

# Cardiac Arrhythmias Identification by Parallel CNNs and ECG Time-Frequency Representation

Jonathan R Torres<sup>1</sup>, K De Los Ríos<sup>2</sup>, Miguel A Padilla<sup>1</sup>

<sup>1</sup> Institute of Applied Sciences and Technology, Universidad Nacional Autónoma de México, Mexico

<sup>2</sup> Institute of Physics, Universidad Nacional Autónoma de México, Mexico

## Abstract

Heart abnormalities cause about 26 % of the deaths of illnesses in the world. Developing computational tools for ECG interpretation plays a critical role in the clinical diagnosis of Cardiac arrhythmias (CAs). **Aims:** This study aimed to develop an automated abnormal pattern recognition method for clinical decision support capable of detecting between 27 possible CAs. **Proposal:** An improved deep learning (DL) model was employed using raw-data and time-frequency representation (TFR) images. **Methods:** A vast set of ECG records were filtered and normalized. They were segmented and transformed into two sets of 2-D images. TFR images were obtained through Wavelet Synchrosqueezing (WS). The VGG-16 network was chosen, modifying the weights of the inner layers to adapt the model to the CAs detection task. A 10-fold cross-validation method was executed. Different training hyperparameters were tested to find the best model. **Results:** With the cross-validation on the training data, the model developed by our team UIDT-UNAM performed identifying CAs, with an overall unofficial S-score of 0.766. This model had a high performance in detecting healthy subjects with an F1 score of 0.83. We obtained these results using only the public training dataset. We plan to test these optimistic results with Physionet private dataset very soon.

## 1. Introduction

Heart abnormalities are the first cause of death for illnesses worldwide [1]. Cardiac arrhythmias (CAs) are the most frequent causes of them and contribute to approximately 15% to 20% of all deaths [2].

The standard 12-lead ECG is the most commonly used to diagnose cardiac abnormalities [3]. The early detection of CAs and its treatment for sudden cardiac deaths (SCDs) prevention, represents a significant opportunity to reduce mortality further [1]. However, ECG manual interpretation is slow, requires training personnel with a high degree of technical knowledge, and suffers subjectivity.

Physicians detect the action potentials of the signals and

analyses the absence of P wave, morphologies irregular of QRS complexes, and irregularity of the segments. Recent results have shown that physicians have an accuracy rate of 75% in the detection procedure of some CAs [4].

Computational tools that using automatic detection and classification of CAs can assist physicians in the ECG diagnosis. Recently, there have been increasing numbers of research focused on 12-lead ECG classification through machine learning (ML) and deep learning (DL) algorithms. Theoretically, many of these algorithms have been accurate in the identification of CAs. However, the successful result of those tests is a consequence of using small and homogeneous datasets. The Physionet/CinC 2020 challenge has provided a vast dataset for this purpose [5].

## 2. Material and methods

The CAs detection method proposed here consists of the following stages: ECG data pre-processing (noise removal and data segmentation) and CAs classification (Signals transformation and final classification). In the first stage, the wavelet transform (WT) method was applied to denoise the ECG signal. Then, the signals were segmented in equal periods of duration, taking into account the signal morphology. In the second stage, the Wavelet synchrosqueezing (WS) method was applied to obtain the time-frequency representation (TFR) images, and the images of the raw signals were obtained. Finally, the arrangement of parallel convolutional neural networks (CNNs) based on VGG-16 was trained from scratch to identify 27 types of CAs.

### 2.1. Data set

Six public datasets of 12-leads ECG records coming from four different sources were used [5]. The datasets have a total of 111 identified CAs, of which just 27 general classes were chosen to be assessed [5]. The dataset was divided randomly into three sets, training, validation, and local testing, with 70 %, 15 %, and 15 % of the data, respectively.

## 2.3. Signals denoising and segmentation

Signals were purged, and irrelevant information was discarded. A denoising process was performed to eliminate outside signals related to the sampling procedure. The Wavelet transform (WT) method was used for this purpose with the Daubechies4 (db4) function since it allows decomposing the input signal into low and high-frequency components [5].

