medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

DeepFake electrocardiograms: the beginning of the end for privacy issues in medicine

Vajira Thambawita[1,2,+], Jonas L. Isaksen[3,+], Steven A. Hicks[1,2], Jonas Ghouse[3], GustavAhlberg[3], Allan Linneberg[3,4], Niels Grarup[3], Christina Ellervik[3], Morten Salling Olesen[3], Torben Hansen[3], Claus Graff[5], Niels-Henrik Holstein-Rathlou[6], Inga Strümke[1], Hugo L. Hammer[1,2], Molly Maleckar[1,2], Pål Halvorsen[1,2], Michael A. Riegler[1+*], and Jørgen K. Kanters[3+*]

[1]SimulaMet, Oslo 0167, Norway

[2]Oslo Metropolitan University, Oslo 0167, Norway

[3]University of Copenhagen, DK2200 Copenhagen N, Denmark

[4]Bispebjerg and Frederiksberg Hospital, DK2400 Copenhagen NV, Denmark

[5]Aalborg University, Aalborg, Denmark

[6]Novo-Nordisk Foundation, Copenhagen, Denmark

Corresponding Authors *jkanters@sund.ku.dk and michael@simula.no

+these authors contributed equally to this work

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Summary

Big data is needed to implement personalized medicine, but privacy issues are a prevalent problem for collecting data and sharing them between researchers. A solution is synthetic data generated to represent real dataset carrying similar information.

Here, we present generative adversarial networks (GANs) capable of generating realistic synthetic DeepFake 12-lead 10-sec electrocardiograms (ECGs). We have developed and compare two methods, namely WaveGAN* and Pulse2Pulse GAN. We trained the GANs with 7,233 real normal ECG to produce 121,977 DeepFake normal ECGs. By verifying the ECGs using a commercial ECG interpretation program (MUSE 12SL, GE Healthcare), we demonstrate that the Pulse2Pulse GAN was superior to the WaveGAN to produce realistic ECGs. ECG intervals and amplitudes were similar between the DeepFake and real ECGs. These synthetic ECGs are fully anonymous and cannot be referred to any individual, hence they may be used freely. The synthetic dataset will be available as open access for researchers at OSF.io and the DeepFake generator available at the Python Package Index (PyPI) for generating synthetic ECGs.

In conclusion, we were able to generate realistic synthetic ECGs using adversarial neural networks on normal ECGs from two population studies, i.e., there by solving the relevant privacy issues in medical datasets.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Introduction

The use of artificial intelligence has increased in medicine over the past years. The overall aim of artificial intelligence in medicine is to aid clinicians based on complex data with decisions that are more accurate and improving personalized medicine. The prerequisite and foundation for artificial intelligence is the large amount of high-quality clinical data.

With updates of the General Data Protection Regulation (GDPR) regulative in the EU, the free flow of data has been restricted to ensure patient consent and anonymity[1]. Even anonymized deidentified data cannot be shared between research groups in different countries, because combining a few variables in an anonymized dataset, may allow for individual identification[2]. For example, knowing the zip code, birthday and sex is enough to identify 87% of US citizens[3]. On the other hand, large-scale, publicly available open-access medical datasets are required for personalized medicine to improve data-heavy machine learning solutions in medicine.

A solution to the privacy issue may be generation of synthetic realistic data. Synthetic data are data, which contain all the desired characteristics of a specific population, but without any sensitive content, making it impossible to identify individuals. Therefore, properly generated synthetic data are a solution to the privacy problem and will enable data sharing between research groups.

In this paper, we showcase synthetic electrocardiograms (ECG) as a complex example of medical data. An ECG is a voltage time series reflecting the electric currents within the heart, a widely used easy applicable and inexpensive clinical screening procedure to detect cardiac diseases. Using multiple electrodes, 3D propagation of cardiac electric impulses can be obtained and plotted as a standard 10-sec 12-lead ECG. Synthetic ECGs have been a topic of interest and research for many years. McSharry et al.[4] and Sayadi et al.[5] proposed mathematical dynamic models to generate continuous ECG signals, but these models were restricted to one

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

lead and did not reflect the distribution found in the normal population, nor did they give any insight in the mechanisms behind the disease.

