ECGAN: Self-supervised generative adversarial network for electrocardiography

Lorenzo Simone[1][0000−0002−5010−7733] and Davide Bacciu[1][0000−0001−5213−2468]

Department of Computer Science, University of Pisa, Pisa 56122, Italy

Abstract. High-quality synthetic data can support the development of effective predictive models for biomedical tasks, especially in rare diseases or when subject to compelling privacy constraints. These limitations, for instance, negatively impact open access to electrocardiography datasets about arrhythmias. This work introduces a self-supervised approach to the generation of synthetic electrocardiography time series which is shown to promote morphological plausibility. Our model (ECGAN) allows conditioning the generative process for specific rhythm abnormalities, enhancing synchronization and diversity across samples with respect to literature models. A dedicated sample quality assessment framework is also defined, leveraging arrhythmia classifiers. The empirical results highlight a substantial improvement against state-of-the-art generative models for sequences and audio synthesis.

Keywords: generative deep learning · self-supervised · generative-adversarial networks · electrocardiography

1 Introduction

Nowadays, the usage of electronic health records and the digital health (e-health) market have grown significantly due to technological advancements. Across the field, there is an upward trend in collecting heterogeneous data via diagnostic tests, clinical trials and wearable devices, supported by demographic insights. The recording of heart's electrical activity (ECG), represents the most common non-invasive diagnostic tool for the early detection of life threatening conditions. Machine learning can play a key role in assisting clinicians by efficiently monitoring and stratifying the risk of patients ([2]). Although fully supervised models could seem a flawless end-to-end solution for clinical research, they come with various drawbacks. For instance, such models are typically trained leveraging clinical datasets often failing in adequately covering rare occurrences. Furthermore, data annotation burden highly skilled physicians requiring exceptional time resources. Therefore, the quality of the predictions is strictly influenced by these preliminary phases, emphasizing the demand for plentiful rigorously labeled samples.

Deep generative models (DGMs), by approximating real conditional distributions, are usually capable of guiding the generative process towards a specific

class of samples. For this reason, they represent a relevant tool for reducing unbalance from rare diseases via data augmentation or patient-specific synthesis ([6]). The employment of DGMs as model agnostic synthesizers could implicitly solve another issue: patient's privacy and data anonymization. Potential threats arise from having a large amount of training data containing highly sensitive medical records.

In this paper, we propose a novel architecture referred to as ECGAN, combining two fields of research: self-supervised learning (SSL) and deep generative models for time series. Our model has been devised specifically with electrocardiography data in mind. The role of self-supervised learning in this research stands in exploiting the underlying time series dynamics via recurrent autoencoders. The features learned through a preliminary reconstruction task are transferred via weight sharing to the generator and a latent space projection. The assessment of the proposed model is compared with state of the art generative models for time series and audio synthesis ([3,12,19]). It yields competitive results concerning structural fidelity, sampling diversity and data applicability to heart rhythm classification tasks.

The main contributions of this work are: (1) We introduce a deep generative model, specifically designed for electrocardiography, intersecting between self-supervised learning and the generative adversarial framework. (2) We propose a parsimonious transfer learning framework requiring lesser resources than its generative counterparts. (3) We evaluate different methods for quantitatively assessing generated ECGs, inspired by the evaluation of image generative models, we inspect the latent projection of an ECG arrhythmia classifier $\mathcal{C}$.

2   Related work

The complexity of modeling electrical heart's activity requires experts to cooperate across several fields of research. A pioneering research from [18] theoretically investigated the application of dynamical systems to electrocardiography. The approach consisted in approximating heart's rhythm via a complex network of relaxation oscillators. This initiating idea spurred on upcoming methods built upon: non-linear coupled oscillators ([15]); three-dimensional trajectories [11,14] and function compositions [13]. Those early efforts in dynamical models were able to produce highly realistic heartbeats, however they result oversimplified for augmentation procedures.

Modern approaches often leverage generative deep learning, which sacrifices model explainability in favor of the avoidance of complex differential equations. One of the initial attempts from [12], consisted in applying long-short term memories to continuous sequential music generation. The input to each cell is treated recurrently as a combination of noise and output from the previous time step. A similar conditional recurrent approach (RCGAN, [4]) replaced the association with previous time steps with conditional input embeddings. Notwithstanding the merit of those approaches, they merely focus on the adversarial minimax, without capitalizing on additional sources. On the other hand [19] recently pro-
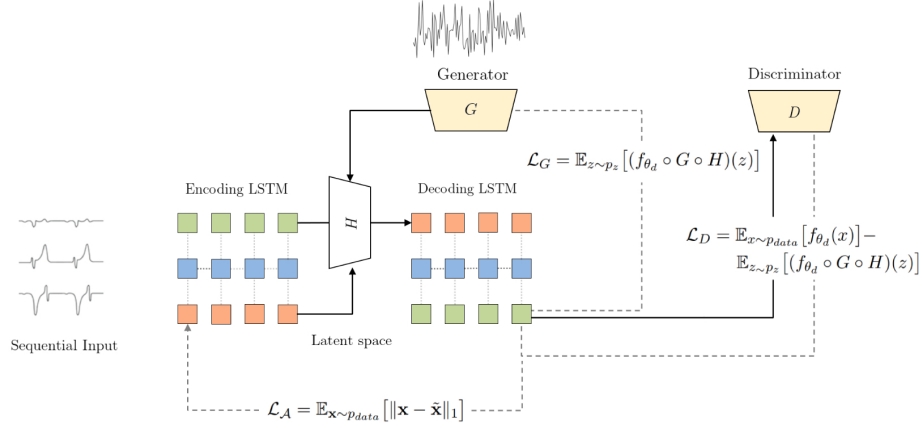
Fig. 1: Visual summary of the workflow of ECGAN. The main components (trapezoids) are correlated with their objective functions (dashed arrows). The sequential ECG input is either fed into the SSL block for preliminary reconstruction or to the discriminator D for adversarial training.

posed TimeGAN, consisting of four components simultaneously learning to encode, generate, and iterate across time in an autoregressive procedure.

