Improving ECG Classification Using Generative Adversarial Networks

Tomer Golany
Technion
Israel Institute of Technology
Tomer.golany@cs.technion.ac.il

Gal Lavee
eBay Research
glavee@ebay.com

Shai Tejman Yarden
Cardiology Department
Sheba Medical Center
tegmanya@gmail.com

Kira Radinsky
Technion
Israel Institute of Technology
Kirar@cs.technion.ac.il

Abstract

The Electrocardiogram (ECG) is performed routinely by medical personell to identify structural, functional and electrical cardiac events. Many attempts were made to automate this task using machine learning algorithms. Numerous supervised learning algorithms were proposed, requiring manual feature extraction. Lately, deep neural networks were also proposed for this task for reaching state-of-the-art results. The ECG signal conveys the specific electrical cardiac activity of each subject thus extreme variations are observed between patients. These variations and the low amount of training data available for each arrhythmia are challenging for deep learning algorithms, and impede generalization. In this work, the use of generative adversarial networks is studied for the synthesis of ECG signals, which can then be used as additional training data to improve the classifier performance. Empirical results prove that the generated signals significantly improve ECG classification.

1 Introduction

Electrocardiography (ECG) is a non-invasive tool used for diagnosis and followup of cardiac anomalies, functional disorders and cardiac arrhythmias. The cardiac cycle is composed of electrical depolarization-repolarization patterns and presents the evolution of the heart's electrical activity over time. Any anomaly regarding the heart rhythm or the morphological pattern of the cardiac heart beats as sampled by the ECG, can indicate acute functional emergencies such as acute myocardial ischemia or acute rhythm disturbances, meaning an arrhythmia. Many studies were conducted in an attempt to reach high performance in ECG classification (Kass and Clancy 2006). Such models were used to reduce interpretation errors, namely by identifying life-threatening arrhythmias. Automatic ECG analysis may also reduce interpretation time by different medical personell while investigating non-life-threatening arrhythmias.

Most methods today for ECG classification focus on applying classical supervised machine learning methods. They perform feature engineering over the ECG wavelet (Chazal and Reilly 2007) and subsequently apply a supervised learning algorithm to predict the class. Classifiers such as SVM (Ye, Kumar, and Coimbra 2012), linear discriminant (LD) (Chazal and Reilly 2007), and Reservoir Computing With Logistic Regression (RC) (Escalona-Moran et al. 2014), have been explored. The performance of the algorithms strongly varied based on the engineered features, as they have to be informative, discriminating, and independent based on expert knowledge. Discounting the tedious effort needed for such a task, in practice, these methods have not been able to scale well across different patient's types of ECGs (Kiranyaz, Ince, and Gabbouj 2016). The changing nature of the ECG signal dynamics and morphological characteristics are significantly different across patients. It strongly depends on the patient's physical condition. For example, even for healthy patients, the R-R interval changes across time between each heart beat. Additionally, such an algorithm must also recognize the distinct wave types and discern the complex relationships between them over time in the presence of noise. To address these issues, deep learning approaches have been proposed (Al Rahhal et al. 2016). Methods, such as fully-connected networks (Al Rahhal et al. 2016), have shown significant performance boosts over the state-of-the-art with regard to known medical benchmarks. In this work, we experiment with a sequence-based classifier, recurrent neural network (RNN), that was recently used for ECG arrhythmias classification (Hou et al. 2019). This model is considered state-of-the-art for the task of ECG arrhythmia classification.

Although deep learning methods are showing promising results for ECG classification, they require large amount of training examples per class. Most life-threatening arrhythmias are extremely rare, limiting the amount of examples available to train deep learning models. Given the high variability between patients and internal-variability of heartbeat classification for same patients, building deep ECG models that might be used in practice has been limited. In this work, we overcome the sparseness of data by learning to synthetically generate ECG signals of different arrhythmias, which we later use to train the deep learning models. Intuitively, we build a model that learns to create synthetic ECG signals that present variability across the ECG signal and across different generated synthetic signals. The inter-variability and cross-variability through the generated signals might help deep learning algorithms to learn significant features and prevent overfitting to a specific patient's variability.