Generative adversarial networks (GAN) were introduced in 2014 by Goodfellow et al. to generate synthetic data[6] using multi-layer perceptrons. Basically, a GAN consists of two neural networks: A generator network making signals (here ECGs) from random noise and a discriminator network evaluating whether the ECG is real or fake. During training, a mix of real ECGs and DeepFake ECGs are presented for the discriminator, which evaluates the ECGs assigning a score; high score for a likely real ECG, and low score for a supposed DeepFake ECG. As training proceeds, both the generator and discriminator improve until an equilibrium is reached[7]. Later, Radford et al.[8] developed a convolutional neural network GAN to generate synthetic images well suited for images like the ECG.

Since ECGs basically are time series, an initial approach was to use a WaveGAN[9] which is capable of generating sound signals. The classical WaveGAN is only able to output a single channel time series, so we modified the WaveGAN to generate all ECG channels (leads) (denoted WaveGAN*) instead of audio signals. We also introduce a novel DeepFake ECG U-net generative model, called Pulse2Pulse inspired by WaveGAN published by Donahue et al.[9] and compare our Pulse2Pulse generator to the WaveGAN generator.

In this paper, we present two GANs with the ability to generate an infinite number of 10-sec 12-leads synthetic "DeepFake" ECGs as a solution to overcome privacy issues related to real ECG data. These DeepFake ECGs can be openly distributed and freely downloaded as open access to be used by other scientists to develop ECG algorithms.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Results

We used ECGs from two population studies (GESUS[10] and Inter99[11]). To avoid chimeras between normal and abnormal ECG, we only trained the neural network with ECGs classified as normal by the MUSE 12SL. As shown in Table 1, both the WaveGAN* and Pulse2Pulse improved during training expressed as the percentage of DeepFake ECGs classified by the MUSE 12SLas normal ECGs. The Pulse2Pulse GAN trained faster than the WaveGAN* and had a better performance (expressed as fraction of ECGs classified as normal by the MUSE) than the WaveGAN* at their respective optimal number of training epochs (Table 1). Figure 1 shows a comparison of real and DeepFake ECGs, and the supplementary Figure S1 shows twenty randomly chosen DeepFake ECGs. Figure 2 shows the distribution of heart rates in the DeepFakes. By clinical definition Normal ECGs heart rates are between 60 and 99 beats per minute. The MUSE 12SL[12] classified 129 DeepFakes (0.5%) as sinus tachycardia (fast heart rate≥100) and 2863 (10.2%) as sinus bradycardia (slow heart rate<60). Figure 4 shows that cross correlation between as an example the QT interval and the RR interval were preserved. All cvariance structures can be seen in Supplementary Figure S2.

All DeepFake ECGs can be downloaded at OSF.io (https://osf.io/6hved/) with the corresponding ground truth parameters for the QT, RR, PR and QRS intervals and the P, STJ, R, and T amplitudes (see Figure 3 for ECG wave/interval naming terminology) delivered by the MUSE 12SL system (version 2.43). The DeepFake ECGs may be freely used for scientific use or commercial algorithm development if this paper is properly cited.)

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 1. Quantitative difference between WaveGAN* and Pulse2Pulse GAN in the initial training for determining the optimal network and optimal number of epochs. The best values are bolded for each GAN.

	Fraction of ECGs classified as Normal (%)	
Checkpoint (epochs)	WaveGAN*	Pulse2Pulse
500	20.9	78. 7
1000	69.5	81.2
1500	71.2	78.8
2000	**72.5**	79.7
2500	71.3	**81.6**
3000	65.3	81.5

Using the Pulse2Pulse model from the optimal number of epochs (2500), we generated 150,000 DeepFake ECGs. To ensure, that these ECGs were realistic, we uploaded the 150,000 ECGs to the GE MUSE system and analyzed them using the 12SL algorithm. We found that 81.3% of the 150,000 DeepFake ECG were classified as "Normal ECG" (vs. 81.6 % in the initial training). Table 2 compares real vs. DeepFake ECGs using eight ECG properties (heart rate, P duration, QT interval, QRS duration, PR interval, STJ amplitude, R amplitude, and T amplitude extracted using MUSE 12SL. See Figure 3 for ECG nomenclature). The real data included all ECGs from GESUS and Inter99 classified as "Normal ECG" which were used for training. DeepFake ECGs are presented both as all 150.000 generated ECGs and the subset classified as Normal ECG. The supplementary Table S4 summaries the most common reasons for classifying DeepFake ECGs as Non-Normal ECGs.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 2. Mean, standard deviation (std), 2.5%, and 97.5% percentile for standard ECG parameters in real and fake ECGs. BPM = beats per minute