Each file in each dataset contains the information corresponding to the 12-leads ECG. The first 6 s of each of the entire dataset signals were segmented in time intervals of 1.2 s. This interval size allows obtaining the relevant information around each peak, regardless of the type of CA that the patient presents, thus having each cycle of beats.

## 2.3. Signals and TFR images

Both sets of images were constructed by taking the segments of each signal. First, the segments were plotted as time series and saved as 64 x 64 greyscale images. The second set was built by transforming the signals using the WS method to obtain the TF features. With these features, the TFR images were obtained (see Fig. 2 b), d), and f)).

The WS method used for TFR is based on the continuous wavelet transform (CWT) [6]. In this transformation, concentrated high-resolution TF patterns are obtained, from which instantaneous frequency lines can be identified. The instantaneous frequency  $\omega(a, b)$  for any point  $(a, b)$  of the original signal with  $W_{a,b} \neq 0$  is given by:

$$\omega_{a,b} = -i(W_{a,b})^{-1} \frac{\partial}{\partial b} W_{a,b} \quad (1)$$

Where  $a, b$  and  $W_{a,b}$  are the scale factor, translational value, and WT, respectively. From this instantaneous frequency, the Synchrosqueezing discrete transform is determined at a local frequency point given by the transformation [6]:

$$T_{\omega_l, b} = (\Delta\omega_{a,b})^{-1} \sum_{a_k: |\omega_{a_k, b} - \omega_l| \leq \Delta\omega/2} W_{a_k, b} a_k^{-\frac{3}{2}} (\Delta a)_k \quad (2)$$

Finally, the representation of the signal in the TF space at high resolution is obtained. Fig. 1 shows both the raw signal and the TFR of three different types of CAs with the same window length of 1.2 s.

Different features can be observed for each segment, especially the amplitude in the TFR and the morphology in the raw signal images. The representation of the signal segments in these two types of images gives relevant information no longer observed when using only one of the

two representation types shown here. Then, it is possible to observe notable differences both in each class and to each kind of representation

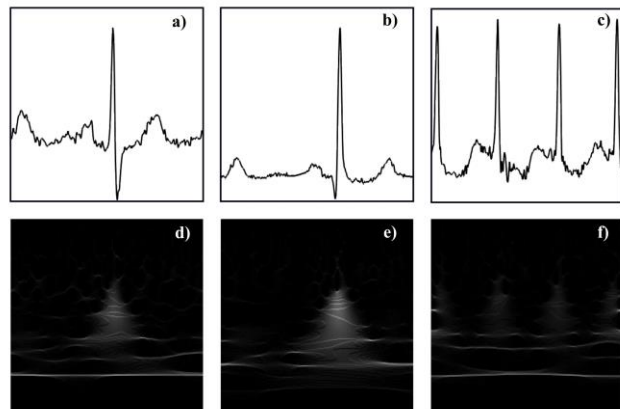


Figure 1. Transformation of the 1.2 s segments extracted from the signals to images: a, d) segment of normal sinus rhythm (NSR), b, d) segment ST depression (STD), and c, f) Premature atrial contraction (PAC).

As a result, we obtained 120 images for each patient from the arrhythmia datasets used here.

## 2.4. Deep learning model

CNNs have proven useful for automatic feature extraction in detecting abnormal patterns in clinical images without pre-processing algorithms or manual intervention [7].

A CNN is composed of an input and an output layer and many hidden layers composed of convolutional, pooling, and fully connected (FC) layers. The convolutional layers are locally connected to extract the features by applying a set of weights called kernels. The ReLU function for an input value  $x$  is generally used as activation functions and is defined as:

$$f(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases} \quad (3)$$

Relevant high-level features can be extracted with an increasing number of convolutional layers. The weights of the convolutional kernel parameters in each layer are trained with the backpropagation (BP) algorithm [8].

