. 2 revealed a clear separation between AF and sinus rhythm representations. This indicates model successfully learned discriminative latent features associated with each rhythm type.

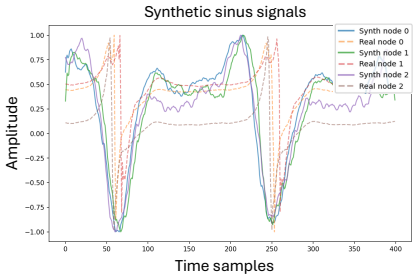

Figure 1. Synthetic signals produced by VAE-S. Dashed lines: real signals (reference). Solid lines: synthetic signals. The x-axis represents time samples (corresponding to 2 seconds), the y-axis to normalized amplitude.

Compared to VAE-S, sinus generation with VAE-C was degraded, with lower reconstruction fidelity and higher distributional divergence. Thus, class conditioning enabled AF generation but at the cost of reduced realism in sinus signals, as expected.

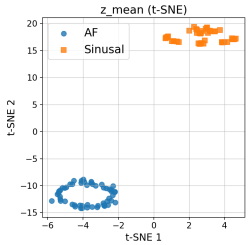

Figure 2. t-SNE projection showing two distinct clusters corresponding to AF and sinus classes

3.2. Downstream Task

The results for $VAE - S@k$ (Table 1) show that adding synthetic sinus EGMs improved reconstruction over the baseline ($k = 0$). Gains were most evident with moderate augmentation ($k = 14$–18), while larger sets ($k = 20$–25) plateaued or slightly degraded, suggesting redundancy or bias in the training data.

Table 1. Mean correlation (Corr) and RMSE of ground truth vs estimated EGMs from BSPMs for VAE-S@k, VAE-C@k_S, and VAE-C@k_S, k_{AF} (mean $\pm$ std).

	VAE-S@k		VAE-C@k_S		VAE-C@k_S, k_{AF}	
k	Corr	RMSE	Corr	RMSE	Corr	RMSE
0	0.160 ± 0.169	0.559 ± 0.103	0.160 ± 0.169	0.559 ± 0.103	0.160 ± 0.169	0.559 ± 0.103
10	0.183 ± 0.175	0.543 ± 0.095	0.156 ± 0.163	0.542 ± 0.086	0.146 ± 0.146	0.543 ± 0.092
14	0.175 ± 0.177	0.527 ± 0.071	0.166 ± 0.169	0.527 ± 0.082	0.151 ± 0.155	0.523 ± 0.072
18	0.207 ± 0.201	0.519 ± 0.082	0.169 ± 0.180	0.523 ± 0.082	0.132 ± 0.125	0.556 ± 0.086
20	0.198 ± 0.201	0.555 ± 0.096	0.175 ± 0.175	0.546 ± 0.087	0.138 ± 0.149	0.521 ± 0.066
25	0.214 ± 0.206	0.541 ± 0.097	0.186 ± 0.182	0.535 ± 0.088	0.116 ± 0.130	0.540 ± 0.072

The results in Table 1 shows mean correlation and RMSE in test set for $VAE-C@kS$ and $VAE-C@k_S+k_{AF}$. Augmenting with sinus-only generations in $VAE-C@kS$ shows a gradual improvement over the baseline, with correlation increasing from 0.16 to 0.186 at $k = 25$, and RMSE remaining relatively stable. This suggests that adding synthetic sinus signals enhances overall reconstruction, even when the generative model has been trained on both classes. In contrast, the symmetric augmentation in $VAE - C@k_S + k_{AF}$ did not yield consistent benefits: performance fluctuates across k, and correlation values remain generally lower than for $VAE - C@kS$. These results indicate that, at least in this setting, targeted augmentation of sinus signals is more effective than symmetric augmentation across classes, likely because the variability introduced by AF generations does not translate into global reconstruction gains.