		Real – Normal (7,233)				Pulse2Pulse – Normal (121,977)				Pulse2Pulse – All (150,000)			
		Mean	std	2.5%	97.5%	Mean	std	2.5%	97.5%	Mean	std	2.5%	97.5%
Heart rate	BPM	70	8	60	90	70	7	60	88	70	8	60	89
P Duration	ms	105	12	82	130	117	17	86	152	118	17	84	152
QT Interval	ms	395	21	352	436	395	20	354	436	395	22	352	436
QRS Duration	ms	90	9	74	110	92	9	78	112	93	10	78	114
PR Interval	ms	156	19	120	198	158	17	126	192	159	19	124	194
STJ amplitude (V5)	µV	2	27	-44	58	18	33	-44	87	16	36	-54	87
R Amplitude (V5)	µV	1287	402	600	2163	1275	367	620	2026	1273	402	566	2094
T Amplitude (V5)	µV	343	137	126	664	366	135	156	668	361	141	141	673

A sample real ECG:

A sample DeepFake ECG:

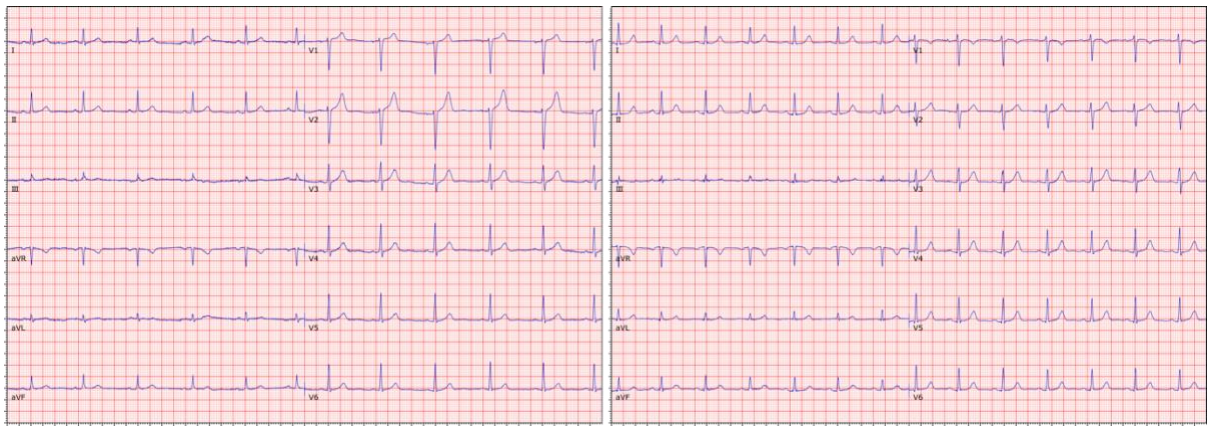

Figure 1. Comparison of examples of a real ECG (left lane) and a DeepFake ECG (right lane). See supplementary Figure S1 for 20 more randomly chosen pairs of real and DeepFake ECGs.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

