Title

AFE-GAN: Synthesizing Electrocardiograms with Atrial Fibrillation Characteristics Using Generative Adversarial Networks *

Authors

Wang, Xianglong

Sahiner, Berkman

Scully, Christopher G

et al.

Publication Date

2023-07-01

DOI

10.1109/embc40787.2023.10340565

AFE-GAN: Synthesizing Electrocardiograms with Atrial Fibrillation Characteristics Using Generative Adversarial Networks*

Xianglong Wang[1], Berkman Sahiner[2], Christopher G. Scully[2], and Kenny H. Cha[2]

Abstract— **Labeled ECG data in diseased state are, however, relatively scarce due to various concerns including patient privacy and low prevalence. We propose the first study in its kind that synthesizes atrial fibrillation (AF)-like ECG signals from normal ECG signals using the AFE-GAN, a generative adversarial network. Our AFE-GAN adjusts both beat morphology and rhythm variability when generating the atrial fibrillation-like ECG signals. Two publicly available arrhythmia detectors classified 72.4% and 77.2% of our generated signals as AF in a four-class (normal, AF, other abnormal, noisy) classification. This work shows the feasibility to synthesize abnormal ECG signals from normal ECG signals.**

Clinical significance - The AF ECG signal generated with our AFE-GAN has the potential to be used as training materials for health practitioners or be used as class-balance supplements for training automatic AF detectors.

I. Introduction

Atrial fibrillation (AF) is the most common type of cardiac rhythm disorder with an increasing prevalence [1]. The estimated prevalence of AF is 2.8% globally and 1-2% in both Europe and North America, leading to the presence of AF in 3-6% of hospital patients with acute conditions [1]. Presence of AF is linked to higher mortality rates [2] and an increased risk of stroke [3] and dementia [4]. Current diagnosis of AF is largely based on electrocardiogram (ECG) analysis by medical practitioners [5]. In this paper, we propose a method for synthesizing atrial fibrillation-like ECGs since the prevalence of AF in real data sets is low.

The earliest model of ECG synthesis used a set of dynamics-based differential equations with user-set characteristics such as the mean and standard deviation of the heart rate [6]. An extended version of the above method was implemented as the ECG-SYN [7]. Another model by Cao et al. [8] examined the feasibility of synthesizing 12-lead ECG from 3-lead ECG signals and electrode placing data. These modeling equation-based methods offer great insights to visual characteristics of ECG, but are not capable of accounting for noise that presents in real signals and synthesizing ECG signals in a particular kind of diseased state, such as AF.

Generative adversarial networks (GANs) [9], along with its variants, is one successful technique to generate new samples from the distribution learned from a given training set. GANs typically contain two networks: a generator that produces new data, and a discriminator that attempts to discriminate the generated data from real training samples [10]. The two networks are trained in unison, with the goal of generating data that shares characteristics of the training samples as described in section II. Recently, GANs have been applied towards generating ECG signals, including synthesizing Lead-V ECG from Lead-I [11] or synthesizing whole ECG signals from small databases [12], [13].

We present the AFE-GAN (Atrial Fibrillation-like ECG Generative Adversarial Network), which generates AF-like ECG from normal ECGs by using a GAN. Unlike previous studies based on the MIT-BIH arrhythmia database that only contain 8 AF records, our approach is based on a large database of 5,442 signals collected by a modern portable ECG recording system. In addition, our model does not require pairing normal and AF ECG from the same patient and works with general pools of normal and AF ECG signals. Our approach will have more potential to scale and translate to real-life applications. We envision our generated signals to be potentially used in clinical training situations and/or supplementing the development of automatic AF detectors. In this paper, section II highlights the characteristics of AF and normal ECGs, followed by our strategy and methods of generating AF ECG from normal ECG signals. We present the results of our generated ECG signals and evaluate them on a beat-by-beat basis and on a record basis in section III. We discuss our results and evaluations in section IV and conclude our paper in section V.