To learn such signals, we leverage Generative Adversarial Networks (GANs) (Goodfellow et al. 2014). GANs are a class of machine learning algorithms used in unsupervised machine learning, usually implemented by two deep networks (a generator and a discriminator). The two networks compete with each other in a zero-sum game framework. The generator attempts to learn a latent representation of a distribution, in a way such that the discriminator network, trained to discriminate between instances from the true data distribution and the ones produced by the generator, will have a high loss. Generative models have been applied to many tasks, e.g., image generation. For example, (Radford, Metz, and Chintala 2015) synthesized new images which resembled the database on which the models were trained on. Evaluation of these tasks is usually done by presenting the generated images to human labelers which are asked to distinguish between the GAN generated images and real images. To the best of our knowledge, few successful attempts were performed in leveraging the generated instances for improving supervised learning tasks. Specifically, we are not aware of any such attempts outside the image synthesis domain. We show empirical results that indicate that the newly proposed approach for ECG classification by generating synthetic training examples using GANs provides significant accuracy improvements.

Our contribution in this work is twofold: (1) We present an algorithm for synthetically generating ECG signals using GANs. We show that utilizing the generated instances significantly improves ECG classification using the state-of-the-art deep learning techniques. We believe this is the first attempt in applying GAN methods for ECG classification. To the best of our knowledge, this is the first application, where the instances generated by GANs have been shown to improve supervised classification tasks outside the domain of image synthesis. (2) We share the code online for further research and experimentation https://bitbucket.org/tomerGolany/ecg_dl.

2 Background

Electrocardiography (ECG) is a non-invasive tool to monitor the rhythm of the human heart. ECG is frequently applied for diagnosis and monitoring of cardiac conditions in patients. Anomalies in the heart rhythm are potentially life threatening. Thus, using ECG measurements to detect these anomalies can have a meaningful impact on the care of patients at risk of these anomalies. Using automatic methods to analyze ECG data, and specifically for anomaly detection has the potential to act as a quality control on expert diagnoses, reducing the chance of human error. Further, such methods can increase the value of home-ECG technology by providing a indication to patients on the status of their heart rhythms without having to seek expert advice. Finally, such automatic systems can improve the triage process and prioritize patients with a higher risk for abnormal heart rhythms. In each of these scenarios, such systems are potentially lifesaving.

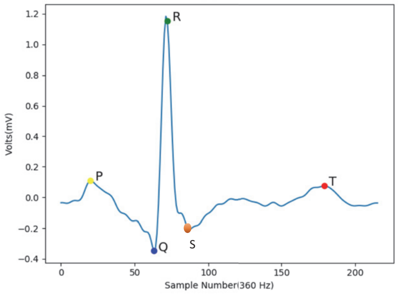


Figure 1: A normal cardiac cycle with P,Q,R,S and T waves.

The ECG signal is composed of a series of events, traditionally labeled with the letters P,Q,R,S and T (Figure 1), which correspond to the physiological stages of contraction and relaxation of the heart. While the sequence of these events is fixed the ECG signals vary across different heartbeats for the same patient. Even larger variance is observed when comparing ECG signals across different patients.

Deviations from the normal heart rhythms are known as arythmias. While arythmias are often benign, in rare cases arythmias can be life threatening.

Heartbeats and their corresponding ECG signals are classified into one of five heartbeat *types*: Normal beats (N) , Supraventricular ectopic beats (S), Ventricular ectopic beats (V) , Fusion beats (F), and Unknown beats (Q). A fusion beat is a bit different from the others as it occurs when a supraventricular beat and a ventricular beat happen simultaneously.

3 Methods

Data Sources

The ECG data used in this study are taken from MIT-BIH dataset (Moody, Mark, and Goldberger 2001). This dataset is considered the gold-standard evaluation data for ECG classification tasks. The dataset contains 48 half-hour ECG records, obtained from patients studied by the BIH Arrhythmia Laboratory between 1975 and 1979. Each record contains two 30-minutes ECG lead signals digitized at 360 samples per second. The dataset contains annotations for both heartbeat class and timing information. These annotations are verified by and independent expert. Twenty-three of the recordings are intended to serve as a representative sample of routine clinical ECG. The remaining twenty-five recordings contain complex ventricular, junctional, and supraventricular arrhythmias. Table 1 presents the five different classes of the heartbeat classes present in the dataset.