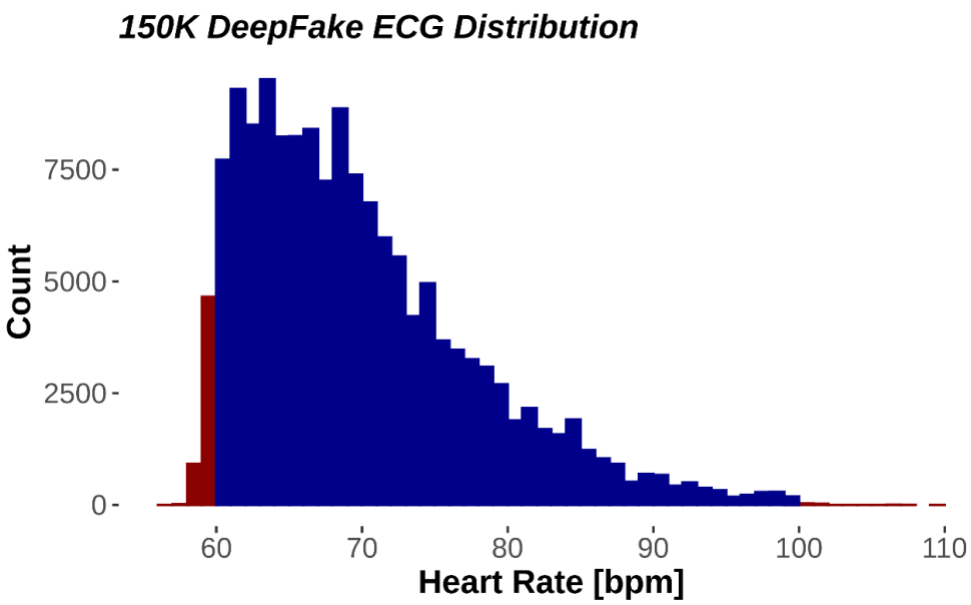

Figure 2. Distribution of heart rates in all 150.000 DeepFake electrocardiograms. Red fill denotes outside the normal heart rate range. Blue fill is within normal heart rate range (60-100).

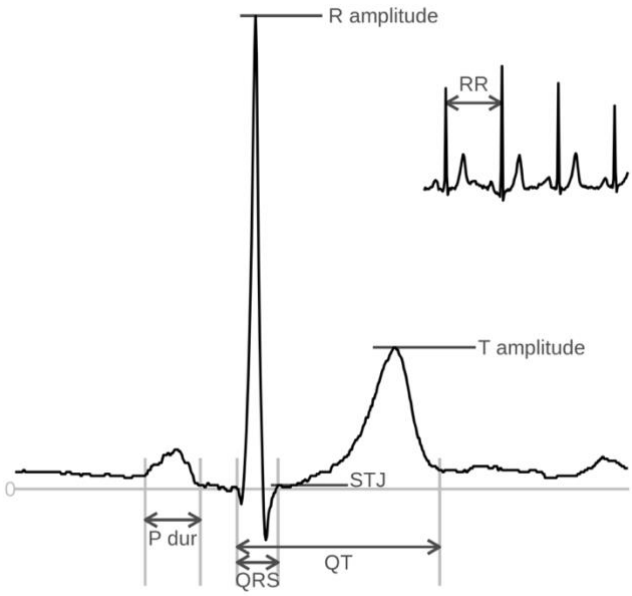

Figure 3. An ECG complex with the nomenclature of intervals (QT, QRS, P duration) and Amplitudes (STJ, R, T) and RR-interval (which can be converted to heart rate (HR) as HR=60/RR interval.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