II. Methods

There are three main morphological waveform components in ECG: the P wave from atrial depolarization, the QRS complex from ventricular depolarization, and the T wave from ventricular repolarization [14]. In a normal ECG signal, all components of the ECG are clearly visible and the RR intervals are regular. Two main visual differences separate AF ECG from normal ECG, summarized below:

1) Beat-wise: The P wave is suppressed or completely absent. The T wave is suppressed and fibrillatory f waves are present.
2) Record-wise: The interval between the R-peaks (RR interval) is more irregular, indicating irregular heart rate. Although not present on all AF signals, the overall heart rate often appears faster.

Features from above and are the main features used for visual and conventional algorithmic AF detection [15], [16].

*This work was supported by CDRH Critical Path funding and in part by an appointment to the Research Participation Program at the CDRH administrated by ORISE through an interagency agreement between the US Department of Energy and the US Food and Drug Administration.

[1]Xianglong Wang is with the Department of Biomedical Engineering, University of California, Davis, xlowang@ucdavis.edu

[2]Berkman Sahiner, Christopher G. Scully, and Kenny H. Cha are with Center for Devices and Radiological Health, U.S. Food and Drug Administration, kenny.cha@fda.hhs.gov

The objective of the AFE-GAN is to transform normal ECG into "AF-like" ECG. The differences between normal ECG and AF ECG signals set two main goals for AF ECG synthesis: beat morphology and RR-interval variability. Shi et al.'s WarpGAN model [17], which performed style transfer and warping to convert photos into caricature images in 2D, inspired our design of the AFE-GAN.

A. Network Structure

The AFE-GAN is effectively a WarpGAN [17] adapted for 1D ECG conversion. The AFE-GAN consists of a deformable generator, a style encoder, and a discriminator. An overview of the generator can be found in Figure 1. The generator consists of a content encoder, a decoder, and a warp controller.

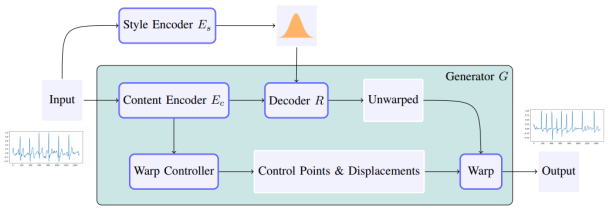


Fig. 1. Generator of 1D-adapted WarpGAN. Given a normal ECG signal, the generator performs a style transfer and learns the control points and their corresponding displacements for the input signal. The post-style transfer signal is then warped with a differentiable warping module and generates a synthetic AF ECG signal. Data and parameters are represented by rectangle nodes without borders and neural network components are represented with rounded rectangle nodes.

The encoder-decoder network fulfills our first requirement to transform normal beats to AF-like ones by performing local style transfer. The content encoder E_c has three convolutional layers of kernel size 7, 4, 4, followed by three residual blocks with size 4. The decoder R contains three residual blocks with Adaptive Instance Normalization (AdaIN), followed by two transpose convolutional layers with kernel size 4 and a convolutional layer with size 7 to generate the output. The content encoder E_c transforms the input signals into feature maps. A normally-distributed style code $s \sim \mathcal{N}(0, 1)$ is drawn for each input signal and fed into the decoder. The decoder R processes the feature maps and the style code to perform style transfer and generate an unwarped signal.

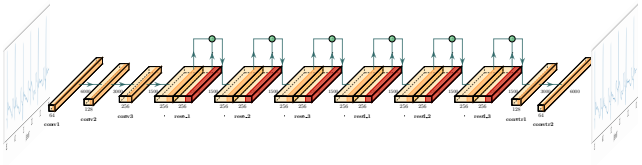


Fig. 2. Generator of the AFE-GAN. The content encoder includes the input to rese_3; the decoder consists of the resd_1 to the generated (unwarped) ECG waveform. The residual blocks resd in the decoder use AdaIN normalization with parameters fed from the style encoder/controller.

The warp controller, which is not present in most GANs, addresses the second requirement so that RR variability is introduced to the generated signals. The warp controller learns a fixed number of control points and their corresponding displacements using the feature maps from the content encoder. The control points are locations on the signals that are placed equally across the time axis. Following the style transfer, the control points are modified by their displacements provided by the warp controller using a differentiable warping module. Warping is performed by displacing the region of a signal associated with a given control point in the time axis. The warping points are generated using three fully connected layers. The displacement generated by the warping controller is amplified by a constant scaling factor α at the end.

