Identifying spatiotemporal dispersion in catheter ablation of persistent atrial fibrillation: a comparative study of machine learning techniques using both real and realistic synthetic multipolar electrograms

Sara Frusone [1], Rafael Costa de Almeida [1], Douglas Almonfrey [2], Fabien Squara [3], Vicente Zarzoso [1]

[1] Université Côte d'Azur, CNRS, I3S Laboratory, Sophia Antipolis, France
[2] Federal Institute of Espírito Santo, Vitoria, Brazil
[3] Université Côte d'Azur, Pasteur Hospital, Cardiology Department, Nice, France

Abstract

Atrial fibrillation (AF) is a common heart condition affecting the elderly population and is a significant risk factor for strokes, making it a growing public health concern. Catheter ablation (CA) is the most effective long-term treatment for persistent AF. Recently, a novel CA approach based on spatiotemporal dispersion (STD) has been proposed. This technique targets STD patterns associated with active zones responsible for sustaining the arrhythmia, aiming to improve the effectiveness of treating persistent AF. We present three datasets to be used to train and test different machine learning models in automatically identifying STD patterns from multipolar electrograms (EGM). Two different real dataset have been acquired from Nice Pasteur University Hospital and labelled by experts. To address the challenging scenario presented by the real data, a synthetic dataset has been created generating EGM records resembling real-world scenarios, using openCARP cardiac electrophysiology simulation software. We evaluate 13 machine learning techniques to demonstrate the challenging scenario of the real data, and we analyze their performance in the proposed datasets. This work indicates that the synthetic data is promising to be used as training set for classifiers to be evaluated in real data.

1. Introduction

Atrial fibrillation (AF) is a common irregular heart rhythm that primarily affects the elderly population and is a major cause of stroke. As the population continues to age, AF is becoming a significant public health concern [1]. To improve the understanding and management of this complex condition, physiological signal analysis and machine learning techniques are being used. Catheter ablation (CA) is currently the most effective long-term treatment for persistent AF, but its success rates may vary [2]. Although an increasingly larger number of patients are eligible for CA, the optimal ablation strategy for persistent AF remains elusive. Some past studies, based on the visual selection of target electrograms (EGMs), have suggested that applying lesions targeting complex fractionated atrial electrograms (CFAE) is beneficial to patients with persistent AF [3]. CA based on spatiotemporal dispersion (STD) has been recently proposed to treat persistent AF effectively based on the Pentaray multielectrode mapping catheter (Biosense Webster Inc., Irvine, CA, USA). STD patterns are thought to be associated with active zones sustaining arrhythmia, and are targets for successful ablation [4]. STD areas are defined as clusters of EGMs, either fractionated or non-fractionated, that display interelectrode time and space dispersion at a minimum of two adjacent bipoles such that activation spreads over $> 70\%$ of the AF cycle length. STD patterns are identified visually by the interventional cardiologist following those rules. The manual classification of the signals represents a substantial limitation to the standardization of EGM-based approaches with large operator differences in experience and learning curve profiles [3]. Also the communication between the software engineer recording the signal and the cardiologist doing the annotation in parallel is a challenging step, which unavoidably produces errors in the annotation process, and consequently in the data labels. Another important limitation concerns the strong inbalance in the available datasets between the two classes, STD patterns do not occur as often as non STD.

In the literature, a technical classification analysis of novel machine learning (ML) models trained to automatically classify STD patterns (such as VX1 software from Volta medical [3]) is missing. In this context we present an automatic classification analysis in two real datasets and a synthetic one. The first contains raw data from Nice Pasteur University Hospital (CHU). The second one is a curated version of the first. The third dataset is synthetic. The objective of using different datasets is to help in the as-

sesses of machine learning models, trained on multipolar EGMs for automatically locating STD patterns. In fact, the models are trained and validated in all the different configurations of the raw, curated and synthetic data. The final end would be providing to operators an automatic labelling tool that identifies the presence or absence of STDs, being a support for the human based procedure.
This study relays on multiparametric ML algorithms trained on annotated signals from intracardiac EGMs and tested on the mentioned different data configurations. In a further step, those models are expected to support interventional cardiologists, helping EGMs interpretation and guiding CA procedures based on STDs identification patterns.