### 2.4.1. Model description

This deep network model provides the automatic classification of input segments through an end-to-end structure without the need for any hand-made feature

extraction or selection steps.

The deep network structure is composed of an arrangement based on the VGG-16 network [8], where the feature extraction stages (convolutions) are duplicated and arranged in parallel.

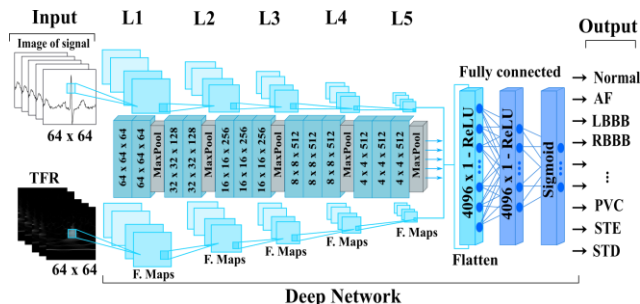


Figure 2. The architecture of the classification model developed.

This allows entry of the two independently constructed image sets. At the end of this arrangement, an FC network takes the output of convolution/pooling and predicts the best label to describe the image. The designed parallel CNNs model is shown in Fig. 2.

The network input size parameters were modified to support the two sets of 64 x 64 grayscale images used as input of the model modified. During the training, the 120 images constructed per patient of each of the classes pass through two 64 x 64 x 64 convolutional layers in the input layer and later through a max-pooling layer. The information is subsequently transferred to the following general layers. The second layer consists of two convolutional layers and one max-pooling layer. The next general layers are composed of three convolutional layers (see model characteristics in Fig. 2). At the end of the convolution layers of each available layer, there is a max-pooling layer. In convolutional layers are used filters with a kernel size of 3 x 3 and in the max-pooling layers a kernel size of 2 x 2.

After the five general layers of parallel CNNs, there is a FC of three layers with a different architecture. The first layer of FC receives the 4096 features obtained. The last layer, called a sigmoid layer, contains 27 channels, and it is in charge of classifying 27 labels (one for each class, see network configuration in Fig. 2). Finally, the network gives the signal probability to present each of the 27 CAs to identify a CA.

## 2.5. Model evaluation

The proposed method was applied to the set of datasets obtained. Some parameters were calculated from the classification results with the designed model: the F1 score

and average loss. The un-official S-score challenge was also computed [5] to assess the proposed parallel model effectiveness and reliability for the CAs identification. Although the dataset size is large, it is necessary to carry out the 10-folds cross-validation technique to stabilize the statistical model performance.

## 3. Results

The results of the implementation of the proposed model are shown. Also, a comparison of evaluation metrics is made with different training parameters tested to find the best model.

The cross-validation method of 10-folds was carried out to assess the model. In the local testing process, an F1 score of 82.5% was reached with a loss of 0.0617 and an unofficial S-score of 76.56%.

### 3.1. Model optimization

For the optimization of the CNN, model parameters such as learning rate and the batch size are taking into account. The step of model parameter optimization is indispensable to achieve the best classification performance.

Table 1. F1 score and average loss (A. Loss) when performing batch size variation.

Batch size	L. Rate	F1-Score	A. Loss
1024		0.804	0.0645
512		0.825	0.0617
256	0.001	0.818	0.0706
128		0.807	0.0641
64		0.812	0.0770

A set of variations in the batch size value was carried out to assess the proposed model's best performance. The batch size was modified five times, and the learning rate was automatically updated in the training process.

The number of epochs was set in 50. As a result, it was found that the best batch size value for the training process was 512. A lower loss and a higher overall F1 score were observed (see Tab. 1).

### 3.2. Comparison with other models

As an additional test, the classification of the CAs was performed by implementing two other widely-used CNNs to compare them with the proposed model.

Table 2. Comparison of the performances of three CAs classification algorithms.

