Benchmark of deep learning algorithms for the automatic screening in electrocardiograms transmitted by implantable cardiac devices

Narimane Gassa, Benjamin Sacristan, Nejib Zemzemi, Maxime Laborde, Juan Garrido Oliver, Clara Matencio Perabla, Guillermo Jimenez-Perez, Oscar Camara, Sylvain Ploux, Marc Strik, et al.

Benchmark of deep learning algorithms for the automatic screening in electrocardiograms transmitted by implantable cardiac devices

Narimane Gassa[1,2,4], Benjamin Sacristan[1], Nejib Zemzemi [1,2,4], Maxime Laborde [1], Juan Garrido Oliver [3], Clara Matencio Perabla [3], Guillermo Jimenez-Perez [3], Oscar Camara [3], Sylvain Ploux [1], Marc Strik [1], Pierre Bordachar [1], Remi Dubois [1]

[1] Electrophysiology and Heart Modeling Institute (IHU-Lyric), Pessac, France
[2] Institute of Mathematics, University of Bordeaux, Talence, France
[3] Universitat Pompeu Fabra, Barcelona, Spain
[4] INRIA Bordeaux Sud-ouest, CARMEN Team, Talence, France

Abstract

The objective of this work was to benchmark different deep learning architectures for noise detection against cardiac arrhythmia episodes recorded by pacemakers and implantable cardioverter-defibrillators (PM/ICDs) and transmitted for remote monitoring. Up to now, most signal processing from ICD data has been based on classical hand-crafted algorithms, not AI or DL-based ones.

The database consist of PM/ICD data from 805 patients representing a total of 10471 recordings from three different channels: the right ventricular (RV), the right atria (RA), and the shock channel.

Four deep learning approaches were trained and optimized to classify PM/ICDs' records as actual ventricular signal vs noise episodes. We evaluated the performance of the different models using the $F2$ score.

Results show that the use of $2D$ representations of $1D$ signals led to better performances than the direct use of $1D$ signals, suggesting that the detection of noise takes advantage of a spectral decomposition of the signal, which remains to be confirmed in other contexts.

This study proposes deep learning approaches for the analysis of remote monitoring recordings from PM/ICDs. The detection of noise allows efficient management of this large daily flow of data.

1. Introduction

Remote monitoring of pacemakers and implantable cardioverter defibrillators (PM/ICDs) identifies early signs of lead failure, which lowers patient morbidity and mortality.

Remote monitoring generates a large amount of data with low yield. It is estimated that only 7 to 9% of the transmitted data requires a medical opinion, and less than 2% motivate early patient management [1]. In case of non-sustained ventricular arrhythmia, a major challenge is to identify lead noise (4% of the tracings). Noise episodes may eventually reveal electrode failure, the consequences of which can be disastrous (syncope, inappropriate shock, death)[2–4]. The analysis of these large datasets requires a lot of human resources, thus the deployment of remote monitoring surveillance in clinical practice remains difficult.

The ultimate goal of our project was to develop an assistance algorithm, based on a machine learning strategy, allowing an automated analysis of cardiac arrhythmia traces transmitted by PM/ICDs upstream of human evaluation. In this work, we compared four different deep learning approaches for noise detection against arrhythmia episodes in order to test three key aspects of the classifier: i) the input space for handling $1D$ signals; ii) the imbalanced data-set with missing signals; and iii) the deep network architecture.

2. Methods

2.1. Dataset

The database was composed of PM/ICD remote records collected at *Bordeaux University Hospital* and managed by *IHU Liryc*. A total of 805 patients has been followed-up, creating 10,471 records, each composed of right ventricular bipolar signals, and possibly atrial, left ventricular and/or far field EGM signals, depending on the type of device (single, dual, or triple-chamber pacemaker or implantable cardioverter defibrillator).

The database was randomly divided into 3 subsets:

- Training set: 417 patients, 4998 recordings
- Validation set: 142 patients, 2621 recordings
- Test set: 328 patients, 2843 recordings

The records labeled by the device as non sustained ventricular tachycardia (NSVT) may actually be relevant signals or noises. The database was then manually annotated, distinguishing two classes: relevant ventricular signal (CLASS **n**) and noise (CLASS **y**).