Formalization

For a patient p the ECG signal is formalized as the vector:

$$\mathbf{V}_p = [v_{p,1}, \ldots, v_{p,n}]$$

where $v_{p,t}$ denotes the measured electrical activity of the patient's heart at discrete time step t. The ECG signal is further divided into fixed length segments, referred to as heartbeats. Formally, each heartbeat of signal $\mathbf{V}_p$ is a sub-sequence of

AAMI heartbeat class	N	S	V	F	Q
Description	Any heartbeat not in the S, V, F or Q class	Supraventricular ectopic beat	Ventricular ectopic beat	Fusion beat	Unknown beat
MIT-BIH heartbeat types	normal beat (NOR) left bundle branch block beat (LBBB) right bundle branch beat block (RBBB) atrial escape beat (AE) junctional escape beat (NE)	artial premature beat (AP) aberrated atrial premature beat (aAP) junctional premature beat (NP) supraventricular premature beat (SP)	premature ventricular contraction (PVC) ventricular escape beat (VE)	fusion of ventricular and normal beat (fVN)	paced beat (P) fusion of paced and normal beat (fPN) unclassified beat (U)

Table 1: Mapping the MIT-BIH dataset heartbeat types to the AAMI heart beat classes. Given an ECG signal from a single patient the task is to classify each beat from the signal to one of the 5 different beat classes.

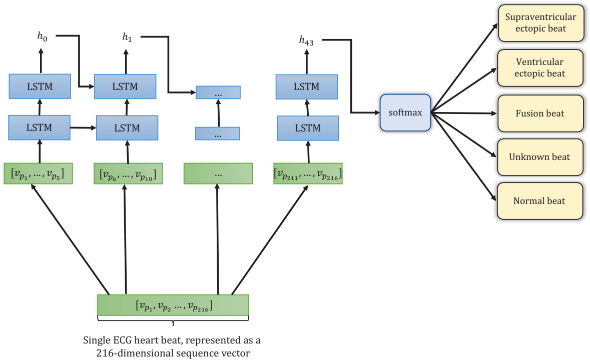

h_0
h_1
h_{43}
LSTM
LSTM
...
LSTM
LSTM
LSTM
...
LSTM
$[v_{p_1}, \dots, v_{p_5}]$
$[v_{p_6}, \dots, v_{p_{10}}]$
...
$[v_{p_{211}}, \dots, v_{p_{216}}]$
softmax
Supraventricular ectopic beat
Ventricular ectopic beat
Fusion beat
Unknown beat
Normal beat
$[v_{p_1}, v_{p_2} \dots, v_{p_{216}}]$
Single ECG heart beat, represented as a 216-dimensional sequence vector

Figure 2: LSTM architecture to classify ECG heartbeats.

length 216 (corresponding to 600 milliseconds at 360 samples per second.

A classical supervised training setup requires a dataset of labeled examples $\{\mathbf{x}_i, y_i\}$, where $\mathbf{x}_i$ denotes the *feature vector* associated with the i-th example and y_i denotes the associated label. Our problem attempts to classify a fixed length vector representing a heartbeat into one of the 5 heartbeat types. More formally, each $\mathbf{x}_i \in \mathbb{R}^{216}$ and $y_i \in \{N, S, V, F, Q\}$ for $i = 1, \dots N$. N denotes the size of the dataset.

LSTM-based Heartbeat Classification

In this work we propose a heartbeat classification approach based on neural sequence model, specifically Long Short Term Memory (LSTM) models. These models naturally consider the temporal nature of ECG signals while avoiding the need for strong modeling assumptions required by preprocessing and feature engineering approaches.

Each feature vector $\mathbf{x}_i$ is segmented into vectors of length 5. This sequence of vectors is the input into the LSTM classification network. The classification network consists of two LSTM layers. A softmax layer is applied to the last hidden output to yield a distribution over the 5 possible heartbeat types. Figure 2 visualizes the LSTM classification network architecture.

ECG Generation Using GANs

In this section we describe our approach for generating ECG signals based on Generative Adversarial Networks.