Recently, deep generative applications to electrocardiography have shown to be an emerging trend, mainly for improved heartbeat classification. Researchers have favourably adapted the promising image-based GAN framework to ECG settings ([16]). At the same time, the latter have been combined with former ideas involving ordinary differential equations (ODEs) [7,5].

3   The ECGAN Model

ECGAN Architecture: The architecture encapsulates a latent space projection for the sequential domain within the adversarial framework typical of GANs. The model is composed of: an encoding recurrent mapping $E_\phi : \mathfrak{X} \to \mathfrak{H}$, which produces a latent representation $\mathfrak{H} \in \mathbb{R}^{h \times n'}$ through H and a decoding recurrent block, sharing its weights with the generator $G(\mathfrak{H}; \theta_g)$. We outlined each component in addition to their paths and training losses by dashed arrows in Figure 1. There are two distinctive possible workflows, both traversing the latent space. The first path starts from the ECG input sequence and learns to produce faithful reconstructions without involving adversarial components. Finally, the generative process samples directly from the latent feature projection which is processed by G and evaluated by D.

Sequence generation is conditioned at global level using embedding layers. During training and sampling, the hidden state of the generator is initialized with a non-linear embedding of the corresponding ECG label. Likewise, the dis-

criminator's output is perturbed by linearly combining its output and the associated embedding. Hence, they retain an initial conditional representation about the generated/classified sequence. This allows to both condition the generative process and influence the prediction based on prior class information.

Training procedure: The ECGAN training procedure comprises two independent phases. At an early stage, and differently from standard GANs, the generator and the discriminator are kept in an idle state. Throughout this phase, unlabeled samples are used for a self-supervised reconstruction of the input. The main purpose is guiding the generative process towards the synthesis of partly recognizable ECGs instead of white noise. Finally, prior knowledge is transferred to the generator, henceforth sampling directly from the latent space projection. The full procedure summarising both phases is detailed in Algorithm 1.

for $t = 0, 1, \ldots epochs_{ssl}$ **do**
    Sample $\{x^{(i)}\}_{i=1}^{m} \sim p_{data}$;
    $\delta_g \leftarrow \nabla_\theta [\frac{1}{m} \sum_{i=1}^{m} \|x^{(i)} - \tilde{x}^{(i)}\|_1]$;
    $\theta_g \leftarrow \theta_g - \alpha_s \cdot \mathrm{Adam}(\theta_g, \delta_g)$;
end
for $e = 0, 1, \ldots epochs_{adv}$ **do**
    for $d = 0, 1, \ldots steps$ **do**
        Sample $\{x^{(i)}\}_{i=1}^{m} \sim p_{data}$;
        Sample $\{z^{(i)}\}_{i=1}^{m} \sim p(z)$;
        $\delta_d \leftarrow \nabla_{\theta_d}[\frac{1}{m} \sum_{i=1}^{m} f_{\theta_d}(x^{(i)}) - \frac{1}{m} \sum_{i=1}^{m}(f_{\theta_d} \circ G \circ H)(z)^{(i)}]$;
        $\theta_d \leftarrow \theta_d + \alpha_d \cdot \mathrm{RMSProp}\ (\theta_d, \delta_d)$;
        $\theta_d \leftarrow \mathrm{clip}(\theta_d, -c, c)$;
    end
    Sample $\{z^{(i)}\}_{i=1}^{m} \sim p(z)$;
    $\delta_g \leftarrow -\nabla_\theta \frac{1}{m} \sum_{i=1}^{m}(f_{\theta_d} \circ G \circ H)(z)^{(i)}$;
    $\theta_g \leftarrow \theta - \alpha_g \cdot \mathrm{RMSProp}(\theta_g, \delta_g)$;
end

Algorithm 1: Training procedure.

The preliminary self-supervised phase is kept entirely modular depending on the downstream generative task. It requires an additional hyperparameter regulating the number of epochs and it employs Adam optimizer for the objective. Lastly, the adversarial training is carried out by using RMSProp for both the generator and the discriminator gradients. The discriminator is constantly kept within a compact parameter space through gradient clipping (an uniform window of $[-0.001, 0.001]$ have been selected for all the experiments). Among possible combinations, the best working training ratio between the generator and the discriminator has been keeping it balanced ($g : d = 1 : 1$).

Self-supervised module: Training starts with a preliminary self-supervised task avoiding the usage of labeled patterns. We are solely interested in capturing high-level features through the manifold, exploiting them later for the downstream conditional generative task. It should be noted that the flexibility

of the approach allows the extension to further SSL tasks such as signal denoising. The process can be summarized by the following three steps:

1. Sampling a batch of samples from the real data distribution $\{x^{(i)}\}_{i=1}^{m} \sim p_{data}$
2. Encoding the signal $x^{(i)}$ in a latent feature space through the embedding E and the projection map H, thus producing $\mathcal{H} \in \mathbb{R}^{h \times n'}$.
3. Recovering the original signal.

Alternatively, we also investigated the effect of perturbing the signal with Additive White Gaussian Noise as an additional and more challenging SSL task. The perturbed signal is treated as $x'_j = x_j + z_j$ where $z_j \sim \mathcal{N}(0, \lambda^2 \sqrt{N})$ and we will refer to the corresponding model as ECGAN_λ throughout the experiments.