Model	F-1 Score	A. Loss
VGG-19	0.813	0.075
ResNet-50	0.764	0.102
Proposed	0.825	0.062

From the results, it was observed that although the training time of the model proposed here is longer than that of Resnet-50 and that the VGG-19 network is more robust than the VGG-16, the evaluation of test signals gives a very accurate result with modified VGG-16 (see the comparison in Tab. 2).

#### 4. Discussion

The comparison of results was performed using only TFR images, only the raw data images, and both. Using only the TFR images, an F1 score of 76 +/- 2 was obtained and with the raw images of 66 +/- 3. For this reason, it was convenient to place them in parallel. Putting them in parallel allowed obtaining feature patterns in a time window as doctors do. At the same time, the features in the time-frequency window were extracted from their instantaneous energy.

The classification interferences found were mainly attributed to the morphology similarity between the different kinds of cardiac abnormalities present in the dataset, despite the class size difference. This characteristic in the classification makes this model suitable for unbalanced classes. However, it needs to be improved to differentiate between classes with similar characteristics properly. The realization of these adjustments in the classification methodology is proposed as future work presented here.

#### 5. Conclusion

In this paper, our team *UIDT-UNAM* proposed a useful CAs classification model using a parallel CNN with ECG images based on the VGG-16 network. As an input, two sets of 64 x 64 grayscale images were transformed from dataset ECG records. Over 37134 12-leads ECG records were processed, and near to 4456080 ECG beat images were obtained with 26 types of CAs and the normal rhythm. The optimized CNN model was designed with considering essential concepts such as 10-fold cross-validation. As a result, our proposed scheme achieved 91.57% SP, 82.50% F1-score, and 76.60% S-score. Our ECG classification result indicates that the identification of

arrhythmia with 2-D images and the VGG-16 model would be a practical approach to detecting CAs from 12-leads ECG signals. It is important to mention that this approach, with its encouraging results, will be verified very soon with the private dataset that is hosted on Physionet servers [5].

#### Acknowledgments

This project was developed with the support of DGAPA-PAPIIT TA100920. J.R. Torres and K. De Los Rios would like to thank CONACYT-Mexico and UNAM-PAEP for the support received for their Ph.D. studies. This research was partially developed in the Miztli-UNAM Supercomputer.

#### References

- [1] N. T. Srinivasan and R. J. Schilling, "Sudden cardiac death and arrhythmias," *Arrhythmia Electrophysiol. Rev.*, vol. 7, no. 2, pp. 111–117, 2018.
- [2] S. S. Virani *et al.*, "Heart disease and stroke statistics - 2020 update: A report from the American Heart Association," *Circulation*, pp. E139–E596, 2020.
- [3] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.*, vol. 127, pp. 144–164, 2016.
- [4] A. Y. Hannun *et al.*, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nat. Med.*, vol. 25, no. 1, pp. 65–69, 2019.
- [5] E. A. Perez Alday *et al.*, "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiol. Meas.*, no. (Under Review), p. 2020.08.11.20172601, Aug. 2020.
- [6] G. Thakur and H. T. Wu, "Synchrosqueezing-based recovery of instantaneous frequency from nonuniform samples," *SIAM J. Math. Anal.*, vol. 43, no. 5, pp. 2078–2095, 2011.
- [7] A. L. P. Ribeiro *et al.*, "Tele-electrocardiography and bigdata: The CODE (Clinical Outcomes in Digital Electrocardiography) study," *J. Electrocardiol.*, vol. 57, pp. S75–S78, 2019.
- [8] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-VGG16 CNN model for big data places image recognition," *2018 IEEE 8th Annu. Comput. Commun. Work. Conf. CCWC 2018*, vol. 2018-Janua, pp. 169–175, 2018.

Address for correspondence:

Jonathan R Torres-Castillo & Miguel A Padilla-Castañeda  
 Instituto de Ciencias Aplicadas y Tecnología (ICAT), UNAM,  
 Cto. Exterior, Cd. Universitaria, Mexico City, 04510, Mexico  
 jonathanrtc@comunidad.unam.mx  
 miguel.padilla@icat.unam.mx