Our approach is to create a generative model for each type of heartbeat type, for a total of five models. Each such model includes a generator network and a discriminator network. The architecture of the generator and discriminator networks is the same across all heartbeat types, but each has unique values for the parameters. A generator network takes as input a random latent vector and outputs a synthetic ECG signal. A discriminator network takes as input an ECG signal and outputs a binary decision representing whether the input signal came from collected training data or was generated by a generator network.

More formally, let us a consider the architecture for a particular heartbeat type. We denote the generator network as $G(z; \theta_G) : \mathbb{R}^d \rightarrow \mathcal{X}$ and the discriminator network as $D(\mathbf{x}; \theta_D) : \mathcal{X} \rightarrow [0, 1]$, where $\mathcal{X}$ denotes the domain of ECG signals, and d is the dimension of the latent vector input to the generator. θ_G and θ_D, represent the parameters of the Generator and Discriminator networks, respectfully.

Using this notation, consider the following objective:

$$V(\theta_G, \theta_D) = \mathbb{E}_{\mathbf{x} \sim p_{data}} [\log D(\mathbf{x}; \theta_D)] + \tag{1}$$

$$\mathbb{E}_{\mathbf{z} \sim p_{noise}} [1 - \log D(G(\mathbf{z}; \theta_G); \theta_D)] \tag{2}$$

where p_{data} denotes the (unknown) distribution over the training data, and p_{noise} denotes some known noise distribution. $\mathbb{E}[\cdot]$ denotes expected value of a random variable.

The discriminator network objective is to *maximize* this expression with respect to parameters θ_D. The generator network objective is to *minimize* this objective with respect to parameters θ_G.

GAN Model Architecture Design

The network architecture for each of our five generative models is identical. We describe the network architecture for one of these models without loss of generality. The architecture choices are inspired by the DCGAN architecture.

The generator network architecture is designed to create an ECG heartbeat signal from a latent vector representation. This input vector is fed into seven 1d transpose convolution layers. Between each layer we apply batch normalization (Ioffe and Szegedy 2015) followed by a Rectified Linear Unit (ReLU) activation function. The final output of the network is a vector of length 216, the length of the ECG heartbeat signal. Figure 3 illustrates the generator architecture.

The discriminator network receives as input an ECG signal and outputs a binary decision. The input signal (a 216 × 1 ECG heartbeat signal) is fed through six convolutional layers. Between each layer we perform batch normalization and apply a LeakyReLU (Maas, Hannun, and Ng 2013) activation function, where the slope of the leak is set 0.2. The final activation function is a sigmoid activation. Figure 4 visualizes the architecture of the discriminator network.

4 Experimental Setup

LSTM classification network: Network consists of two LSTM layers with 512 hidden layer units. We use the *RMSprop* optimizer with learning rate of 0.001 to learn network

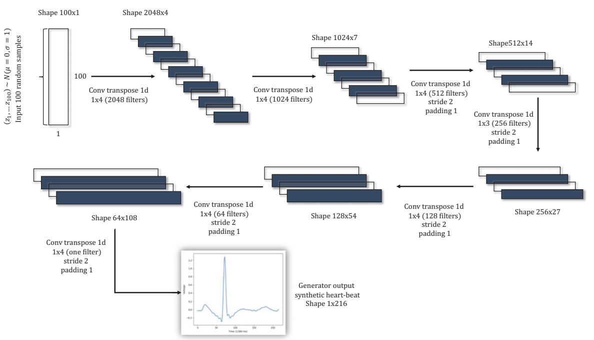


Figure 3: ECG Generator Architecture.

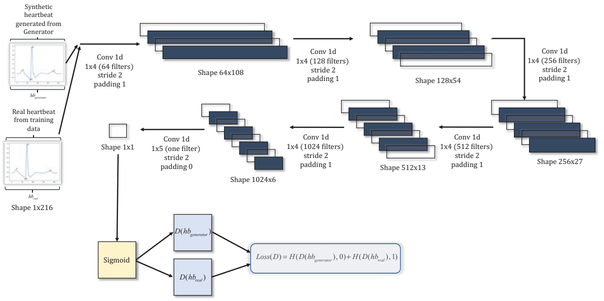


Figure 4: ECG Discriminator Architecture.

weights. The network weights are initialized with Xavier uniform initializer (Glorot and Bengio 2010).

