



UNIVERSITY OF LEEDS

This is a repository copy of *The Effect of Data Augmentation on Classification of Atrial Fibrillation in Short Single-Lead ECG Signals Using Deep Neural Networks*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/176402/>

Version: Accepted Version

---

**Proceedings Paper:**

Hatamian, FN, Ravikumar, N, Vesal, S et al. (3 more authors) (2020) The Effect of Data Augmentation on Classification of Atrial Fibrillation in Short Single-Lead ECG Signals Using Deep Neural Networks. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 04-08 May 2020, Barcelona, Spain. IEEE , pp. 1264-1268. ISBN 978-1-5090-6632-2

<https://doi.org/10.1109/icassp40776.2020.9053800>

---

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# The Effect of Data Augmentation on Classification of Atrial Fibrillation in Short Single-Lead ECG Signals Using Deep Neural Networks

Faezeh Nejati Hatamian<sup>\*</sup>, Nishant Ravikumar<sup>†</sup>, Sulaiman Vesal<sup>†</sup>, Felix P. Kemeth<sup>\*</sup>, Matthias Struck<sup>\*</sup> and Andreas Maier<sup>†</sup>

<sup>\*</sup> *Department of Image Processing and Medical Technology, Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany*

<sup>†</sup> *Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany*

<sup>‡</sup> *Centre for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), School of Computing,*

*LICAMM Leeds Institute of Cardiovascular and Metabolic Medicine, School of Medicine, University of Leeds, United Kingdom*

*Email: andreas.maier@fau.de*

**Abstract**—Cardiovascular diseases are the most common cause of mortality worldwide. Detection of atrial fibrillation (AF) in the asymptomatic stage can help prevent strokes. It also improves clinical decision making through the delivery of suitable treatment such as, anticoagulant therapy, in a timely manner. The clinical significance of such early detection of AF in electrocardiogram (ECG) signals has inspired numerous studies in recent years, of which many aim to solve this task by leveraging machine learning algorithms. ECG datasets containing AF samples, however, usually suffer from severe class imbalance, which if unaccounted for, affects the performance of classification algorithms. Data augmentation is a popular solution to tackle this problem.

In this study, we investigate the impact of various data augmentation algorithms, e.g., oversampling, Gaussian Mixture Models (GMMs) and Generative Adversarial Networks (GANs), on solving the class imbalance problem. These algorithms are quantitatively and qualitatively evaluated, compared and discussed in detail. The results show that deep learning-based AF signal classification methods benefit more from data augmentation using GANs and GMMs, than oversampling. Furthermore, the GAN results in circa 3% better AF classification accuracy in average while performing comparably to the GMM in terms of f1-score.

**Keywords**—atrial fibrillation, data augmentation, GMM, DCGAN

## I. INTRODUCTION

Based on a report released by the World Health Organization (WHO), ischemic heart disease, also known as coronary artery disease, were globally acknowledged as one of the most common causes of death in 2016 [1]. Previous studies have shown that AF increases the risk of embolic stroke by five times [2]. Moreover, detection of atrial fibrillation (AF) in the asymptomatic stage can help prevent strokes and aid in the timely identification of patients that are likely to benefit from anticoagulant therapy [3] [4].

In a healthy heart, the sinoatrial (SA) node, which works as the natural pacemaker, generates electric signals in a periodic pattern. These signals result in the periodic contraction of the heart's atria and ventricles. These consecutive contractions translate into the PQRST peaks in an electrocardiogram (ECG) signal. Malfunction in the SA node may cause the atria to quiver instead of performing a complete contraction. This can potentially result in blood remaining in the atrial chamber, followed by the formation of clots, and ultimately leading to ischemia and strokes [5], [6]. AF is reflected in the missing of the P-peak and its substitution with an inconsistent and chaotic fibrillatory wave (f-wave) and an irregular R-peak to R-peak (R-R) interval in an ECG signal [5].