Table 1: Data profile.

	Training set	Validation set	Test set
Patients	417	142	246
Recordings	4998	2621	2843
NSVT (CLASS n)	4681	2515	2722
Noise (CLASS y)	317	106	121

2.2. Neural Network architectures

2.2.1. 1D-AgResNet

All Recordings were unevenly sampled at the maximum rate 200 Hz for efficient storage in the device. We applied data interpolation so that all time-steps were 5 ms. Considering that the recordings were not of the same length, window sliding was applied. It consisted of taking the first W points from the recording of L points, and then moving forward by S points and taking the following W points, at the end getting **(L-W)/S + 1** samples of each recording.

This way we achieved two things: (i) samples of equal length; and (ii) a larger sample size. Moreover, not all recordings have the same number of channels. As not all recordings had the same number of channels, we artificially added signal lines at zero to have 3 channels in all recordings. We finally obtained samples with the same number of channels, time-step, and number of points.

The architecture of our model was based on the Residual neural network with 3 channels in the input and 2000 points per sample.

2.2.2. 2D-DenseNet

To take advantage of Deep Convolutional Neural Networks, we converted the time series data into images.

In this part, we only considered the recordings of the right ventricular channel, so we obtained two-dimensional images of the temporal traces of the ventricular signal over $2000ms$. Then these images were fed to a Dense neural Network (DenseNet) for image classification. Input image dimension of the network was (248,372,1).

DenseNet is an improved version of convolutional neural networks that minimizes the vanishing gradient problem, and can be trained with few parameters which takes less time and is found to achieve better results. [5]

The figure depicts the architecture of a typical DenseNet model, we have dense blocks composed of interconnected layers : each layer is receiving information from all preceding layers and in between there are non-linear transformation (Batch Normalization, Rectified Linear Unit (ReLU), Convolutions). A transitional block is introduced between two dense blocks as a way of downsampling by applying a batch normalization, a convolution and a global average pooling.

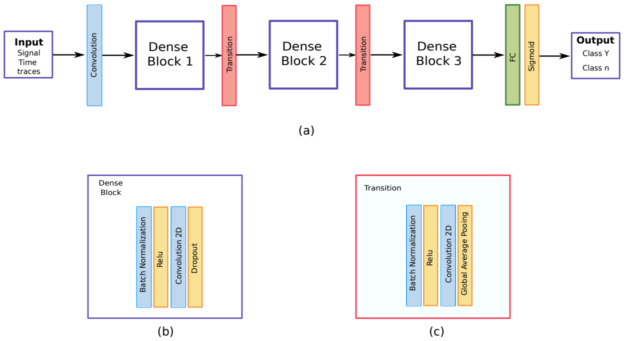

Figure 1: (a) Representation of the neural network architecture (2D-DenseNet), (b) Dense block, (c) transition layer.

2.2.3. 2DTF-CNN

All recordings were sampled at $200Hz$. In this section, we have only considered the first $15000ms$ for each sample and we carried out an interpolation in order to have a fixed time-step equal to $5ms$ and a number of recordings equal to 3000.

A spectrogram is the pointwise magnitude of the fourier transform of a segment of an audio signal. We compute, for each time signal, the corresponding spectrogram. We obtain two-dimensional graphs and a third dimension represented by colors, time is running along the $x-axis$. The vertical axis represents frequency. The amplitude of a particular frequency at a particular time is represented by the third dimension.

The $2D$ time/frequency maps of the ventricular bipolar signal are then fed to convolutional neural for classification purpose.

We propose here a classical CNN architecture constituted of a stacking of two convolutional layers alternating with two pooling layers and terminating with two fully connected (FC) layers and a dropout layer ($p = 0.5$) in-between. The output is a probability distribution obtained utilizing a sigmoid function on the output of the final FC layer. The network architecture is represented in Figure 2.