Generator and Discriminator networks: All weights are initialized from a 0-mean Gaussian distribution with standard deviation 0.02. We set the hyper parameters as suggested in (Radford, Metz, and Chintala 2015). The optimization is performed using Adam optimizer (Kingma and Ba 2014) and a learning rate of 0.0002. Data Processing We process the dataset describe in Section 3 as follows: in agreement with the AAMI recommended practice, the four recordings containing paced beats were removed from the dataset. We apply heartbeat detection to obtain subsequences corresponding to a total of $109,492$ labeled heartbeats. We follow AAMI recommendations for dataset partitioning which (De Chazal, O'Dwyer, and Reilly 2004; Al Rahhal et al. 2016) ensure that patient data is not mixed between the training and the test sets. The training dataset contains $51,020$ unique heartbeats , while the test set contains $49,711$ unique heartbeats. Table 2 presents the statistics of the entire dataset and the test-train division.

Training procedure

Generator and Discriminator networks For each heartbeat class we train our GAN framework using the 2000 update iterations over the training set described in the previous section. In each update iteration, the training procedure maximizes (minimizes) Equation 2 w.r.t. parameters θ_D (θ_G). In particular, first update the parameter weights of the generator (θ_G) twice, and then the parameter weights of the discriminator (θ_D) once. The mini-batch size used for each iteration is 50. That is, at each iteration 50 random vectors are fed to the generator to output 50 fake heartbeats, and 50 real heartbeats are sampled from the training set. The loss function is optimized using Adam optimizer (Kingma and Ba 2014) with learning rate of 0.0002.

Heart beat class	N	S	V	F	Q	total
Total number	90632	2779	7235	803	8043	109492
DS1 Training set	45868	942	3787	415	8	51020
DS2 Test set	44258	1837	3221	388	7	49711

Table 2: Partition of the MIT-BIH dataset to Train and Test. DS1 is used for training and DS2 is used for testing. DS1 comprises data from 22 patients. DS2 comprises data from 22 other patients. 4 patients were not included in dataset 1 or 2.

The expectations in Equation 2 are reconciled by sampling. Uniform sampling of examples from the training data simulates p_{data}, while for p_{noise} we choose a multivariate Gaussian distribution $z \sim N(\mu, \sigma)$ with $\mu = 0$ and $\sigma = 1$ (chosen to allow straightforward sampling).

5 Results

In this section we present the results of our empirical evaluation of the main research questions of this study: (1) The applicability of sequence models to the problem of ECG classification; (2) The impact of augmenting training data with synthetic examples generated using GANs trained on the original data; (3) Are synthetic ECG signals generated by GANs trained on real signals using our approach qualitatively similar to real ECG signals.

Using Sequence Models for ECG Classification

We first recreate the results by (Hou et al. 2019) on our data and show it indeed receives superior results to the previous state-of-the-art models. We report the results of applying the LSTM architecture proposed in Section 3 to the task of ECG classification. As a baseline for this evaluation, we used a recently published deep neural network approach (Al Rahhal et al. 2016) which does not (explicitly) take sequence information into account (a fully-connected architecture). This approach achieves state of the art performance on this problem. We observe that for all heart-beat types, applying a sequence based approach significantly improves AUC on our our test set. We attribute this improvement to the inherent ability of the LSTM architecture towards considering the temporal structure of the signal. The fully connected architecture, does not inherently consider this structure, and thus ,when such structure is informative (as is clearly the case in ECG), suffers from degraded performance.

Increasing Training Size using Synthetic Examples

We evaluate the improvement achieved by augmenting the training data for the ECG classification network by adding

Heartbeat Class	Fully connected Network	LSTM network
N	0.85	**0.87**
S	0.81	**0.82**
V	0.95	**0.97**
F	0.5	**0.95**

Table 3: Results of classifying MIT-BIH ECG data over five classes. Statistically significant results are shown in bold.

additional examples synthetically generated from using the GAN framework described in Section 3. We evaluate the impact of this augmentation on two classification architectures: a state of the art fully-connected architecture (Al Rahhal et al. 2016) and the sequence architecture described in Section 3. Figure 5, shows the AUC performance of the LSTM and Feed-Forward networks respectively as a function of the number of synthetic DCGAN examples added to the training set of the relevant model. We observe no monotonic behavior in the performance as a function of the number of synthesized examples added. However, we observe that both the LSTM model and the fully-connected model trained with added synthetic examples from our DCGAN model significantly outperforms the models trained without synthetic examples added. We conclude that the practice of adding synthesized examples to the training of the LSTM model and to the fully-connected model significantly improves its performance for heartbeat classification when tuning the number of synthesized examples added. The DCGAN model with LSTM architecture outperforms for all heartbeat classes.