Then, we proceed by defining the reconstruction objective:

$$\mathcal{L}_\mathcal{A} = \mathbb{E}_{\mathbf{x} \sim p_{data}} \big[\| \mathbf{x} - (G \circ H \circ E)(\mathbf{x}) \|_1 \big], \tag{1}$$

while the gradient of the expectation is approximated through Montecarlo integration as $\nabla_\theta [\frac{1}{m} \sum\limits_{i=1}^{m} \| x^{(i)} - (G \circ H \circ E)(\mathbf{x}^{(i)}) \|_1]$, allowing us to train G, H and E through Adam optimizer [8]. Formally, the process requires an additional hyperparameter s, i.e. the number of training steps. Anyway, further investigations employing it in an alternating fashion among G and D, are also suggested ($s : g : d = 0 : 1 : 1$ in the original GAN research). According to our experiments, the preliminary objective (1) as well as the self-supervised module introduced two desirable training properties:

– The discriminator is inevitably forced to a rough start, being encouraged to look for finer distinctive features between real and fake samples (Figure 4).
– Empirically, by inspecting pre-training checkpoints, the model grasps a primitive knowledge of ECG peaks location and duration (Figure 3).

Adversarial Training: Ultimately, the focus switches to the synchronized learning between the generator and the discriminator. Our preliminary experiments shared the widely known issues of standard GANs in encountering mode collapse scenarios (Appendix D). The adoption of combining WGAN principles introduced by [1] and the proposed novelties heavily mitigated this phenomenon. Thus, the model minimizes the Wasserstein distance via the Kantorovich-Rubinstein duality:

$$W(p_{data}, \tilde{p}) = \frac{1}{K} \sup_{\|f\|_L \leq K} \mathbb{E}_{x \sim p_{data}}[f(x)] - \mathbb{E}_{x \sim \tilde{p}}[f(x)]. \tag{2}$$

Differently from the original research, we sample a noise vector $\mathbf{z} \sim p_z$ processing it through the pre-trained projection map H, which is subsequently fed to the generator. As a result, the objective for the generator is:

$$\mathcal{L}_G = -\mathbb{E}_{z \sim p_z} \big[(f_{\theta_d} \circ G \circ H)(z) \big]. \tag{3}$$

Meanwhile, the discriminator is a real-valued mapping $f(\mathbf{x}; \theta_d) \colon \mathbb{R}^n \to \mathbb{R}$ using a linear activation function. Its K-Lipschitz continuity constraints in (2) are enforced by gradient clipping and the resulting objective is:

$$\mathcal{L}_D = \mathbb{E}_{x \sim p_{data}}\big[f_{\theta_d}(x)\big] - \mathbb{E}_{z \sim p_z}\big[(f_{\theta_d} \circ G \circ H)(z)\big]. \tag{4}$$

In other words, the discriminator should learn to estimate the distance between real patterns and synthesized ECGs, overcoming the prior advantage of the generator. Differently, the latter focuses exclusively on minimizing the Wasserstein distance between p_{data} and $\tilde{p}$ in (2).

4   Experimental Analysis

Despite the growing interest in generative models for health, their evaluation is still predominantly qualitative. Other than being time-consuming and arduous to reproduce, this is highly subjective. Common metrics adopted for images such as the Inception Score (IS) based on the Inception Network ([17]) are inadequate for biomedical time-series. Since they are mainly pre-trained on datasets comprising objects from the real world, it discourages any fine-tuning on image-like signal representations. Consequently, we pursued three core properties based on the application domain:

- **Diversity**: Generated samples should be equally distributed among different classes of arrhythmias resembling the real data distribution
- **Fidelity**: The morphological properties of generated signals should mirror those from real samples (including segment lengths or complex duration).
- **Functionality**: New data should be adoptable and useful to enhance arrhythmia predictive models.

We propose an approach specifically designed for this sequential domain which includes an ECG arrhythmia classifier $\mathcal{C}$. The Inception feature space is replaced with an intermediate activation layer from the predictive model. The latter allows us to define $\phi(p_{data})$ and $\phi(\tilde{p})$ as Gaussian random variables with empirical means μ_r, μ_g and covariances $\mathbf{C}_r, \mathbf{C}_g$. From now on, we rely on $p_{\mathcal{C}}(y \mid \mathbf{x})$ as the label distribution of $\mathbf{x}$ predicted by $\mathcal{C}$ and $p_{\mathcal{C}}(y) = \int_{\mathbf{x}} p_{\mathcal{C}}(y \mid \mathbf{x})d\tilde{p}$, being the marginal of $p_{\mathcal{C}}(y \mid \mathbf{x})$ over the generative probability distribution $\tilde{p}$. A detailed description and formulation of the metrics used throughout the experiments is covered in Appendix B.

To demonstrate the capability of the proposed model to synthesize realistic ECGs, we employed two different open datasets: MIT-BIH Arrhythmia Database and BIDMC Congestive Heart Failure Database. Each dataset is repeatedly split five times through a randomized hold-out technique, which assigns an equally partitioned 25% dedicated to hyperparameters tuning and model assessment of the classifier $\mathcal{C}$. The remaining portion is entirely dedicated for the generative task (3998 and 11576 samples respectively).

The metrics collected throughout the experiments are compared with state of the art generative models for signals, including WaveGAN and SpecGAN [3].