2. Datasets

2.1. Real datasets

Two different real dataset acquired from the Cardiology Department of Nice Pasteur CHU are considered in this study. Each sample in the raw dataset consists of 10 time series with 2500 timesteps each, acquired from 53 persistent AF patients. The raw dataset presents sometimes wrong labels, due to communication delays in the annotation procedure between the software engineer and the cardiologist. The curated dataset comprises only a subset of samples revisited and relabeled after the intervention (offline) by the same specialist. This solved the delay in the annotation procedure during the ablation procedure using the Carto software.

2.2. Synthetic data generation

The use of synthetic data is the answer to many problems encountered by experts in our domain [5]. In particular, the STD classification, the interoperator subjectivity, the lack of a precise protocol for the labelling, the annotation procedure, i.e. the communication between the doctors and engineers during the intervention, are challenging problems. Other issues are related to privacy, anonimization and security, and not last the amount of curated data. It is difficult to find clinicians to label and check the annotation of real data. In order to generate realistic synthetic multipolar EGMs, tissue patches of the heart are generated by computer simulations. Ionic current and conductivity parameters are set to create fibrosis in the simulated heart tissue, leading to rotor-like propagation patterns linked to STD behavior.
The fibrotic, scar and normal tissue patterns are made by three different cell types based on the Courtemanche model [6] with variations in the ionic charges. The fibrotic patterns were created using clustering algorithms designed to generate block regions with specific conductivity and cell types. An example of the resulting activation map is shown in Figure 1. Each tissue patch measures 273 by 273 pixels (6.7 by 6.7 cm), matching the real scaling. Using the openCARP cardiac electrophysiology simulation software[1], realistic 10-leads EGM records are synthesized from the tissue patches, simulating the pentaray multipolar catheter.

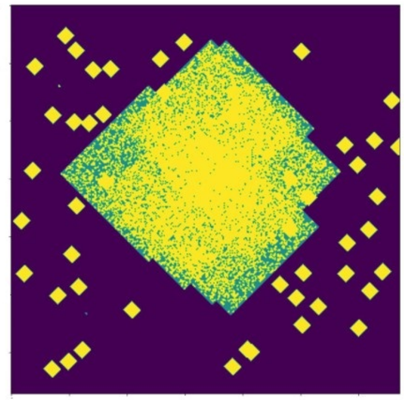

Figure 1. An example conductivity map is shown. The different colors represent different cell types and different conductivity values. Yellow regions represent Courtemanche ionic model with low conductivity, which mimics scar tissue. Green regions represent Courtemanche ionic model with medium conductivity, which mimics fibrotic tissue but differs from scars. Purple regions represent Courtemanche ionic model with high conductivity, which mimics normal tissue.

The advantages provided by the synthetic data generation process are many. It is possible to play with the parameters for the tissue patch generation, to change the degree of fibrosis and electrical conductivity. The catheter position can be moved around in the synthetic tissue, and the signal is then recorded.
We consider five different catheter position in a tissue patch, because in real life scenarios, during CA interventions, cardiologists do not consider very small rotations or movements of the sensor from one recording to another. It would be time consuming and the information recorded would not change much. From each catheter position, per patch we record the 10-lead EGMs. We set the length of the signal recordings to 2.5 seconds to be consistent with the real data annotation procedure performed at Nice Pasteur CHU's Cardiology Department using Carto software. An example of the simulated catheter on the tissue patch is presented in Figure 2, where the sensor poles have lifesize dimension (radius equal to 0.1 cm). At that stage, we were able to produce synthetic signal recordings. A synthetic dataset of multipolar EGM samples is obtained. It

[1] https://opencarp.org

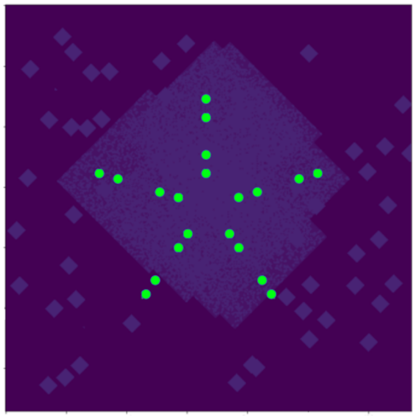

Figure 2. The green dots highlight each pole of the virtual pentaray multipolar catheter sensor on the simulated tissue. There are five branches and 20 poles, generating ten bipolar leads.

was built considering 37 different synthetic patients. The cardinality of the dataset opens new perspective concerning the kind of AI methods that can be applied.
Finally, Table 1 presents the different configurations, in terms of number of samples per class. By the term sample we mean a 2.5 s recording of ten EGM signals acquired from ten bipolar electrodes, one signal per lead.