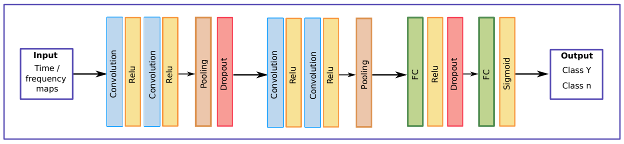

Figure 2: Representation of the neural network architecture (2DTF-CNN).

We used a stochastic gradient descent method (Adam) for training the proposed model with 32 samples for batch size. The hyper-parameters are set as follows: learning rate of 10^{-3}, and number of epochs are set to 100.

2.2.4. 2DTF-VGG

In this part, RV-time/frequency maps are also used as inputs of the network. Each RV recording is first divided into regular time segments each 10 seconds wide. We compute then for each part the spectrogram in the range [0-100Hz] with a frequency resolution of 2Hz, and a 50% overlapping sliding windows of 1s leading to one 50x20 image for each 10-second segment. These 2D representations feed a VGG-like neural network. Final architecture was selected using a 5 fold cross-validation process over the training dataset. In order to manage the unbalanced dataset, we train a cost-sensitive classifier and penalize missclassifications from the minority class y more than the majority class n by introducing weights to our loss function as follows:

$$Loss = \frac{1}{n}\sum_{i=1}^{n} W y_i log(\hat{y}_i) + (1-W)(1-y_i)log(1-\hat{y}_i)$$

Where y_i is the probability of input i to be be a noise event; and n the total number of example in the training database. W was set to 0.66 using the cross validation process.

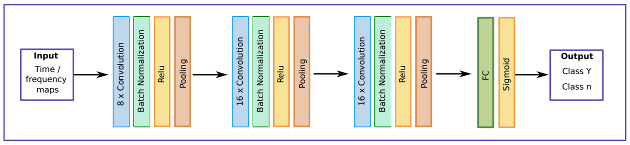

Figure 3: Representation of the neural network architecture (2DTF-VGG).

A record being composed of several parts, we classify the recording as noise if at least one of its parts is classified as noise.

2.3. Evaluation metric

To assess the accuracy of the classification results obtained by the different deep neural networks models, we compute the $F2$ score [6]:

$$F_\beta = (1+\beta^2)\frac{PPV \cdot se}{\beta^2 \cdot PPV + se}, \text{ with } \beta = 2$$

where *se* is the sensitivity to detect noise, and *PPV* the positive predictive value to detect noise.

$$se = \frac{TP}{TP+FN}, PPV = \frac{TP}{TP+FP}$$

For instance, noise (class **y**) is the positive class, and NSTV (class **y**) is the negative class.

3. Results

The CNN-based network (2DTF-CNN) that used 2D time/frequency maps of the ventricular bipolar signal as input gave the best results on the test set ($F2$ = 0.914), outperforming pre-trained VGG (2DTF-VGG) ($F2$ = 0.863). However, a CNN network based on a naive DenseNet architecture trained on 2D images of the ventricular signal time traces (2D-ResNet) also performed very well ($F2$ = 0.906). Both architectures surpassed networks with 1D signals as input (1D-AgResNet: $F2$ = 0.791).

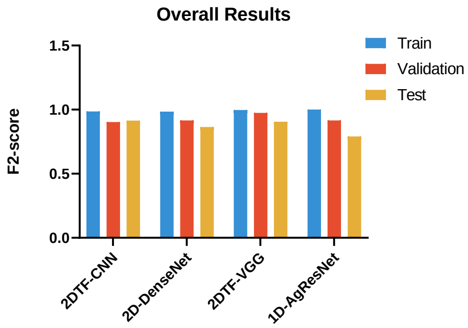

Figure 4: Overall Results.

In table 2 we show confusion matrices of the four methods. In particular, the false negative metric on the test set is coherent with the F2-scores (2DTF-CNN: FN = 9, 2D-DenseNet: FN = 13, 2DTF-VGG: FN = 17, 1D-AgResNet: FN=30). On the other hand, if we consider the number of the missclassified recordings (2DTF-CNN: FN+FP =26 , 2D-DenseNet: FN+FP = 17, 2DTF-VGG: FN+FP = 31, 1D-AgResNet: FN+FP=30) the 2D-DenseNet method outperforms the three others methods but the F2-score gives more weight to sensitivity(*se*) thus we penalize our models more for false negatives then false positives.