During each step of training, we randomly select an AF ECG signal x_A from all AF ECG signals $\mathcal{X}_A$ and a normal ECG signal $\boldsymbol{x}_N$ from all normal ECG signals $\mathcal{X}_N$ in our database. Each signal in training is randomly assigned a style code $s \sim \mathcal{N}(0, 1)$, which is used as the parameters of the AdaIN layers in R that has been shown to effectively control the visual styles [18]. In our case, the style code is used to control the amplitude of each component in the generated ECG and potential absence of beats. The style encoder E_s, a three-layer convolutional network with the same structure as our conv1-conv3 layers in Figure 2 followed by a fully connected layer, learns the mapping between each ECG signal and the randomly-drawn style code s as an unsupervised way to learn the time trends. The style encoder E_s will generate the style code s during testing (signal generation) so that we could obtain variations on visual styles. To avoid losing semantic information during style transfer, we require the AF and normal signals to be capable of being reconstructed from the feature maps. We adopt the identity loss $\mathcal{L}_{idt}$ for this purpose.

$$\begin{aligned} \mathcal{L}_{idt}^{A} &= \mathbb{E}_{\boldsymbol{x}_A \in \mathcal{X}_A}[||R(E_c(\boldsymbol{x}_A), E_s(\boldsymbol{x}_A)) - \boldsymbol{x}_A||_1] \\ \mathcal{L}_{idt}^{N} &= \mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N}[||R(E_c(\boldsymbol{x}_N), E_s(\boldsymbol{x}_N)) - \boldsymbol{x}_N||_1]. \end{aligned} \tag{1}$$

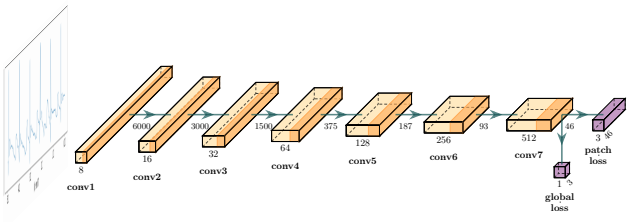


Fig. 3. Discriminator of the AFE-GAN. The discriminator contains 7 convolutional layers of kernel size 4 and step 2. The output of the discriminator is a 46-by-3 matrix for calculating the patch loss and an overall predictor for the global loss. We selected 7 layers to allow a window size of more than 300, which is enough for covering at least one beat.

The structure of the discriminator of the AFE-GAN can be seen in Figure 3. The discriminator is a straightforward 7-layer convolutional neural network for classification. The goal of the training process is to produce a model that applies both style transfer and warping of a normal ECG to generate

an ECG with AF characteristics. A patch-based discriminator and a global discriminator classify the input and generated signals into one of three categories: normal, AF, generated. The global discriminator classifies the whole signal while the patch discriminator classifies continuous samples of a fixed window size. In our case, the global discriminator is charged with identifying the overall appearance as AF; the patch discriminator is used for identifying beat morphology. Let D_{pA}, D_{pN}, D_{pG} denote the logits of AF, normal, and generated signals from the patch-based discriminator; D_{gA}, D_{gN}, D_{gG} from the global discriminator. The patch adversarial losses $\mathcal{L}_p$ and the global adversarial losses $\mathcal{L}_g$ can be written as follows:

$$\begin{aligned}
\mathcal{L}_p^G &= -\mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N, s \in S}[\log D_{pA}(G(\boldsymbol{x}_N, s))] \\
\mathcal{L}_p^D &= -\mathbb{E}_{\boldsymbol{x}_A \in \mathcal{X}_A}[\log D_{pA}(\boldsymbol{x}_A)] - \mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N}[\log D_{pN}(\boldsymbol{x}_N)] \\
&\quad - \mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N, s \in S}[\log D_{pG}(G(\boldsymbol{x}_N, s))] \\
\mathcal{L}_g^G &= -\mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N, s \in S}[\log D_{gA}(G(\boldsymbol{x}_N, s))] \\
\mathcal{L}_g^D &= -\mathbb{E}_{\boldsymbol{x}_A \in \mathcal{X}_A}[\log D_{gA}(\boldsymbol{x}_A)] - \mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N}[\log D_{gN}(\boldsymbol{x}_N)] \\
&\quad - \mathbb{E}_{\boldsymbol{x}_N \in \mathcal{X}_N, s \in S}[\log D_{gG}(G(\boldsymbol{x}_N, s))].
\end{aligned} \tag{2}$$