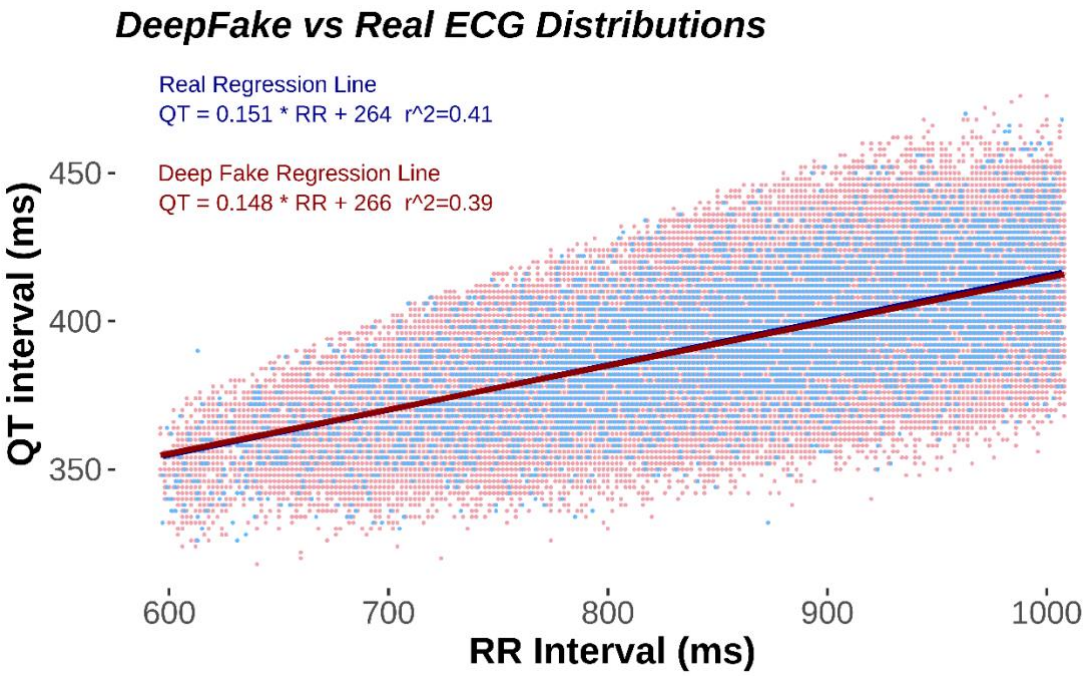

Figure 4. Scatter plot of the QT/RR interval relationship. Real ECG in blue and normal DeepFakes in red. DeepFake dots are nudged 1 ms to the left for visibility. Note that there are 121,977 normal DeepFakes and only 7,233 Real ECG making the DeepFake distribution more pronounced. As seen by the correlation coefficient r^2, the real and the fake DeepFake ECGs are similarly distributed.

Discussion

Although deep learning has been used for ECG analysis before[13,14], this study is the first study to generate realistic synthetic 10-sec 12-lead DeepFake ECGs. We demonstrate that the ECG characteristics from the real ECGs were similar to DeepFake ECGs. Hence, our DeepFake generator was able to construct synthesized ECG with similar intervals and amplitudes as the original population.

Nearly one fifth of the DeepFake ECGs were not recognized as Normal ECGs (Non-Normal) by the commercial MUSE 12SL ECG analyzer (No ECGs were rejected as being invalid).

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Many ECG parameters have hard boundaries in distinguishing between Normal and Non-Normal. For example, a normal heart rate is defined as between 60 to 99 bpm. Since we trained our model only on Normal ECGs, the input distribution for the GAN was a truncated asymmetric distribution. Thus, the clinically defined boundaries are skewed compared to the normal distribution of heart rates. The left truncation (at low heart rates) will discard more individuals than the right truncation (at high heart rates), and the final distribution of the real ECGs will be close to a truncated normal distribution with asymmetric truncations. The GAN will generally learn that heart rates outside 60-99 will not be valid, but small deviations will occur as seen in Figure 2 and Table 2. Since similar boundaries exist for many ECG parameters (for example PR interval >120 ms or QRS Interval<120 ms) sharp truncations would occur with several ECG parameters. This would lead to exclusion of some DeepFake ECGs, simply because the ECG intervals or amplitudes were just outside the normal range. Most ECG amplitudes and intervals were similar between real ECGs and DeepFake ECGs, but it was noteworthy that the STJ amplitude and the P duration had the greatest deviation between real ECGs and DeepFake ECGs. An explanation may be that both STJ and P amplitudes are small, and the network may tend to focus on larger waves such as the R and T waves. Following this theory, the network would to some extent neglect the smaller waves and features thereby introducing a larger uncertainty. Future networks may improve the ECG generation using conditional GANs to give more attention to smaller signal features. The Pulse2Pulse model was able to preserve the covariance structure between different ECG features, as seen in the most important relationship the QT/RR relationship which is known to have prognostic importance[15].

A challenging task is to define the optimal number of epochs for training. GANs tend to become unstable during the training process with the risk of the generator producing unrealistic output. To get an unbiased estimate how well the trained GAN performs, we used the commercial

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

MUSE 12SL system which automatically and reliably evaluates an ECG with a sensitivity of 99.9% and specificity of 100%[16]. Although the ECG discarded by the MUSE 12 SL may only have minimal abnormalities (like a heart rate of 59.9 where 60 is normal), the filtering of the DeepFake ECGs ensures, that the best epoch is chosen without bias, and the resulting ECGs are normal not only according to the discriminator, but also according to one of the most widely used ECG system in hospitals worldwide.