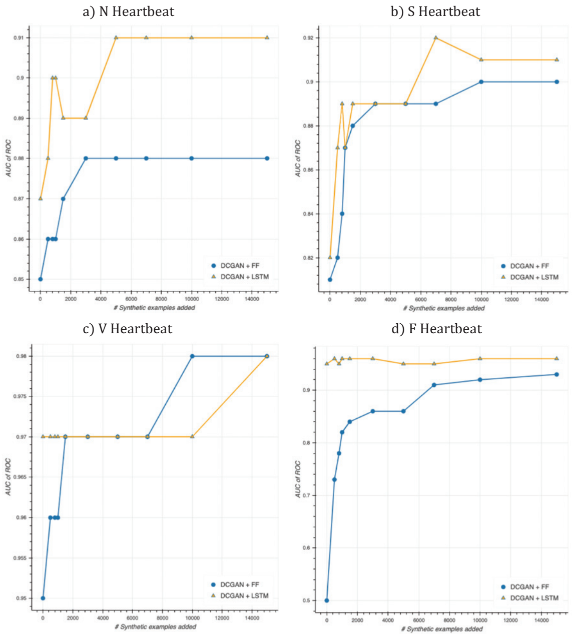


Figure 5: Average AUC comparison for each type of heart beat. We compare using LSTM (yellow) and fully connected networks (blue) with additional synthetic training data produced from the ECG DCGAN.

Heartbeat Class	No Added Examples	Jittering Examples	DCGAN Examples
N	0.87	0.76	**0.91**
S	0.82	0.73	**0.92**
V	0.97	0.92	**0.98**
F	0.95	0.88	**0.96**

Table 4: Results over five classes using LSTM network. We compare adding no synthetic data to two methods for adding synthetic training data: (1) adding additional data from jittering (2) DCGAN-generated data. Statistically significant results are shown in bold.

As increasing the training set size using GAN samples is a type of data augmentation, we evaluate its performance against the well known data augmentation approach of ”jittering”. We experimented with several noise distributions (Gaussian and Uniform) and augmented data sizes for jittering and generated the noise following best practices (An 1996), (Holmstrom and Koistinen 1992). Table 4, provides a comparison of the best jittering augmentation methods and the GAN based augmentation we propose in this work.

Examining Table 4 gives evidence supporting our hypothesis. In all cases the performance of both sequence and fully-connected architectures improves when samples from the GAN are added to the training set. This improvement is beyond the improvement that can be obtained with (the best setting of) a baseline data augmentation approach (jittering).

Comparison of Generated Examples to Real Examples

In lieu of a quantitative analysis of the synthetic ECG signals generated by our trained GANs, we consider several anecdotal examples and compare them to real ECG heartbeats taken from the training set. Figures 6,7 and 8 present these comparisons.

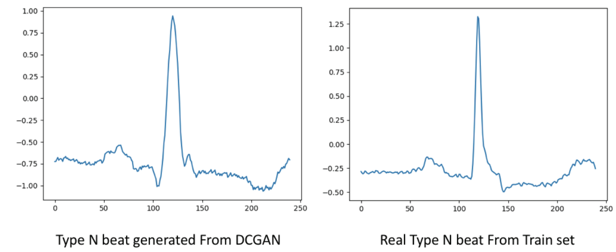
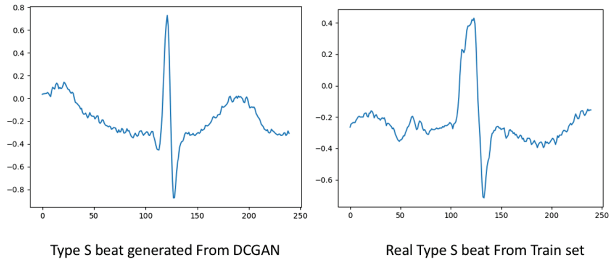


Figure 6: Comparison between DCGAN generation of N heartbeats against real N heartbeats from the training set.