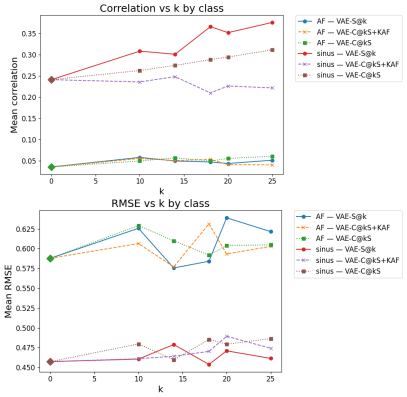

Figure 3. Mean correlation and RMSE for sinus rhythm and AF across augmentation settings

As shown in Fig. 3, reconstruction metrics by class (rhythm) reveal that $VAE-C@k_S$ improves sinus performance but remains below $VAE-S@k$, which achieves the highest correlation and lowest RMSE due to its specialization. In contrast, $VAE-C@k_S+k_{AF}$ introduces fibrillatory variability that slightly benefits AF, though at the expense of sinus accuracy. Overall, this indicates a trade-off: VAE-S is optimal for sinus augmentation, whereas VAE-C provides a slightly better balance across rhythms.

4. Conclusions and Limitations

We presented a VAE-based framework for generating synthetic atrial EGMs to address data scarcity. VAE-S achieved higher fidelity with sinus signals, while VAE-C enabled AF-specific generation with clear class separation but reduced sinus quality. This trade-off suggests that using separate networks for each rhythm type may enhance performance. Synthetic EGMs also improved data augmentation. Main limitations are the small dataset, limited physiological variability, and few test patients. Future work will focus on exploring advanced training strategies, like including spatial information of EGMs.

Funding

This work has been partially supported by Ministerio de Ciencia (PID2022-136887NB-I00).

References

[1] Van Gelder IC, Rienstra M, Bunting KV, et al. 2024 esc guidelines for the management of atrial fibrillation developed in collaboration with the european association for cardio-thoracic surgery (eacts). Eur Heart J 2024; 45(36):3314–3414.

[2] Gómez-Doblasa JJ, López-Garrido MA, Esteve-Ruiz I, et al. Epidemiología de la fibrilación auricular. Rev Esp Cardiol Supl 2016;16A:2–7.

[3] Tsyganov A, Petru J, Skoda J, et al. Anatomical predictors for successful pulmonary vein isolation using balloon-based technologies in atrial fibrillation. J Interv Card Electrophysiol 2015;44(3):265–271.

[4] Atienza F, Almendral J, Jalife J, et al. Real-time dominant frequency mapping and ablation of dominant frequency sites in atrial fibrillation with left-to-right frequency gradients predicts long-term maintenance of sinus rhythm. Heart Rhythm 2009;6(1):33–40.

[5] Zheng Y, Xia Y, Carlson J, et al. Atrial average conduction velocity in patients with and without paroxysmal atrial fibrillation. Clin Physiol Funct Imaging 2017;37(6):596–601.

[6] Rodrigo M, Climent AM, Hernández-Romero I, et al. Noninvasive assessment of complexity of atrial fibrillation: correlation with contact mapping and impact of ablation. Circ Arrhythm Electrophysiol 2020;13(3):e007700.

[7] Knackstedt C, Schauerte P, Kirchhof P. Electro-anatomic mapping systems in arrhythmias. Europace 2008;10:iii28–iii34.

[8] Hernandez-Romero I, Molero R, Fambuena-Santos C, et al. Electrocardiographic imaging in the atria. Med Biol Eng Comput 2023;61(4):879–896.

[9] Bacoyannis T, Ly B, Cochet H, Sermesant M. Deep learning formulation of ecgi evaluated on clinical data. EP Europace 2022;24.

[10] Sviridov I, Egorov K. Conditional electrocardiogram generation using hierarchical variational autoencoders. arXiv preprint arXiv250313469 2025;.

[11] Rodrigo M, Climent AMea. Technical considerations on phase mapping for identification of atrial reentrant activity in direct-and inverse-computed electrograms. Circ Arrhythm Electrophysiol 2017;10(9).

[12] Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv13126114 2013;.

[13] Gutiérrez-Fernández M, Cámara-Vázquez MA, Hernández-Romero I, et al. Egm reconstruction from bsps in atrial fibrillation using deep learning. In Comput. Cardiol. (CinC), volume 50. 2023; 1–4.

Author for correspondence: Miriam Gutiérrez Fernández
mgutierrezf@vicomtech.org | Zorrotzaurreko Erribera, 2, 48014, Spain