Personalized medicine will be dependent on big data, which demands international cooperation to ensure large datasets for researchers and the industry to work with. However, privacy and general data protection regulation rules are major obstacles for sharing data between researchers from different institutions and countries, or with the industry[17]. In conclusion, we show that we may overcome privacy and ethical[18] issues by constructing synthetic signals from real patients keeping the same clinical information as in the real dataset. The synthetic data generated by our Pulse2Pulse GAN makes it impossible to identify any patients, but still the ECGs remain useful for data scientists or industry to use for generating novel algorithms for ECG analysis. The approach is not limited to ECGs but should be expandable to all medical multichannel data, e.g., electroencephalography and electromyography. Therefore, the DeepFake ECGs generated from the Pulse2Pulse model can be used as a replacement to overcome the privacy constraints in real datasets.

Methods

The WaveGAN model is an evolution from the first GAN model introduced by Goodfellow et al.[6]. There are two deep neural networks named generator (G) and discriminator (D) to achieve the generation task from these GANs models. The main goal of the generator is to produce a data sample input (ECG(z)) from random noise (z) to the generator. The discriminator's task is to differentiate between real and fake data. We train the generator and discriminator together

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

as a competition (minmax game) between them. When a steady state is reached, the training halts, and the generator will generate realistic synthetic ECGs.

Data preparation

We have used two combined datasets: the Danish General Suburban Population Study[10] (GESUS) and the Inter99 study[11] (CT00289237, ClinicalTrials.gov). GESUS consists of 8,939 free-living subjects, and Inter99 consists of 6,667 free-living subjects with an available digital ECG. To avoid generation of hybrid ECG with mixed ECG abnormalities not occurring in real persons (e.g., to both be in sinus rhythm and atrial fibrillation at the same time which is impossible), we excluded ECGs who were not classified as normal (n=8,348) leaving 7,233 Normal ECGs for training.

A 12-lead 10-sec ECG consists only of 8 independent channels since 4 of the channels are simply trigonometric rotations. Therefore, the input ECG signal is 5,000x8 data points (corresponding to 10 sec with 500 samples per sec. x 8 channels). In addition to the up-scaling, we calculate the missing four channels with trigonometric functions to create the classic 12-channels ECG.

WaveGAN:* The input to WaveGAN* is a 1D random noise vector sampled from the uniform distribution (mean = 0, std = 1) with 100 x 1 passes through six deconvolution blocks to generate the desired output of 5000 x 8 samples. The deconvolution blocks are built from a series of four layers: an up-sampling layer, a constant padding layer, a 1D-convolution layer, and a ReLU activation function, consecutively. This implementation is deeper than the original architecture which use five deconvolution blocks used to generate synthetic music samples. Table S1 has comprehensive details of our WaveGAN* generator network.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Pulse2Pulse: The implementation of the Pulse2Pulse architecture (Figure 4) is inspired by the U-Net architecture[19] which is used for image segmentation. However, our Pulse2Pulse implementation is different than the original U-net implementation because the Pulse2Pulse implementation use 1D CNN for ECG signal generation rather than the 2D CNN used for original image segmentation task. The Pulse2Pulse network takes an 8×5000 noise vector which has the same dimension as the output dimension of a generated ECG. Then, we pass the noise through six down-sampling blocks followed by six up-sampling blocks as illustrated in Figure 3b. Each down-sampling block consists of a 1D-convolution layer followed by a Leaky ReLU activation. The up-sampling block is similar the deconvolution block used in WaveGAN*. In down-sampling, we have used Leaky ReLU instead of the ReLU layer used in the up-sampling to match the down-sampling operations to the discriminator. In addition to the up-sampling and down-sampling, the major modification is a bypass with down-sampling block features concatenating into the up-sampling block features represented by the black arrows in Figure 3b. To facilitate for this concatenation, we doubled the input size of up sampling blocks compared to WaveGAN* up sampling blocks. More details about Pulse2Pulse architecture are shown in the supplementary Table S1.