The existence of a distinct pattern in ECG signals with AF disorder makes machine learning an appropriate tool for the task of distinguishing the normal vs. AF signals. The possibility of automatically detecting AF in ECG signals along with the potential clinical impact of doing so, have led to the PhysioNet/CinC challenge organized in 2017, which focused on the classification of ECG signals with AF anomaly [7, 8]. Numerous

studies have adopted the traditional machine learning pipeline using handcrafted features and reported models with accurate results. In these approaches, expert knowledge of cardiologists was taken into account to calculate representative and distinctive features. These works either focus on detecting the presence of the f-wave instead of the P-peak [9] or are based on anomaly detection in the R-R intervals in ECG signals [10].

Hand crafting descriptors and engineering features, however, are very expensive, time-consuming and tedious. Moreover, they demand the availability of expert knowledge. Convolutional Neural Networks (CNNs) provide the means to automate this feature extraction process without the need for expert domain knowledge. Moreover, the hierarchical data processing nature of CNNs produces highly descriptive and informative features. Deep CNNs have shown competitive performance in different computer vision and medical applications and in some cases matched or even surpassed the human accuracy [11].

Deep CNNs have also been adopted for ECG signal classification. Rajpurkar et al. [12] proposed a 34-layer CNN for classifying 12 cardiac disorders by mapping the ECG signals to their corresponding rhythmic disorder classes. Andreotti et al. [13] used ResNets [14] for AF detection in ECG signals. The residual connections in such networks help tackle the issue of vanishing gradients, enabling deeper networks to be trained. An alternative to the use of CNNs for supervised feature learning, and the use of hand-crafted features, was proposed in [15]. In the representation block, they feed the raw ECG signals to a stacked denoising autoencoder and train it in an unsupervised manner. The result is then fed into a softmax regression layer. Authors in [16] use the famous AlexNet for feature extraction. They fed AlexNet with R-T segments, which are extracted from the ECG signals, and direct the result into a 2-layer feed-forward neural network for detecting normal beats, paced beats and Rbbb (Right Bundle Branch Blocks) beats. The work of Ziehlmann et al. [17] achieved one of the best classification performances in the PhysioNet 2017 challenge [8]. Their best model is a 24-layer CNN followed by an LSTM. The input to the network is the logarithmic spectrogram of the ECG signals. In a similar work, Chauhan et al. [18] proposed to stack multiple LSTM layers for the task of cardiovascular anomaly detection from ECG signals.

Previous studies on ECG signal classification using deep learning focus on proposing new and powerful architectures that can better handle the aforementioned task. However, to our knowledge, there is neither a universal agreement nor a comprehensive and thorough study on the most effective data augmentation algorithm to be employed to address the issue of the class imbalance. In this work, we aim to close this gap. The rest of this paper is organized as follows: In section II we go over our classification pipeline, discuss our contribution and

review the theoretical background of the algorithms we employ. Section III is dedicated to describing the experimental setup and summarizing the results. Finally, section IV concludes the paper.

## II. PROCESSING FRAMEWORK AND THEORETICAL BACKGROUND

### A. Classification Pipeline

The majority of previous studies employing deep learning for ECG signal classification, follow a common pipeline [17, 18]: First the 1-D signals are pre-processed. The ECG signals can be segmented into equal-size segments. Moreover, it is common to transform the 1-D signals to 2-D logarithmic spectrograms. In the next step, various data augmentation schemes are typically employed to reduce class imbalance. It is worth noting that not all studies employ data augmentation. Finally, the pre-processed and augmented data is fed into the deep network architecture.

In this work, we focus on the data augmentation block and evaluate various data augmentation schemes, in order to identify the most suitable ones to enhance classification performance. In the literature and within this community, oversampling is the most common data augmentation algorithm. In this paper, we introduce and study the effectiveness of two more augmentation algorithms, namely Gaussian Mixture Model (GMM) and Generative Adversarial Networks (GANs).

### B. Gaussian Mixture Model (GMM)

1) *Gaussian Probability Distribution Function*: Many random events in nature follow Gaussian probability distribution functions (pdf). A Gaussian pdf is parameterized by two parameters, i.e., the mean and the variance. For one dimensional data  $\mathbf{X}$ , the Gaussian pdf is represented by  $\mathcal{N}(\mu, \sigma)$  as in Eqn. 1.

$$\mathcal{N}(\mathbf{X}|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mathbf{X} - \mu)^2}{2\sigma^2}\right). \quad (1)$$

However, for  $d$ -dimensional data  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_d]$ , the mean is a vector  $\boldsymbol{\mu}$  and the variance becomes a covariance matrix  $\boldsymbol{\Sigma}$ . The multivariate Gaussian pdf in this case is mathematically expressed as in Eqn. 2.

$$\mathcal{N}(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right), \quad (2)$$

where  $|\boldsymbol{\Sigma}|$  represents the determinant of the covariance matrix. Thus, for this pdf to be well defined, the covariance matrix  $\boldsymbol{\Sigma}$  needs to be positive-definite.