We observe that samples produced from the GAN are similar to real ECG signals in the training data. We Observe that for all type of heartbeats our GAN learns to generate synthetic heartbeats which have the same PQRST morphology as their corresponding real heartbeats from the training set.

6 Conclusion

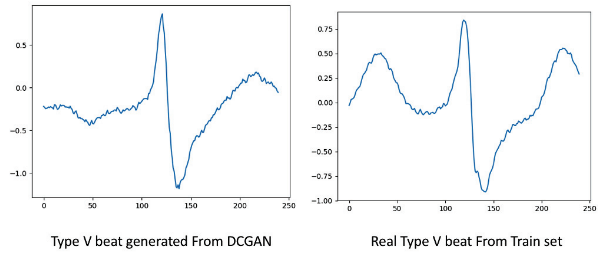
In this work, we consider the problem of ECG classification. We leverage a sequence model architecture to capture the inherent temporal structure in ECG signals (Hou et al. 2019). We propose a Generative Adversarial Network (GAN) architecture that can be trained to generate ECG signals which are



Type S beat generated From DCGAN

Real Type S beat From Train set

Figure 7: Comparison between DCGAN generation of S heartbeats against real S heartbeats from the training set.



Type V beat generated From DCGAN

Real Type V beat From Train set

Figure 8: Comparison between DCGAN generation of V heartbeats against real V heartbeats from the training set.

similar to real ECG signals observed in the data. We show that the performance of sequence based models is significantly improved by using synthetic examples, generated using GANs, to augment the labeled training data.

These results alongside further research improve ECG classification have the potential to enable automated ECG analysis technology. This type of technology can support applications such as quality control on expert diagnoses and home-ECG kits.

References

Al Rahhal, M. M.; Bazi, Y.; AlHichri, H.; Alajlan, N.; Melgani, F.; and Yager, R. R. 2016. Deep learning approach for active classification of electrocardiogram signals. *Information Sciences* 345:340–354.

An, G. 1996. The effects of adding noise during backpropagation training on a generalization performance. *Neural Computation* 8(3):643–674.

Chazal, P., and Reilly, R. 2007. A patient-adapting heartbeat classifier using ecg morphology and heartbeat interval features. 53:2535–43.

De Chazal, P.; O'Dwyer, M.; and Reilly, R. B. 2004. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *IEEE transactions on biomedical engineering* 51(7):1196–1206.

Escalona-Moran, M.; Soriano, M.; Fischer, I.; and R Mirasso, C. 2014. Electrocardiogram classification using reservoir computing with logistic regression. 19.

Glorot, X., and Bengio, Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249–256.

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

Holmstrom, L., and Koistinen, P. 1992. Using additive noise in back-propagation training. *IEEE Transactions on Neural Networks* 3(1):24–38.

Hou, B.; Yang, J.; Wang, P.; and Yan, R. 2019. Lstm based auto-encoder model for ecg arrhythmias classification. *IEEE Transactions on Instrumentation and Measurement*.

Ioffe, S., and Szegedy, C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Kass, R., and Clancy, C. 2006. *Basis and Treatment of Cardiac Arrhythmias*. Handbook of Experimental Pharmacology. Springer Berlin Heidelberg.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kiranyaz, S.; Ince, T.; and Gabbouj, M. 2016. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering* 63(3):664–675.

Maas, A. L.; Hannun, A. Y.; and Ng, A. Y. 2013. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, 3.

Moody, G. B.; Mark, R. G.; and Goldberger, A. L. 2001. Physionet: a web-based resource for the study of physiologic signals. *IEEE Engineering in Medicine and Biology Magazine* 20(3):70–75.

Radford, A.; Metz, L.; and Chintala, S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Ye, W.; Kumar, B.; and Coimbra, M. 2012. Heartbeat classification using morphological and dynamic features of ecg signals. 59:2930–41.