	Train	Valid	Test
TN	4677	2508	2705
FN	5	11	9
FP	4	7	17
TP	312	95	112

(a) 2DTF-CNN

	Train	Valid	Test
TN	4680	2513	2718
FN	1	3	13
FP	1	2	4
TP	316	103	108

(b) 2D-DenseNet

	Train	Valid	Test
TN	4679	2502	2708
FN	0	8	17
FP	2	13	14
TP	317	98	104

(c) 2DTF-VGG

	Train	Valid	Test
TN	4681	2511	2722
FN	0	10	30
FP	0	4	0
TP	317	96	91

(d) 1D-AgResNet

Table 2: Confusion matrices

4. Discussion and conclusion

In this work, we suggest a deep learning approach to automatically identify PM/ICD episodes with lead noise, results show that the use of $2D$ representation and a convolutional neural network (2DTF-CNN, 2D-DenseNet, 2DTF-VGG) outperforms the use of $1D$ representation of the signal (1D-AgResNet).

The 2DTF-CNN has the best performance in terms of $F2$-score reaching 91% on the test set. This may suggest that noise detection can also benefit from spectral decomposition of the signal, which remains to be confirmed in other contexts.

Regarding the imbalanced data, we used a weighted loss function which potentially improved the F2-score. It will also be interesting to work on resampling the training set as shown in [7] where the authors propose an under-sampling approach to make a balanced dataset out of the imbalanced one. Adding an arrhythmia class, [8], may also help the model distinguish noise episodes from arrhythmia episodes, thereby minimizing the false positive rate.

Acknowledgments

This Project has received funding from the European Unions Horizon research and innovation programme under the Marie Skodowska-Curie grant agreement No. 860974 and by the French National Research Agency, grant references ANR- 10-IAHU04- LIRYC and ANR-11-EQPX-0030.

References

[1] Vogtmann T, Stiller S, Marek A, Kespohl S, Gomer M, Kühlkamp V, Zach G, Löscher S, Baumann G. Workload and usefulness of daily, centralized home monitoring for patients treated with cieds: results of the monic (model project monitor centre) prospective multicentre study. Europace 2013; 15(2):219–226.

[2] Ploux S, Swerdlow CD, Strik M, Welte N, Klotz N, Ritter P, Haïssaguerre M, Bordachar P. Towards eradication of inappropriate therapies for icd lead failure by combining comprehensive remote monitoring and lead noise alerts. Journal of cardiovascular electrophysiology 2018;29(8):1125–1134.

[3] Nishii N, Miyoshi A, Kubo M, Miyamoto M, Morimoto Y, Kawada S, Nakagawa K, Watanabe A, Nakamura K, Morita H, et al. Analysis of arrhythmic events is useful to detect lead failure earlier in patients followed by remote monitoring. Journal of cardiovascular electrophysiology 2018; 29(3):463–470.

[4] Ploux S, Varma N, Strik M, Lazarus A, Bordachar P. Optimizing implantable cardioverter-defibrillator remote monitoring: a practical guide. JACC Clinical Electrophysiology 2017;3(4):315–328.

[5] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 4700–4708.

[6] Van Rijsbergen C. Information Retrieval. 2nd edition. Butterworth, 1979.

[7] Yoon D, Lim HS, Jung K, Kim TY, Lee S. Deep learning-based electrocardiogram signal noise detection and screening model. Healthcare informatics research 2019;25(3):201–211.

[8] Rajpurkar P, Hannun AY, Haghpanahi M, Bourn C, Ng AY. Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv preprint arXiv170701836 2017;.

Address for correspondence:

Narimane Gassa
INRIA Bordeaux sud-ouest
200 Avenue de la Vieille Tour, 33405 Talence
narimane.gassa@inria.fr