2) *GMM*: Despite the diversity of the Gaussian pdf, not all the processes follow the Gaussian distribution. Such arbitrary data distributions can, however, be approximated using a GMM, i.e., by the weighted sum of more than one Gaussian component, as in Eqn. 3:

$$p(\mathbf{X}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, w_i) = \sum_{i=1}^c w_i \mathcal{N}(\mathbf{X}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad (3)$$

where  $\boldsymbol{\mu}_i$ ,  $\boldsymbol{\Sigma}_i$  and  $w_i$  represent the mean, covariance and mixture weight of the  $i$ -th component. Furthermore, mixture weights  $w_i$  need to sum up to one. Given the data  $\mathbf{X}$ , the parameters of a GMM are usually estimated via the expectation maximization (EM) algorithm in an iterative manner.

### C. Generative Adversarial Networks (GANs)

Generative adversarial networks (GANs) [19] are composed of two sub-networks, i.e., a generator and a discriminator. The generator takes a random noise vector  $z$  initially as input, which represents the latent space to be learned, and learns the mapping

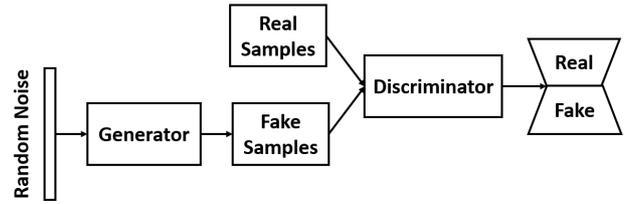


Figure 1: High level GAN architecture.

between this latent vector  $z$  and the data distribution. The discriminator plays the role of an investigator, who determines the legitimacy of the data generated by the generator from the learned latent representation. The main objective of the generator is to fool the discriminator. These two networks are trained alternatively, which makes them grow together and therefore, compete. The high level GAN architecture is depicted in Fig. 1.

GAN's training, however, strongly depends on the two competing networks, i.e., the generator and the discriminator, and can be highly unstable. Convolutional neural networks (CNNs) have shown great performance in supervised learning [20]. Radford et al. [21] proposed to combine CNNs and GANs and called the final framework Deep Convolutional GAN (DCGAN). In particular, they eliminated all the fully connected layers and substituted them by convolutional layers in the discriminator and transposed convolutional layer in the generator. Moreover, max-pooling and upsampling are replaced by strided convolution and strided transposed convolution, respectively. In addition, in the discriminator all the activation functions are LeakyReLU. In the generator, all the activation functions are ReLU except for the last layer which uses tanh as the activation function. Finally, for further stabilizing the training, Batch Normalization (BN) is used after all layers except the output of the generator and the input of the discriminator.

## III. EVALUATION

### A. Dataset

1) *Data Statistics*: The data used in this study is publicly available from the PhysioNet/CinC challenge 2017 [7, 8]. It comprises 8528 single-lead ECG signals in four classes, namely 5154 Normal (N), 771 atrial fibrillation (AF), 2557 Noisy and 46 Other rhythm signals. Following the main objective in this work, we kept the Normal and AF classes and discarded the Noisy and Other rhythm ones. The ECG signals' recording times vary between 9 and 61 seconds, which are sampled at the rate of 300 Hz. The majority of the signals in the provided dataset is 30s long, while a smaller subset is 60s long.

2) *Data Segmentation*: As different ECG signals in the dataset have different length and the majority of deep classifiers require the samples to have equal size, it is common to segment the signals into smaller segments with equal sizes [22, 23]. In order to do so, a segment length (SL) is chosen. In particular, ECG signals longer than the chosen SL are cropped, whereas signals shorter than the SL are discarded. In this work, we follow [24] and use SL=1500 being equivalent to 5 seconds.