Discriminator: The same discriminator is used by WaveGAN* and Pulse2Pulse to discriminate between real and fake ECGs (Figure 3c). We used seven convolution layers (the original WaveGAN[9] has five layers), and each convolution layer is followed by a Leaky ReLU activation and the phase shuffle layer introduced in the original WaveGAN paper[9]. The discriminator takes an ECG as input (5000 samples * 8 channels) and outputs a score how close the ECG are to be determined fake or real.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

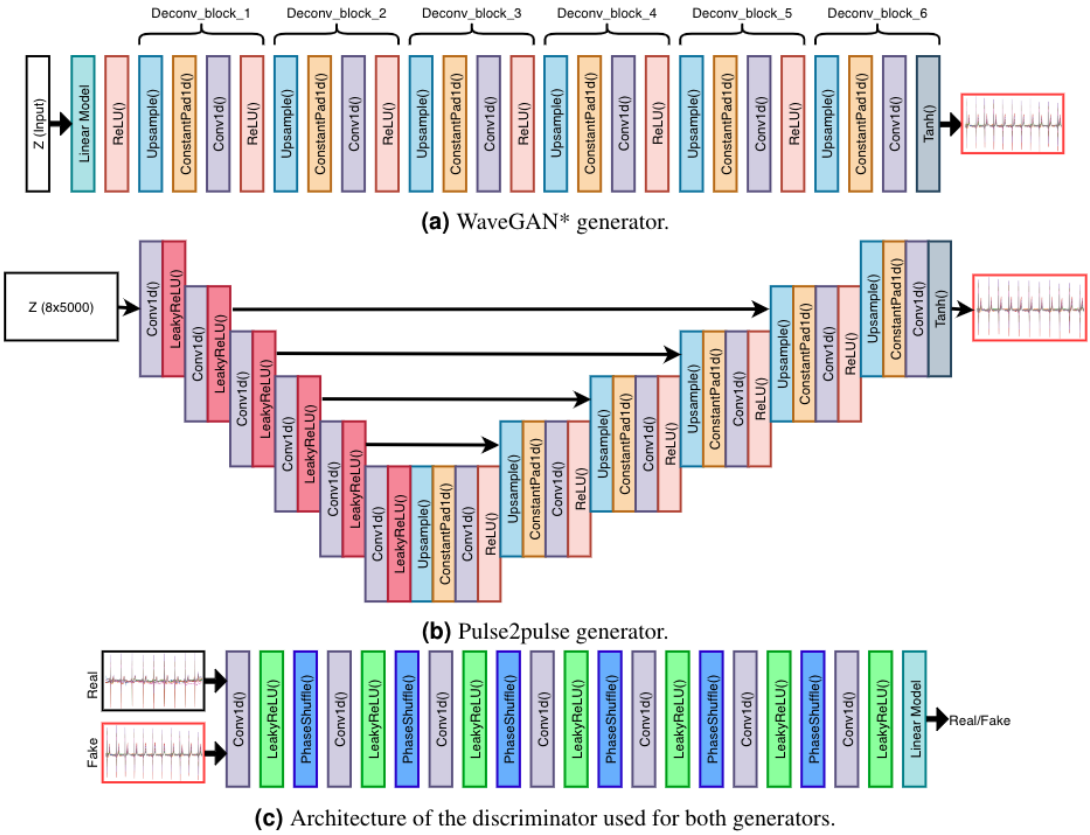

Figure 4. Model architectures of the generators and the discriminator used to generate synthetic ECGs. WaveGAN* uses a 1D noise vector with 100 points. Pulse2Pulse uses a 2D noise vector with size of 8×5000 as input, same as the output ECG size.

Training: The models were trained on a Ubuntu workstation with a double Xeon processor and a GeForce NVIDIA RTX 2080 running the Pytorch deep learning framework[20]. We ran all our experiments (generators + discriminator) using the Adam[21] optimizer with a learning rate of 0.0001, β1value of 0.5, and β2 value of 0.9. As loss function, we used gradient clipping WGAN-GP[22], to ensure faster and better convergence. Similar to the audio generation paper of WaveGAN[9], we updated (backpropagated) the discriminator five times per update of the generator. We used a batch size of 32, which is half of the original batch size of 64 used in the original WaveGAN paper, as a result of using larger networks than the WaveGAN networks.

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

We kept the training process until 3000 epochs (~10 days computing time) because we experienced unstable training curves for both WaveGAN* and Pulse2Pulse afterwards.