The network is optimized end-to-end with the following overall cost functions:

$$\begin{aligned}
G^* &= \arg\min_G [\lambda_p \mathcal{L}_p^G + \lambda_g \mathcal{L}_g^G + \lambda_{idt}(\mathcal{L}_{idt}^A + \mathcal{L}_{idt}^N)] \\
D^* &= \arg\min_D [\lambda_p \mathcal{L}_p^D + \lambda_g \mathcal{L}_g^D].
\end{aligned} \tag{3}$$

B. Dataset and Training

We used a subset of the publicly available training set from the 2017 Physionet Challenge as our training set [15], [19]. The challenge dataset consists of 8,528 short (9-61s) single-lead ECG records from one of three portable ECG devices. All the records are recorded at a sampling frequency $f_s = 300$ Hz. These records are labeled as one of the four classes: normal, AF, other abnormal, and noisy, mainly by a single expert. For the current work we only used records labeled as either normal or AF [15] of the reference labels and were longer than 20 seconds. Our dataset contains 87.6% (4768) normal signals and 12.4% (674) AF signals out of the 5,442 signals that satisfy our criteria.

Twenty (20) seconds of each ECG signal is extracted. Since the beginning of these handheld ECG recordings are often noisy, we extracted the last 20 seconds from signals of length 20-25 seconds and 5th to 25th second on signals longer than 25 seconds. In addition, the complexity of the signal that would need to be generated could be reduced by shortening the signal in this preliminary study. The extracted signal is then filtered with a bandpass filter with cutoff frequencies at 0.25 and 55 Hz to remove baseline wander and power line noise. We inverted signals that have larger negative amplitudes since these signals are likely to have been inverted during recording with the portable machine. The filtered signals make up our database and were consumed as the training set.

We optimized the GAN using an ADAM [20] optimizer. Results presented below use data trained at 130,000 steps with a learning rate of 2.5×10^{-5}. We set $\lambda_p = 2.0$, $\lambda_g = 1.0$, and $\lambda_{idt} = 10.0$ for the components of the cost function. We used 12 control points for the warp controller to allow approximately 1-2 beats between each control point. A scaling factor $\alpha = 2.5$ was chosen to obtain the most AF-like results.

C. Results Analysis

We performed beat-wise and signal-level characterizations for the generated AF-like signals. For beat-wise analysis, we detected the locations of the QRS complexes in all real and generated signals with the XQRS algorithm in the wfdb-python package from PhysioToolkit [19]. Each RR interval is then split in a 2:1 ratio to obtain the individual beats. Each beat is resampled to 250 time points using cubic interpolation and normalized to the amplitude of the R peak. An overall median beat is extracted for each of the three categories of signals, namely normal, real AF, and generated AF. The median beats are obtained by taking the median value at all 250 time points for every beat from each type of record. We then extracted a median beat for each generated AF signal and compared the signal-level median beats to the overall median beat of normal and real AF signals. We also extracted numerical values of the RR intervals from XQRS detector. In our analysis, we excluded the RR intervals that are less than 0.167s, more than 2.33s, or not within 1.66 [21] and $1/1.66$ times of the average RR interval of the signal. 97.8% of normal, 97.4% of real AF, and 83.8% of generated AF RR-intervals remained after this exclusion.