3) *Class Imbalance and Data Augmentation*: Using SL=1500, AF class has 4674 signals and Normal class has 31939 signals that is seven times the AF class, which indicates the presence of significant class imbalance. This is undesirable in supervised learning as it would push the classifier towards being biased on the larger class. In order to tackle this problem, we use data augmentation on the AF class. The data augmentation algorithms that we use in this paper are oversampling, GMM

and DCGAN and we compare them with no data augmentation, referred to as None. When using GMM and DCGAN for data augmentation, we follow Davari et al. [25]. That means, we train these generative models using only the AF class samples, draw virtual samples from the trained models and combine them with the original samples so that the number of AF class samples equals the Normal class samples.

For training the DCGAN and GMM, we tried both the 1-D and 2-D (logarithmic spectrogram) variants of the ECG signals. DCGAN performed better when trained using the 2-D spectrograms, while GMM performed better when trained on the 1-D signals. The GANs have been shown to have high capacity in producing high quality images [21]. However, GMMs are unable to accurately capture spatial correlations in 2-D images and are used to generate 1-D signals [25].

#### 4) Division of Data into Train, Validation and Test Subsets:

In order to divide our data into training set, validation set and the test set, we opt for the following policy: We take 20% of whole dataset as the test set. Then we divide the remaining 80% into 90% training and 10% validation. We want to preserve the original data statistics and keep the imbalanced class ratio during the division. Therefore, in order to take  $x\%$  from the data, we take  $x\%$  from the AF class and  $x\%$  from the Normal class and then concatenate them.

5) *2-D Data Representation*: In order to convert our 1-D signals to 2-D using logarithmic spectrogram, we use the *signal.Spectrogram* module from the Scipy toolbox [26] in python. As for the spectrogram’s hyper-parameters, we follow Zihlmann et al. [17], i.e., Tukey window of length 64 and hop length of 32 (i.e., 50% window overlap), and shape parameter of 0.25.

## B. Deep Learning Architectures

1) *Classification Architectures*: For the evaluation purposes in this paper, we use one of the leading deep learning-based works in the PhysioNet challenge 2017, by Zihlmann et al. [17]. They proposed two different architectures, which are all adapted towards the usage of 2-D spectrogram representation of the ECG signals. SOA\_CNN, which is depicted in Fig. 2a, comprises six convolutional blocks. Each block consists of four sub-blocks with a convolutional layer with  $5 \times 5$  kernel size, then a batch normalization followed by ReLU activations and a dropout layer. The last sub-block contains a max-pooling layer. The first three convolutional layers have 64 filters while, the last convolutional layer starts with 64 filters in the first block, with the filter number being increased by 32 for each subsequent block. The output of the convolutional blocks is then averaged on the temporal axis, flattened and fed into a sigmoid classifier.

Their next proposed architecture, SOA\_CRNN, is very similar to SOA\_CNN. Their differences boil down to the usage of 4 convolutional blocks instead of 6, and the substitution of the temporal averaging layer by an LSTM layer with 200 neurons. This network is illustrated in Fig. 2b. They trained SOA\_CRNN in three phases. In the first phase, they replaced the LSTM layer by temporal averaging, and trained the resulting network. Then, they substituted the temporal averaging with an LSTM layer and trained this layer while keeping the rest of the network frozen. At this point, the convolutional blocks and the LSTM layer are trained independently. In the final training phase, the whole network is trained. This training policy effectively initializes parts of the network far better than random weight initialization.

2) *DCGAN Architecture*: We investigated different architectures for the generator and discriminator in the DCGAN. The best performing combination is depicted in Fig. 3. In our imple-

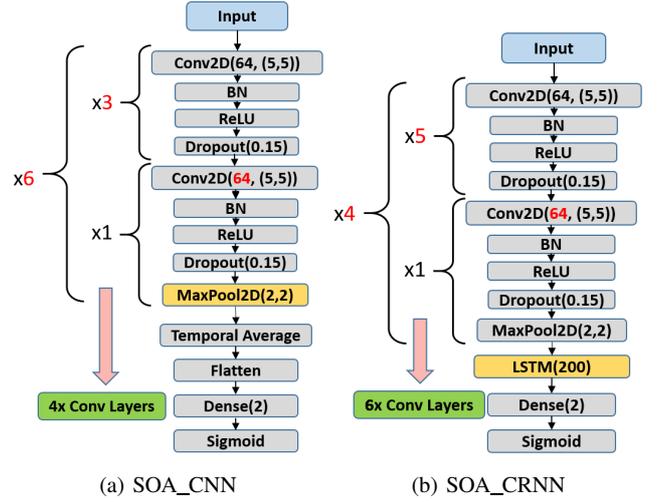


Figure 2: Zihlmann et al. [17] architectures: (a) SOA\_CNN, (b) SOA\_CRNN.