DeepFake ECGs: For evaluation of our two GAN models, we initially generated 10,000 ECGs from every 500 epochs until 3000 epochs from each GAN model. The DeepFake ECGs were transferred to the MUSE system and evaluated by the MUSE 12SL algorithm v. 2.43[12] using the fraction of DeepFake ECGs described as Normal (similar to the Real ECGs used for training). Using the best epoch for the best GAN, we generated 150.000 DeepFake ECGs. These DeepFakes were similar evaluated by the MUSE 12SL.

Data Availability: The Normal DeepFake ECGs are available at OSF (https://osf.io/6hved/) with corresponding MUSE 12SL ground truth values freely downloadable and usable for ECG algorithm development. The DeepFake generative model is available at https://pypi.org/project/deepfake-ecg/ to generate only synthetic ECGs.

Code Availability: The complete source code of all networks discussed in paper are available at GitHub (https://github.com/vlbthambawita/deepfake-ecg).

References

1. Voigt, P. & von dem Bussche, A. *The EU General Data Protection Regulation (GDPR)*. (Springer International Publishing, 2017). doi:10.1007/978-3-319-57959-7.
2. de Montjoye, Y.-A., Radaelli, L., Singh, V. K. & Pentland, A. S. Identity and privacy. Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science* **347**, 536–539 (2015).
3. El Emam, K., Jonker, E., Arbuckle, L. & Malin, B. A systematic review of re-identification attacks on health data. *PloS One* **6**, e28071 (2011).

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

4. McSharry, P. E., Clifford, G. D., Tarassenko, L. & Smith, L. A. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Trans. Biomed. Eng.* **50**, 289–294 (2003).
5. Sayadi, O., Shamsollahi, M. B. & Clifford, G. D. Synthetic ECG generation and Bayesian filtering using a Gaussian wave-based dynamical model. *Physiol. Meas.* **31**, 1309–1329 (2010).
6. Goodfellow, I. *et al.* Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **27**, 9.
7. Osborne, M. J. & Rubinstein, a: *A Course in Game Theory*. (MIT Press, 1994).
8. Radford, A., Metz, L. & Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *ArXiv151106434 Cs* (2016).
9. Donahue, C., McAuley, J. & Puckette, M. Adversarial Audio Synthesis. in *International Conference on Learning Representations* (2019).
10. Juhl, C. R., Miller, I. M., Jemec, G. B., Kanters, J. K. & Ellervik, C. Hidradenitis suppurativa and electrocardiographic changes: a cross-sectional population study. *Br J Dermatol* **178**, 222–228 (2018).
11. Ghouse, J. *et al.* Rare genetic variants previously associated with congenital forms of long QT syndrome have little or no effect on the QT interval. *Eur Heart J* **36**, 2523–2529 (2015).
12. GE Healthcare. Marquette™ 12SL™ ECG Analysis Program Physician's Guide 2056246-002 Revision C. (2015).
13. Attia, Z. I. *et al.* An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* (2019) doi:10.1016/S0140-6736(19)31721-0.
14. Hicks, S. A. *et al.* Explaining Deep Neural Networks for Knowledge Discovery in Electrocardiogram Analysis. *MedRxiv* **2021.01.06.20248927**, (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.04.27.21256189; this version posted April 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

15. Jensen, B. T. *et al.* QT dynamics in risk stratification after myocardial infarction. *Heart Rhythm* **2**, (2005).

16. Froelicher, V. F., Marcus, R. & Heidenrich, P. Prognostic Value of Computer Electrocardiography in Veteran Outpatients. *Fed. Pract.* **21**, 11–20 (2004).

17. El Emam, K., Rodgers, S. & Malin, B. Anonymising and sharing individual patient data. *BMJ* **350**, h1139 (2015).

18. Ienca, M. *et al.* Considerations for ethics review of big data health research: A scoping review. *PloS One* **13**, e0204937 (2018).

19. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *International Conference on Medical image computing and computer-assisted intervention* 234–241 (Springer, 2015).

20. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. *ArXiv191201703 Cs Stat* (2019).

21. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).

22. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. Improved training of Wasserstein GANs. *Adv. Neural Inf. Process. Syst.* **30**, 5767–5777 (2017).