Fig. 2: A T-distributed Stochastic Neighbor Embedding (t-SNE) visualization of the activations $\phi\left(p_{data}\right)$ and $\phi\left(\tilde{p}\right)$ obtained by the classifier $\mathcal{C}$. The distribution of synthesized samples from ECGAN (third column) and the top performing IS model (second column) is compared with the original datasets (first column).
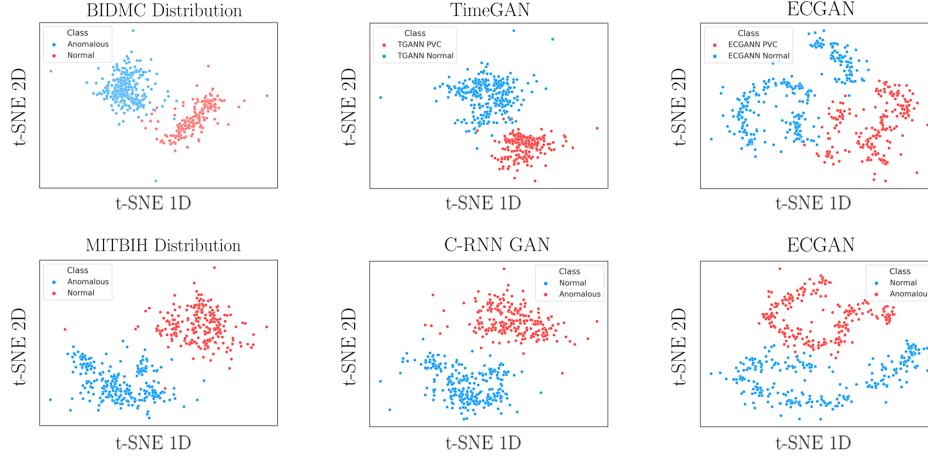

Table 1: Summary of the metrics evaluated for each generative model on both datasets. We report: the inception score (IS), the Frechèt Inception Distance (FID), the linear maximum-mean-discrepancy (MMD) and the Wasserstein distance (W).

Model	$\text{FID}(p_{data},\tilde{p}) \downarrow$	$\text{IS}(\tilde{p}) \uparrow$	$\text{MMD}(p_{data},\tilde{p}) \downarrow$	$\text{W}(p_{data},\tilde{p}) \downarrow$
		BIDMC Dataset		
ECGAN	$\mathbf{233.86 \pm 17.24}$	$\mathbf{1.97 \pm 0.09}$	$\mathbf{31.71 \pm 4.43}$	$\mathbf{0.52 \pm 0.01}$
ECGAN$_\lambda$	289.06 ± 10.94	1.96 ± 0.05	60.40 ± 3.50	0.79 ± 0.01
TimeGAN	328.03 ± 2.13	1.60 ± 0.01	144.90 ± 3.51	1.03 ± 0.00
WaveGAN	839.42 ± 7.26	1.32 ± 0.01	332.59 ± 8.93	0.81 ± 0.02
C-RnnGAN	917.22 ± 63.54	1.35 ± 0.03	788.67 ± 27.85	2.07 ± 0.03
SpecGAN	942.10 ± 8.25	1.26 ± 0.02	349.88 ± 6.56	0.79 ± 0.00
		MITBIH Dataset		
ECGAN	$\mathbf{45.49 \pm 1.82}$	$\mathbf{1.41 \pm 0.03}$	$\mathbf{17.41 \pm 1.62}$	0.40 ± 0.01
ECGAN$_\lambda$	78.87 ± 0.56	1.33 ± 0.01	37.66 ± 0.49	0.53 ± 0.01
C-RNNGAN	91.35 ± 3.00	1.40 ± 0.01	27.57 ± 1.15	0.52 ± 0.01
TimeGAN	113.81 ± 2.37	1.11 ± 0.01	60.51 ± 2.23	0.68 ± 0.01
WaveGAN	151.22 ± 0.94	1.03 ± 0.00	64.56 ± 0.84	$\mathbf{0.32 \pm 0.00}$
SpecGAN	151.21 ± 0.95	1.01 ± 0.00	64.97 ± 0.58	0.35 ± 0.01

Lead II Sinus rhythm

White noise   Epoch 15   Epoch 30   Synthesized   Real

Lead II CHF

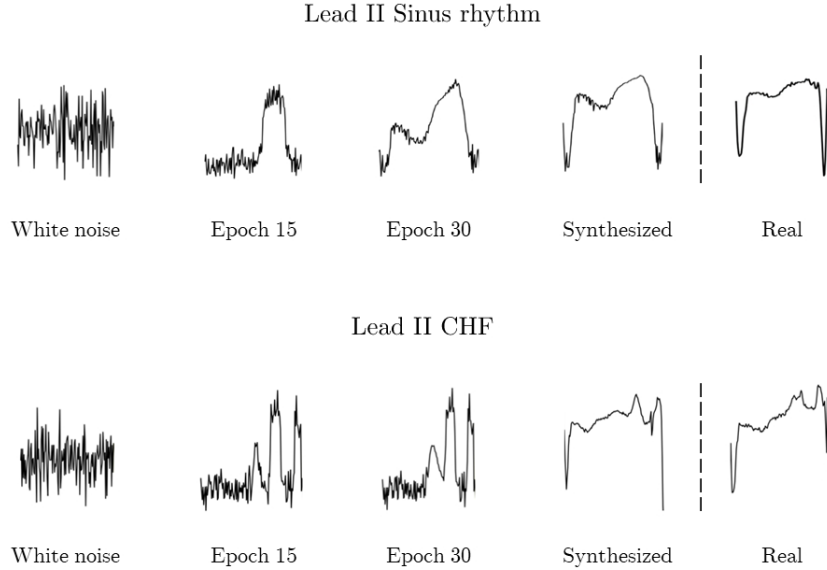White noise   Epoch 15   Epoch 30   Synthesized   Real

Fig. 3: Sampling progression over training epochs from ECGAN. The process starts with noise vectors and ends with a fully trained (SSL and adversarial) model sampling from sinus rhythm and congestive heart failure instances on the forth column.