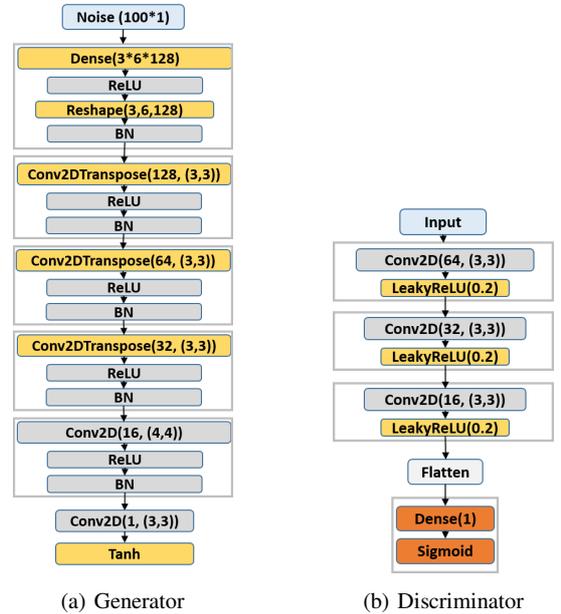


Figure 3: The proposed DCGAN generator and discriminator architectures.

mentation of the DCGAN architecture, we follow the general tips and tricks that are suggested in [21] and explained in Sec. II-C.

3) *Hyper-parameters*: As preprocessing, we apply mean-std normalization on the data. That means, we subtract the mean of the training data from all data subsets, i.e., training, validation and test subsets. Then we divide the result by the standard deviation (std) of the training data. The mean of the resulting data is zero and standard deviation is one. For training the deep neural networks, we use the ADAM optimizer with learning rate equal to  $10^{-3}$  and decay rate of  $10^{-5}$ , and run them for 150 epochs. However, while training the generator and the discriminator of the DCGAN, we use a learning rate equal to  $2 \times 10^{-4}$ , as suggested in [21]. We use binary cross entropy (BCE) loss function for training all networks. As for the weight initialization in the network, we use Xavier uniform initialization technique, also known as Glorot uniform.

We implemented all the neural network architectures using

Table I: Quantitative results associated with AF class augmentation using different data augmentation algorithms. F1 stands for f1-score, AF stands for AF class accuracy and N represents the Normal class accuracy.

Classifier	None			oversampling			GMM			DCGAN		
	F1[%]	AF[%]	N[%]	F1[%]	AF[%]	N[%]	F1[%]	AF[%]	N[%]	F1[%]	AF[%]	N[%]
SOA_CNN	84.47	75.61	99.50	85.91	79.25	99.23	<b>87.16</b>	85.67	98.40	86.77	<b>86.31</b>	98.15
SOA_CRNN	86.02	81.60	98.81	85.81	80.86	98.89	<b>87.79</b>	82.67	99.17	86.90	<b>88.34</b>	97.81

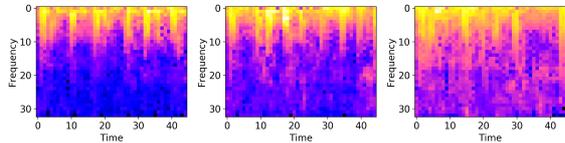


Figure 4: Sample DCGAN-generated AF ECG spectrograms. The DCGAN is trained on spectrograms of 5 seconds long signals. Thus, it is expected to see 4-7 peaks.