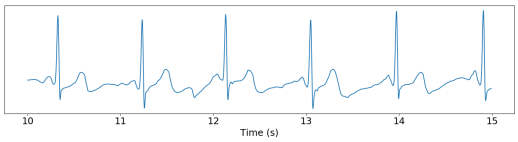
On a signal level, we first analyzed the QRS data obtained above to compare the signal-level distribution of the RR intervals. To evaluate the overall quality of our generated signals, we analyzed the sensitivity on two of the four winning AF detectors in the 2017 Physionet Challenge [15]. These detectors classify each signal into one of four categories: normal (N), AF (A), other abnormal (O), or noisy ($\sim$). We selected Hong et al.'s [22] and Datta et al's [16] detectors due to limitations of the computing environment. We also evaluated the normal and real AF signals in our training set with these two detectors for comparison. The reported accuracy values on our training set are slightly different from the reported training set accuracy due to our using 20-second segments instead of the whole signals during evaluation.

III. Results

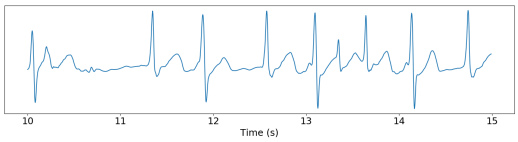
Figure 4 illustrates sample AF-like ECGs that we were able to obtain from our network. In this section, we present the beat-level and signal-level results that we obtained.

A. Beat-Level Analysis

The overall median beats for all three types of ECG signals are shown in Figure 5 (one median beat is computed from all signals from the same category, e.g. normal). The PQRST components of ECG beats are clearly visible in the normal median beat. The overall median beats of real and generated AF signals, as expected, show no P and suppressed T waves.

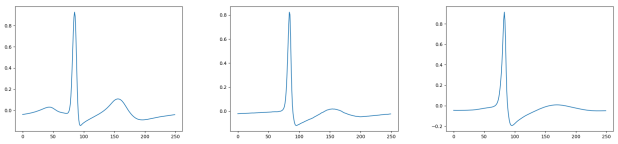


(a) Source Normal ECG



(b) Generated AF-like ECG

Fig. 4. Generated AF-like ECG (b) from source normal ECG signal (a). The 10-15th seconds of the signals are shown. The GAN is capable of introducing new QRS complexes and adjust the R-R variability.



(a) Normal (b) Real AF (c) Generated AF

Fig. 5. Median beat from (a) Normal, (b) Real AF, (c) Generated AF ECG Signals. The real AF and generated AF median beats have absent P waves and suppressed T waves compared to the normal median beat.

B. Signal-Level Analysis

We analyzed the whole generated signals with two distinct measures: distribution of RR intervals and sensitivity (positive AF rates) by two challenge-winning AF detectors from Physionet Challenge 2017 [15].

The Poincaré plot maps out the relationship between consecutive RR intervals [23]. We also calculated the goodness of fit r with a fixed model $RR_n = RR_{n+1}$ for all three kinds of signals. The real and generated AF signals show desirably similar R^2 values: real AF $R^2 = 0.924$; generated $R^2 = 0.906$. The normal signals have fairly regular RR intervals and mostly cluster around the diagonal line ($R^2 = 0.995$).

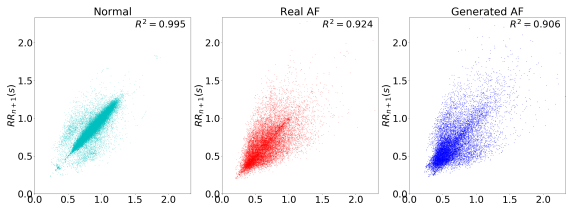


Fig. 6. Poincaré plot for all three types of signals - an RR interval RR_n is plotted against the next interval RR_{n+1}. Diagonal clusters depict regular RR intervals. The goodness of fit R^2 with respect to a fixed model $RR_n = RR_{n+1}$ is provided on the figure.

The two AF detectors detected 95.4% and 96.5% of normal ECG as normal, but after applying the AFE-GAN, most (77.2% and 72.4%) of the generated signals were detected as AF, as shown in Table I. Table I indicates that the sensitivity on the generated AF signals is approximately 13% lower compared to real AF signals. Only about 1% of the generated signals are still identified as normal, while the rest are identified as other or noisy.

TABLE I

EVALUATION OF NORMAL, GENERATED AF, AND REAL AF DATA ON HONG AND DATTA DETECTORS. BOLDED MEANS DESIRED CATEGORY.