These models were originally designed for audio synthesis, hence we performed a Short-time Fourier transform (STFT) to obtain spectrograms. We report the results over five different runs of the sampling procedure in Table 1. The proposed model consistently outperforms models selected from literature in both tasks considering the proposed metrics. Despite the increasing performance of the discriminative measures for the last task, the Inception Score suggests how spectrogram-based models are hardly producing well-distributed patterns. This is attributable to: an higher diversity of samples and the difficulty in masking zero padding from the original dataset.

In order to visually understand the capability of the classifier to distinguish synthesized patterns from original samples we computed a T-distributed Stochastic Neighbor Embedding (t-SNE [9]) of the activations $\phi(p_{data}), \phi(\tilde{p})$. The targeted real distribution in the first column exhibit two clearly distinguishable clusters for both datasets. On one hand, the competing models, according to the highest IS, follow the original separable arrangement (Figure 2). On the other hand, ECGAN provides a unique spatial distribution of patterns. Hence, sample diversity across instances is maximized while maintaining a positional reference to the original classes.

A contextual estimate of the impact of the SSL module and the subsequent adversarial phase over different training epochs is shown in Figure 3. We show,

from left to right, the progression of a fixed random vector from the latent space through the SSL process. The fourth column represents the synthesized ECG pattern after the adversarial training. Lastly, the fifth column belongs to the best match, according to the DTW distance, between the generated and real pattern.

5   Discussion

We have introduced ECGAN, a generative model for ECG data leveraging self-supervised learning principles. Throughout the experiments, ECGAN has been shown to be able to produce morphologically plausible ECGs, including specific pattern abnormalities, by global conditioning of the input. Our self-supervised generative adversarial framework encourages the sampling process to synthesize patterns coherently to real samples' dynamics. Anyway, a proper trade-off between preliminary and adversarial phases by a rigorous hyperparameter selection is needed. Ultimately, we obtained competitive results for the expected properties over the monitored metrics.

The synthesized patterns are also suitable for data augmentation procedures as suggested by the assessment of their functionality. We believe that our contribution could pave the way to several other biomedical applications. As a first further development, we advise the usage of the proposed approach for multi-lead ECGs. This would potentially contribute with an heavier impact on improving deep-learning arrhythmia classifiers. A 12-lead ECG synthesizer could involve other recording systems such as vectorcardiography (VCG), which requires multiple leads. Encouraging trends emerge from investigating the explainability of generative biomedical models, besides the scope of improving predictive models. Furthermore, an additional promising application field for DGMs comes from electroencephalography (EEG). Similarly to our domain, data availability is inevitably affected by complex experimental setups and prolonged wearing sessions from the patients.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proceedings of the 34th International Conference on Machine Learning - Volume 70. p. 214–223. ICML'17, JMLR.org (2017)
2. Beaulieu-Jones, B.K., Yuan, W., Brat, G.A., Beam, A.L., Weber, G., Ruffin, M., Kohane, I.S.: Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? NPJ digital medicine **4**(1), 1–6 (2021)
3. Donahue, C., McAuley, J., Puckette, M.: Adversarial audio synthesis. In: International Conference on Learning Representations (2019)
4. Esteban, C., Hyland, S.L., Rätsch, G.: Real-valued (medical) time series generation with recurrent conditional gans. arXiv preprint arXiv:1706.02633 (2017)
5. Golany, T., Freedman, D., Radinsky, K.: Ecg ode-gan: Learning ordinary differential equations of ecg dynamics via generative adversarial learning. Proceedings of the AAAI Conference on Artificial Intelligence **35**(1), 134–141 (May 2021)

6. Golany, T., Radinsky, K.: Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification. Proceedings of the AAAI Conference on Artificial Intelligence **33**(01), 557–564 (Jul 2019). https://doi.org/10.1609/aaai.v33i01.3301557

7. Golany, T., Radinsky, K., Freedman, D.: SimGANs: Simulator-based generative adversarial networks for ECG synthesis to improve deep ECG classification. In: III, H.D., Singh, A. (eds.) Proceedings of the 37th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 119, pp. 3597–3606. PMLR (13–18 Jul 2020)

8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

9. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**, 2579–2605 (Nov 2008)

10. Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S.H.A.: NeuroKit2: A python toolbox for neurophysiological signal processing. Behavior Research Methods **53**(4), 1689–1696 (feb 2021). https://doi.org/10.3758/s13428-020-01516-y

11. McSharry, P.E., Clifford, G.D., Tarassenko, L., Smith, L.A.: A dynamical model for generating synthetic electrocardiogram signals. IEEE transactions on biomedical engineering **50**(3), 289–294 (2003)

12. Mogren, O.: C-rnn-gan: A continuous recurrent neural network with adversarial training. In: Constructive Machine Learning Workshop (CML) at NIPS 2016. p. 1 (2016)

13. Roonizi, E.K., Sameni, R.: Morphological modeling of cardiac signals based on signal decomposition. Computers in biology and medicine **43**(10), 1453–1461 (2013)

14. Sameni, R., Clifford, G.D., Jutten, C., Shamsollahi, M.B.: Multichannel ecg and noise modeling: Application to maternal and fetal ecg signals. EURASIP Journal on Advances in Signal Processing **2007**, 1–14 (2007)

15. dos Santos, A.M., Lopes, S.R., Viana, R.R.L.: Rhythm synchronization and chaotic modulation of coupled van der pol oscillators in a model for the heartbeat. Physica A: Statistical Mechanics and its Applications **338**(3-4), 335–355 (2004)

16. Sarkar, P., Etemad, A.: Cardiogan: Attentive generative adversarial network with dual discriminators for synthesis of ecg from ppg. Proceedings of the AAAI Conference on Artificial Intelligence **35**(1), 488–496 (May 2021)

17. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence (2017)

18. Van Der Pol, B., Van Der Mark, J.: Lxxii. the heartbeat considered as a relaxation oscillation, and an electrical model of the heart. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **6**(38), 763–775 (1928)

19. Yoon, J., Jarrett, D., Van der Schaar, M.: Time-series generative adversarial networks. Advances in neural information processing systems **32** (2019)

A   Dataset Preparation

We considered two ECG classification datasets, both publicly accessible from PhysioNet. The first dataset have been used as a baseline, since it has a lower diversity across patterns and it has been automatically annotated.