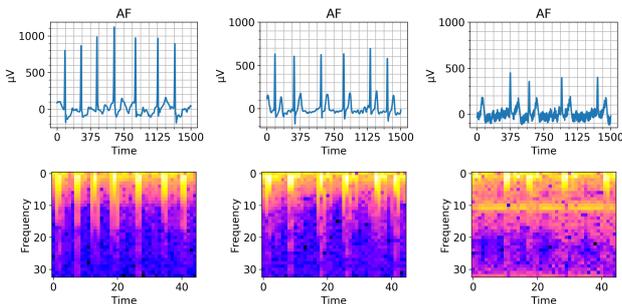


Figure 5: Original AF signals from the training set (first row) and their corresponding logarithmic spectrogram (second row).

the Keras toolbox with Tensorflow backend [27] in Python. Apart from the previously explained hyper-parameters, we used the default hyper-parameters of Keras. We refer the reader to the Keras documentation [27] for detail set of parameter values for reproduction. Furthermore, in the network architecture visualizations in this work, we use the Keras syntax. For example for a 2-D convolution, we may write  $\text{Conv2D}(a,(b,c))$ , which  $a$  represents the number of convolutional filters and  $(b,c)$  represents the convolutional kernel size.

### C. Experimental Results

Table I presents the classification results that are computed using different data augmentation algorithms, namely oversampling, GMM and DCGAN. For the sake of comparison, we report the results without augmentation (None) as well. It can be observed that using data augmentation always outperforms no augmentation with respect to f1-score. Another interesting observation is that in the majority of the cases, applying augmentation on the AF class results in an improvement in both the AF accuracy and the total f1-score. However, in some cases the Normal class accuracy slightly deteriorates ( $\approx 1\%$ ). Among the augmentation algorithms investigated in this study, the lowest improvement in performance is achieved by oversampling, while GMM and DCGAN enhance the performance the most. Although GMM marginally outperforms DCGAN in terms of the overall f1-score, it turns out that DCGAN results in better AF accuracy while producing comparable Normal class accuracy to GMM.

Figure 4 depicts some spectrograms that are generated by the DCGAN’s generator. For the sake of visual comparison, Fig. 5 illustrates three real AF ECG signals and their corresponding logarithmic spectrograms.

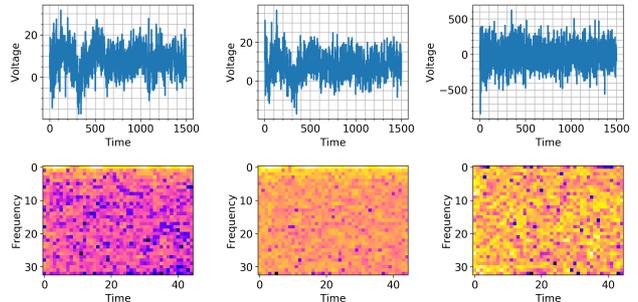


Figure 6: Sample GMM-generated AF ECG spectrograms.

Figure 6 illustrates the logarithmic spectrogram of two samples that are generated by a GMM, which is trained on the AF class. As it can be observed, these samples hardly resemble an ECG signal visually. However, as GMM resulted in high classification performance improvement, it certainly should have learned the underlying distribution of the AF class distinctive characteristic.

## IV. CONCLUSION AND DISCUSSION

Data augmentation is a popular technique to solve the data imbalance problem in supervised learning. In this work, we have investigated the impact of various data augmentation algorithms, i.e., oversampling, GMM and DCGAN, on the short single-lead ECG signal classification into AF and Normal classes. The results show that in all cases, data augmentation comes with performance improvement. Oversampling performs comparable to, if not better than, using no augmentation. We should state that we did not expect oversampling to significantly increase the performance because it does not add extra information to the dataset as it simply is a duplication of data. However, it still balances the classes and prevents the classifier to be biased towards the bigger class. On the other hand, the deep learning-based methods seem to really benefit from the DCGAN and GMM. Further, GAN results in better AF accuracy while GMM results in better f1-scores using both classification networks investigated. Both GMM and DCGAN augmentation schemes perform comparably in terms of the Normal class accuracy.

In some cases, using GAN and GMM to augment the AF class causes slight deterioration of the Normal class accuracy. It would be interesting to tackle this problem by applying the learning-based augmentation algorithms, e.g., GMM or GAN, on the Normal class as well. In this manner, the underlying distribution of the Normal class can be learned and this information can be included in the data and exploited by the classifier. Furthermore, other more recent variants of GAN, e.g., improved Wasserstein GAN, outperform DCGAN. In our future work, we intend to conduct a thorough study on the effect of different GAN variants on this application.