	Hong Detector					
	Normal		Generated AF		Real AF	
	Count	%	Count	%	Count	%
Normal	**4547**	**95.4%**	39	0.8%	11	1.6%
AF	17	0.4%	**3682**	**77.2%**	**610**	**90.5%**
Other	178	3.7%	749	15.7%	48	7.1%
Noisy	26	0.5%	298	6.2%	5	0.7%
Total	4768	100.0%	4768	100.0%	674	100.0%
	Datta Detector					
	Normal		Generated AF		Real AF	
	Count	%	Count	%	Count	%
Normal	**4598**	**96.4%**	66	1.4%	28	4.2%
AF	14	0.3%	**3452**	**72.4%**	**575**	**85.3%**
Other	131	2.7%	742	15.6%	61	9.1%
Noisy	25	0.5%	508	10.7%	10	1.5%
Total	4768	100.0%	4768	100.0%	674	100.0%

IV. DISCUSSION

We were able to transform a normal ECG signal into multiple AF-like ECG signals with a GAN. We characterized the generated signals considering both the beat morphology and rhythm variability. The median beat from the generated AF-like signals resembles the median beat of true AF visually, while the mean and variability of record-level RR interval distributions resembles the real AF signals more than the normal signals. 85% of the generated signals are identified as AF by at least one of the two challenge-winning detectors, while the corresponding percentage is 92% for real AF.

During the generation of the AF-like signals, we manually amplified the amount of warping by a scaling factor of $\alpha = 2.5$, since we thought that the resulting signals are visually the best. Investigation of a single optimal scaling factor, or even an optimal scaling factor for each of the source normal signals, may be needed to achieve a better synthetic data set. We will now provide some of our insights to possibly achieve a better scaling factor without using visual inspection.

We use a warp controller to learn the location of 12 control points for a signal 6000 samples long along with their displacements. With 12 control points and 20 seconds of data, it is almost guaranteed to have more than one beat in some of the segments being warped. Unsurprisingly, warping segments containing one beat or more will lengthen or shorten the various interval measurements. This is a limitation on our model since we do not control where the control points lie. The scaling factor also has a tremendous effect on the variability of RR intervals. A possible criterion for selecting an appropriate α may be to check the distributions of the RR intervals on the generated AF signals using a reliable QRS detector.

The results from the AF detectors reveal an additional area of concern for our generated AF signals: noise. 6.2%

and 10.7% of the generated signals are identified as noisy, and about 5% of the signals are identified as too noisy by both detectors. A clean-up routine may be added to remove these noisy signals. The signals identified as "other abnormal" also warrant additional investigation. They may contain physiologically unrealistic signals that we should try to remove. Additional work is needed to potentially restrict the warping to limit the number of "other" and "noisy" detection, potentially by using initial morphology and rhythm-based analysis. These clean-up routines and additional observations may help with the higher fraction of failure by the XQRS algorithm. Naturally, the scaling factor affects the RR variability in the generated signals and will give different sensitivity results in the AF detectors. The sensitivity results from the AF detectors may also be used as a criterion for selecting appropriate α.

There are limitations to the data sets and tools we used in the GAN model and the evaluations: the training ECG were labeled by a single expert without knowing the actual diagnosis [15], potentially affecting network training and generation; the QRS detector can be thrown off by large noise peaks; the two AF detectors we used won the challenge with an overall F1-score (a harmonic average of precision and recall) of 0.83. Our AFE-GAN currently can only generate ECG signals of 20 seconds (native) or less (by cropping), although the duration can be extended by providing more computational resources. All of these factors may affect the overall observed performance of a generated data set. However, the best-level and signal-level characterizations we presented here help us to assess of the generated signals more comprehensively to understand AFE-GAN's limitations.

V. Conclusion

We successfully generated AF-like ECG signals from normal ECG signals using the AFE-GAN. Multiple AF-like signals can be generated from a source normal signal with different styles and scaling factors. The morphology of generated beats resemble more to real AF beats visually. The variability of RR intervals is similar to real AF signals on Poincaré plots. A high percentage of our generated signals are identified as AF (77.2%/72.4%) by two challenge-winning AF detectors. For next steps, we would like to use our synthesized data to supplement training of ECG detectors to potentially improve deep-learning ECG algorithms.