BIDMC Congestive Heart Failure Database (ECG5000):
The open-source dataset *BIDMC-CHFDB* is a collection of long-term ECG recordings from 15 subjects suffering from severe congestive heart failure (NYHA class 3-4). The initial recordings were twenty hours long, and comprised two-lead ECG signals sampled at 250 Hz with a 12-bit resolution over a range of $\pm 10 mV$. They have been collected at Boston's Beth Israel Hospital with an approximate bandwidth of $0.1 - 40$Hz. The collection is composed of normal rhythm beats as well as premature atrial and ventricular complexes (PVC) and R-on-T PVCs.

Preliminary procedures: The first dataset, being patient-specific, serves as a baseline task for the generative models examined. Five thousand singular beats have been extracted following the principles adopted for the dataset "ECG5000". After having centered each fragment by its QRS complex, we obtained sequences of equal length by linear interpolation.

MIT-BIH Arrhythmia Database:
The MIT-BIH Arrhythmia Database is composed of 48 half-hour two-lead ambulatory ECG recordings. It comprises 47 different subjects and it was collected at the BIH Arrhythmia Laboratory. For the generative task, we have used the II lead undersampling it from 360Hz to 125Hz. Differently from the first dataset, it provides five annotated classes performed by two cardiologists (in accordance with the Association for the Advancement of Medical Instrumentation AAMI-EC57.

Preliminary procedures: The initial distribution of samples across the five classes is highly unbalanced. We proceeded by discarding the classes S and F, the first being a collection of several atrial and supra-ventricular events and the second being a mixture of the latter. This could unnecessarily complicate the generative task. The resulting classes: N referring to normal rhythm and V standing for premature ventricular complexes were highly unbalanced (N contains 72470 samples, while V only 5788). In order to balance the two classes for the generative task, we performed a standard undersampling procedure.

B   Metrics definition

The discrepancy measures from the experimental results are detailed utilizing the label distribution $p_{\mathcal{C}}(y|\mathbf{x})$ and its marginal distribution $p_{\mathcal{C}}(y)$ from $\mathcal{C}$.

The Inception Score (IS) is computed as:

$$\text{IS}(\tilde{p}) = e^{\mathbb{E}_{\mathbf{x} \sim \tilde{p}}[KL(p_{\mathcal{C}}(y|\mathbf{x}) \| p_{\mathcal{C}}(y))]} \tag{5}$$

The Fréchet Inception Distance (FID) is:

$$\text{FID}(p_{data}, \tilde{p}) = \|\mu_r - \mu_g\| + \text{Tr}(\mathbf{C}_r + \mathbf{C}_g - 2(\mathbf{C}_r \mathbf{C}_g)^{1/2}) \tag{6}$$

Kernel-based Maximum Mean Discrepancy (MMD) used as a dissimilarity measure between p_{data} and $\tilde{p}$:

$$\text{MMD}^2(p_{data}, \tilde{p}) = \underset{\mathbf{x}_r, \mathbf{x}'_r \sim p_{data},}{\mathbf{x}_g, \mathbf{x}'_g \sim \tilde{p}} [k(\mathbf{x}_r, \mathbf{x}'_r) - 2k(\mathbf{x}_r, \mathbf{x}_g) + k(\mathbf{x}_g, \mathbf{x}'_g)] \qquad (7)$$

,with a linear and radial basis function (RBF) kernel k:

$$k(\mathbf{x}_r, \mathbf{x}_g) = exp(-\frac{||\mathbf{x}_r, \mathbf{x}_g||^2}{2\sigma^2}) \qquad (8)$$

C   Ablation study

The focus in this appendix section is devoted to the empirical estimation of the contribution of the proposed novelties. We start by singularly detaching each component analyzing the impact on the evaluated metrics, following with qualitative improvements over adversarial training stability and finally presenting the latter over the generative process. We will refer, from now on, to a fully adversarial version of ECGAN having the self-supervised component and training removed as ECGAN$_{\text{NO-SSL}}$. The already mentioned set of metrics have been reused for the experiments in Table 2. The main difficulties without self-supervision have been: producing high quality ECGs, avoiding mode collapse and a lack of improvement over predictive metrics for classification tasks.

Table 2: Summary of the metrics for the ablation study of the SSL module evaluated on both datasets.