## REFERENCES

- [1] “World Health Organization (WHO), the top 10 causes of death,” <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>, last visited: 08.09.2019.

- [2] James A Reiffel, "Atrial fibrillation and stroke: epidemiology," *The American journal of medicine*, vol. 127, no. 4, pp. e15–e16, 2014.
- [3] Emelia J Benjamin, Salim S Virani, Clifton W Callaway, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Stephanie E Chiuve, Mary Cushman, Francesca N Delling, Rajat Deo, et al., "Heart disease and stroke statistics-2018 update: a report from the american heart association.," *Circulation*, vol. 137, no. 12, pp. e67, 2018.
- [4] Christine Benn Christiansen, Thomas A Gerds, Jonas Bjerring Olesen, Søren Lund Kristensen, Morten Lamberts, Gregory YH Lip, Gunnar H Gislason, Lars Køber, and Christian Torp-Pedersen, "Atrial fibrillation and risk of stroke: a nationwide cohort study," *Ep Europace*, vol. 18, no. 11, pp. 1689–1697, 2016.
- [5] S Levy, G Breithardt, RWF Campbell, AJ Camm, J-C Daubert, M Allessie, E Aliot, A Capucci, F Cosio, HJGM Crijns, et al., "Atrial fibrillation: current knowledge and recommendations for management," *European heart journal*, vol. 19, no. 9, pp. 1294–1320, 1998.
- [6] K Padmavathi and K Sri Ramakrishna, "Classification of ecg signal during atrial fibrillation using autoregressive modeling," *Procedia Computer Science*, vol. 46, pp. 53–59, 2015.
- [7] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [8] Gari D Clifford, Chengyu Liu, Benjamin Moody, H Lehman Li-wei, Ikaro Silva, Qiao Li, AE Johnson, and Roger G Mark, "Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [9] Manuel García, Juan Ródenas, Raúl Alcaraz, and José J Rieta, "Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation," *computer methods and programs in biomedicine*, vol. 131, pp. 157–168, 2016.
- [10] David T Linker, "Accurate, automated detection of atrial fibrillation in ambulatory recordings," *Cardiovascular engineering and technology*, vol. 7, no. 2, pp. 182–189, 2016.
- [11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.
- [12] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.
- [13] Fernando Andreotti, Oliver Carr, Marco AF Pimentel, Adam Mahdi, and Maarten De Vos, "Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ecg," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [15] Mohamad Mahmoud Al Rahhal, Yakoub Bazi, Haikel AlHichri, Naif Alajlan, Farid Melgani, and Ronald R Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.
- [16] Ali Isin and Selen Ozdalili, "Cardiac arrhythmia detection using deep learning," *Procedia computer science*, vol. 120, pp. 268–275, 2017.
- [17] Martin Zihlmann, Dmytro Perekrestenko, and Michael Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [18] Sucheta Chauhan and Lovekesh Vig, "Anomaly detection in ecg time signals via deep long short-term memory networks," in *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2015, pp. 1–7.
- [19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [20] Andreas Maier, Christopher Syben, Tobias Lasser, and Christian Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 86–101, 2019.
- [21] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [22] Philip Warrick and Masun Nabhan Homsí, "Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [23] Shenda Hong, Meng Wu, Yuxi Zhou, Qingyun Wang, Junyuan Shang, Hongyan Li, and Junqing 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.
- [24] Zhaohan Xiong, Martin K Stiles, and Jichao Zhao, "Robust ecg signal classification for detection of atrial fibrillation using a novel neural network," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [25] Davari AmirAbbas, Hasan Can Ozkan, Andreas Maier, and Christian Riess, "Fast and efficient limited data hyperspectral remote sensing image classification via gmm-based synthetic samples," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2107–2120, July 2019.
- [26] Eric Jones, Travis Oliphant, Pearu Peterson, et al., "SciPy: Open source scientific tools for Python," 2001–, [Online; accessed 09.09.2019].
- [27] François Chollet et al., "Keras," <https://keras.io/>, 2015, Accessed: 2019-09-17.