References

[1] G. Y. H. Lip, L. Fauchier, S. B. Freedman, I. Van Gelder, A. Natale, C. Gianni, S. Nattel, T. Potpara, M. Rienstra, H.-F. Tse, and D. A. Lane, "Atrial fibrillation," *Nature Reviews Disease Primers*, vol. 2, p. 16016, 3 2016.

[2] E. J. Benjamin, P. A. Wolf, R. B. D'Agostino, H. Silbershatz, W. B. Kannel, and D. Levy, "Impact of atrial fibrillation on the risk of death: The Framingham Heart Study," *Circulation*, vol. 98, no. 10, pp. 946–952, 1998.

[3] P. A. Wolf, R. D. Abbott, and W. B. Kannel, "Atrial fibrillation as an independent risk factor for stroke: The framingham study," *Stroke*, vol. 22, no. 8, pp. 983–988, 8 1991.

[4] P. Santangeli, L. Di Biase, R. Bai, S. Mohanty, A. Pump, M. Cereceda Brantes, R. Horton, J. D. Burkhardt, D. Lakkireddy, Y. M. Reddy, M. Casella, A. Dello Russo, C. Tondo, and A. Natale, "Atrial fibrillation and the risk of incident dementia: A meta-analysis," *Heart Rhythm*, vol. 9, no. 11, pp. 1761–1768, 2012.

[5] C. Gutierrez and D. G. Blanchard, "Atrial fibrillation: diagnosis and treatment," *Am Fam Physician*, vol. 83, no. 1, pp. 61–68, 2011.

[6] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE transactions on biomedical engineering*, vol. 50, no. 3, pp. 289–294, 2003.

[7] G. Clifford and P. McSharry, "Generating 24-hour ecg, bp and respiratory signals with realistic linear and nonlinear clinical characteristics using a nonlinear model," in *Computers in Cardiology, 2004*. IEEE, 2004, pp. 709–712.

[8] H. Cao, H. Li, L. Stocco, and V. C. Leung, "Design and evaluation of a novel wireless three-pad ecg system for generating conventional 12-lead signals," in *Proceedings of the Fifth International Conference on Body Area Networks*, 2010, pp. 84–90.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[10] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, "Deep learning in medical imaging and radiation therapy," pp. e1–e36, 1 2019. [Online]. Available: https://doi.org/10.1002/mp.13264

[11] J. Lee, K. Oh, B. Kim, and S. K. Yoo, "Synthesis of electrocardiogram v lead signals from limb lead measurement using r peak aligned generative adversarial network," *IEEE journal of biomedical and health informatics*, 2019.

[12] T. Golany and K. Radinsky, "Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 557–564.

[13] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network," *Scientific reports*, vol. 9, no. 1, pp. 1–11, 2019.

[14] K. F. Tan, K. L. Chan, and K. Choi, "Detection of the QRS complex, P wave and T wave in electrocardiogram," *IEE Conference Publication*, no. 476, pp. 41–47, 2000.

[15] G. D. Clifford, C. Liu, B. Moody, L. H. Lehman, I. Silva, Q. Li, A. E. Johnson, and R. G. Mark, "AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017," in *Computing in Cardiology*, vol. 44. IEEE, 2017, pp. 1–4.

[16] S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A. D. Choudhury, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, and S. Khandelwal, "Identifying normal, AF and other abnormal ECG rhythms using a cascaded binary classifier," in *Computing in Cardiology*, vol. 44. IEEE, 2017, pp. 1–4.

[17] Y. Shi, D. Deb, and A. K. Jain, "Warpgan: Automatic caricature generation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[18] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1501–1510.

[19] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 6 2000.

[20] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 12 2014.

[21] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomed. Eng*, vol. 32, no. 3, pp. 230–236, 1985.

[22] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "ENCASE: An ENsemble ClASsifiEr for ECG classification using expert features and deep neural networks," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

[23] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?" *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 11, pp. 1342–1347, 2001.