Configuration	FID$(p_{data}, \tilde{p}) \downarrow$	IS$(\tilde{p}) \uparrow$	MMD$(p_{data}, \tilde{p}) \downarrow$	W$(p_{data}, \tilde{p}) \downarrow$
		BIDMC Dataset		
ECGAN	233.86 ± 17.24	1.97 ± 0.09	31.71 ± 4.43	0.52 ± 0.01
ECGAN$_{\text{NO-SSL}}$	1156.0 ± 13.83	1.03 ± 0.02	728.91 ± 13.95	1.19 ± 0.01
		MITBIH Dataset		
ECGAN	45.49 ± 1.82	1.41 ± 0.03	17.41 ± 1.62	0.40 ± 0.01
ECGAN$_{\text{NO-SSL}}$	1189.46 ± 12.83	1.02 ± 0.10	752.1 ± 8.45	1.20 ± 0.01

Throughout the experiments we have monitored the loss of inner components from ECGAN, namely the generator, the discriminator and the SSL module as shown in Figure 4. In the first column, over five different runs, the discriminator starts to improve significantly at an early stage. As a consequence, there is a clear downward trend for the generator, which is incapable of improving nor synthesize plausible samples. The pre-training SSL module contributes to mitigating this phenomenon promoting a balance between G and D.
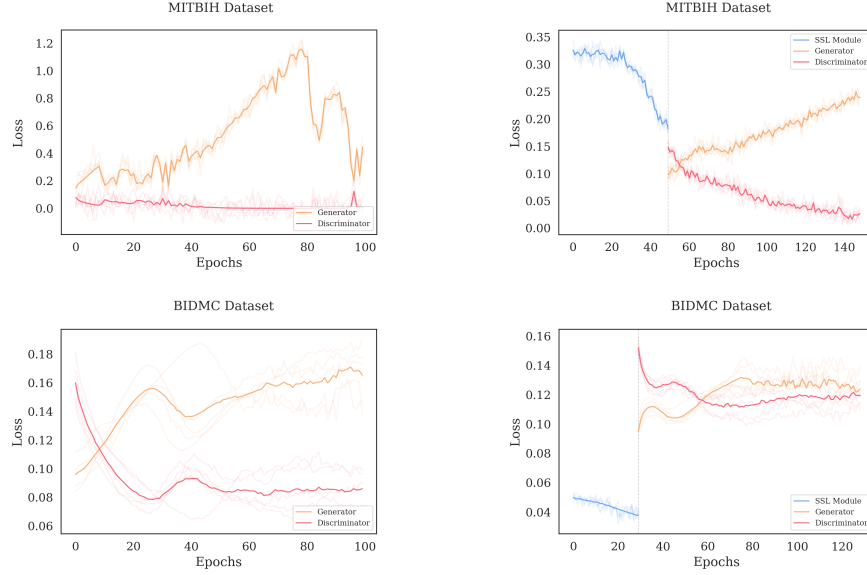
Fig. 4: The loss over training epochs of the generator and the discriminator from standard GAN (first column) is compared with its counterparts and the SSL module from the proposed model (second column).

D   Mode collapse

Standard generative adversarial networks have been proven their strengths, especially in image generative tasks, over the last years of research. Still, they pose several difficulties in having a balanced learning process. We reported the limitations encountered by training them from both qualitative and quantitative perspectives. The latter have been measured computing the dynamic time warping (DTW) distance over 30 generated patterns from ECGAN and standard GAN (Figure 5).

The overall quality of the generative process is unsatisfactory other than suffering from a noticeable mode collapse behaviour (Figure 6). Apart from being capable of producing R peaks, there are not other distinctive features in the produced signal (Figure 6b).
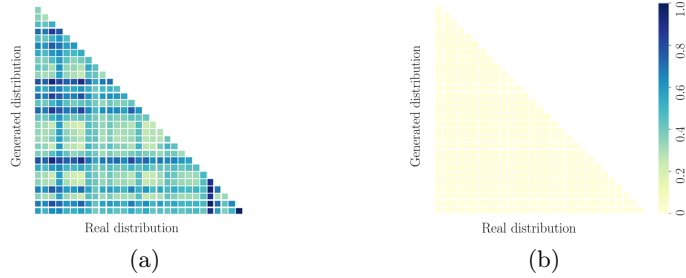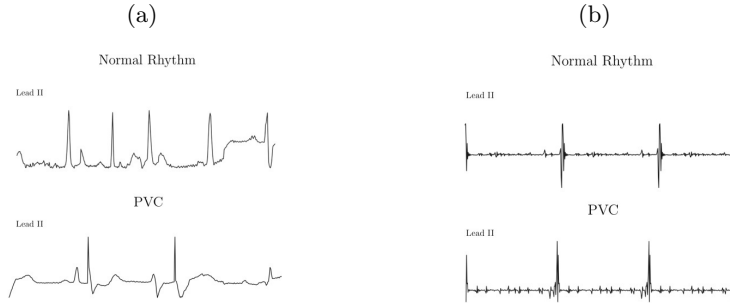
Fig. 5: DTW Correlation matrix over 30 samples from ECGAN (a) and standard GAN (b), which is suffering from mode collapse.

Fig. 6: Synthesized sample from standard GAN (b) failing in learning a reliable approximation of the real data distribution. There is a clear evidence of *mode collapse* with a fair discrepancy from real patterns (a).

E   Segments analysis

As a further quantitative evidence of the overall adherence of the generated signals to the corresponding ECG classes from the original datasets, we retrieved the period of meaningful segments. The latter have been carried out via an automatic feature extraction phase (Figure 7), with the objective of measuring: QRS complexes, QT, ST and PR segments.

Distinctive traits and patterns of PVCs, differently from normal rhythms, emerge also for the generated signals as can be observed in Table 3. These properties are clearly desirable both for patterns fidelity and data augmentation procedures involving classifiers relying on those features.

We proceed by showing the morphological distribution of a subset of synthesized and baseline ECG patterns from the first dataset (Figure 8). The overall adherence to segment lengths is respected without sacrificing sampling diversity.
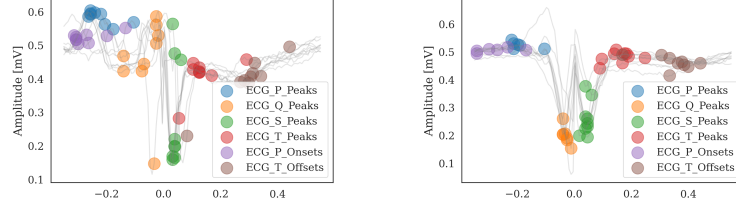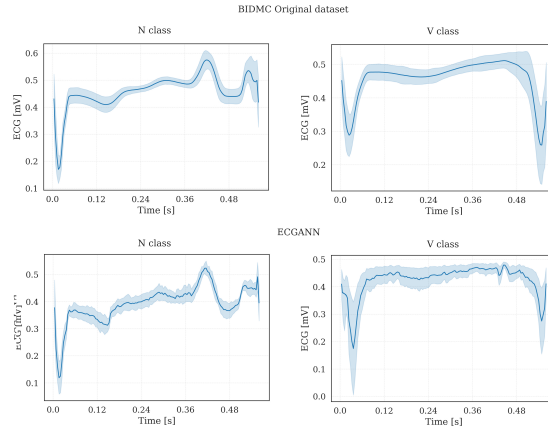
Fig. 7: Visual summary of the segmentation of P,Q,S and T peak onsets and offsets for both normal and anomalous classes ([10]).

Table 3: Summary of the time duration of meaningful segments after feature extracting Q,R,S and T peaks. A comparison between the distribution of segments duration in seconds between real BIDMC samples those produced by ECGAN.

Dataset class	QT (s)	QRS (s)	PR(s)	ST(s)
BIDMC (N)	0.55 ± 0.13	0.11 ± 0.03	0.60 ± 0.12	0.24 ± 0.05
BIDMC (V)	0.52 ± 0.11	0.06 ± 0.04	0.61 ± 0.03	0.13 ± 0.09
ECGAN (N)	0.53 ± 0.15	0.13 ± 0.05	0.59 ± 0.07	0.26 ± 0.08
ECGAN (V)	0.60 ± 0.05	0.12 ± 0.11	0.60 ± 0.05	0.12 ± 0.11

Fig. 8: Morphological distribution of a normal ECG (left) and anomalous (right). The top row is sampled from BIDMC, while the bottom is generated by the proposed model.

F   Functionality assessment

The architecture of inner feature learners is flexible allowing us to utilize stacked LSTMs for an higher level of abstraction. Meanwhile, the discriminator is built upon three stacked convolutional blocks: 1D convolutional layer followed by global average pooling and instance normalization. The latter allowed significant improvements over batch normalization. The choice of adopting LSTMs for our generative task is based on the requirement of learning long-term sequential rather than spatial features. Differently, convolutional blocks are better suited for our discriminative duty by focusing on local patterns in parallel. The details of these architectural choices are reported extensively in Table 4.

Table 4: Architectural details of the layers from the discriminator and the generator.

Generator	
Component	Shape
LSTM	$128 \times h$
LSTM ($\times 4$)	$128 \times n$
Discriminator	
Component	Shape
1D CNN (kernel size = 6)	$128 \times n$
Instance Normalization	$128 \times n$
ReLU Activation	$128 \times n$
1D CNN (kernel size = 6)	$128 \times n$
Instance Normalization	$64 \times n$
ReLU Activation	$64 \times n$
1D CNN (kernel size = 6)	$32 \times n$
Instance normalization	$32 \times n$
ReLU Activation	$32 \times n$
Global average pooling	32×1
Linear activation	1×1

The architecture of the CNN-based arrhythmia classifier $\mathcal{C}$, have been employed also to evaluate generated signals from a functional perspective. The objective of this phase is exploiting the inclusion of synthetic data to the training set to improve generalization capabilities on the initially retained internal test set. In the long run, each of the baselines achieved improved performances except for audio synthesis models (SpecGAN and WaveGAN).

Table 5: Predictive results for the classification task. A fixed number of balanced samples for each of the generative models is randomly inserted within the training set, and later evaluated over the test set.

Dataset	Accuracy	Specificity	Sensitivity	Precision	F1
Training					
Original	0.990 ± 0.01	0.986 ± 0.02	0.996 ± 0.04	0.997 ± 0.03	0.991 ± 0.02
ECGAN	0.997 ± 0.01	0.999 ± 0.01	0.994 ± 0.03	0.996 ± 0.02	0.997 ± 0.01
TimeGAN	0.987 ± 0.02	0.980 ± 0.03	0.997 ± 0.03	0.997 ± 0.02	0.988 ± 0.01
C-RNNGAN	0.991 ± 0.01	0.985 ± 0.02	0.999 ± 0.01	0.999 ± 0.00	0.992 ± 0.01
SpecGAN	0.960 ± 0.03	0.931 ± 0.05	0.998 ± 0.01	0.998 ± 0.01	0.963 ± 0.03
WaveGAN	0.937 ± 0.04	0.925 ± 0.08	0.954 ± 0.07	0.969 ± 0.04	0.943 ± 0.03
Testing					
Original	0.981 ± 0.02	0.988 ± 0.02	0.973 ± 0.09	0.979 ± 0.07	0.991 ± 0.02
ECGAN	0.989 ± 0.01	0.998 ± 0.00	0.977 ± 0.03	0.982 ± 0.02	0.997 ± 0.01
TimeGAN	0.988 ± 0.01	0.995 ± 0.04	0.978 ± 0.05	0.983 ± 0.04	0.988 ± 0.01
C-RNNGAN	0.987 ± 0.01	0.994 ± 0.00	0.979 ± 0.05	0.984 ± 0.04	0.992 ± 0.01
SpecGAN	0.969 ± 0.02	0.956 ± 0.05	0.985 ± 0.00	0.988 ± 0.00	0.963 ± 0.03
WaveGAN	0.950 ± 0.03	0.927 ± 0.07	0.979 ± 0.02	0.983 ± 0.01	0.943 ± 0.04