Table 1. Datasets cardinality

Dataset	Patients	Samples	STD	non STD
raw	53	13888	1035	12853
curated	53	430	112	318
synthetic	37	64680	32340	32340

3. Classification approaches

In our approach to detect STD patterns, we employed a diverse set of features and classifiers. For classical machine learning classifiers, our feature set included voltage-based features. We computed the average of the mean, standard deviation, maximum, and minimum values for each lead in the raw signal. Furthermore, we derived the maximum value across all leads. For the convolutional neural networks (CNN), the 10 leads of the raw signal are treated as an image. Regarding the classification step, we explored 11 different classical techniques, and two neural networks. The classical ML techniques include Random Forest, Support Vector Machines, Logistic Regression, Decision Trees, K Neighbors, Gaussian Naive Bayes, Gradient Boosting, Extreme Gradient Boosting, Ridge, Multi-Layer Perceptron (MLP), and AdaBoost.
For the CNN, we normalized each lead to have unit L2 norm. Prior to the convolutional layer, we subjected the raw signal to average pooling with a 1x2 pool size. The core of the network comprised a single convolutional layer employing a 3x75 kernel to also capture temporal dependencies between leads. Subsequently, we applied average pooling with a 2x110 window. This was followed by one dense hidden layer with 4 units. The first model consists of 395 parameters, representing less than 1/10 of the number of training samples in the raw dataset. In the second network we added one more kernel in the convolutional layer and four more within the hidden layer. This second model includes 1109 parameters. It is expected to better accommodate a higher amount of data from the synthetic dataset.

4. Training and evaluation methodology

In our methodology, we followed a structured approach for training and evaluation. The experiments were performed in the following configurations, for both the classical ML methods, and the CNN approaches:

- train on the raw dataset, test on the curated dataset,
- train and test on the synthetic dataset,
- train on the synthetic dataset, test on the curated dataset.

In all cases, splits of training and test set are patient aware. For the classical ML approaches, we adopted a nested cross-validation strategy to determine the optimal algorithm among classical machine learning algorithms. An inner loop with two-fold cross-validation was combined with a two-fold outer loop. In the inner loop, we conducted a grid-search procedure, exploring classical hyperparameter values for each classifier. Feature scaling and data sampling were also considered. This approach enabled effective model parameter tuning, ensuring robustness in performance assessment through the outer loop. Then, for the CNN, we performed a two-fold cross-validation procedure with standard hyperparameters as a learning rate of 0.01 and the Adam optimizer. Given the model's relatively small size, explicit regularization was omitted. To address class imbalance, we employed oversampling of the minority class during training. Additionally, we incorporated EarlyStopping to mitigate overfitting and enhance training efficiency. Finally, our chosen key performance indicators (KPI) encompassed the F1 score (F1), positive predictive value (PPV), sensitivity (Sens), specificity (Spec), negative predictive value (NPV), and accuracy (Acc).

5. Results and discussion

Table 2 presents a summary of the test performances of the best models, in the different above mentioned configurations. For each datasets configuration, we present the best model in terms of F1 score among the eleven classical

Table 2. Final results of the experiments.

no.	Train	Test	Method	F1 score	Acc	Auc roc	PPV	Sens	NPV	Spec
1	raw	curated	LogisticRegression	0.419 ± 0.084	0.619	0.620	0.384	0.563	0.825	0.660
2	synthetic	synthetic	XGB	0.435 ± 0.075	0.837	0.919	0.296	0.832	0.984	0.838
3	synthetic	curated	GaussianNB	0.430 ± 0.028	0.365	0.604	0.281	0.919	0.858	0.170
4	raw	curated	CNN (#395)	0.430 ± 0.021	0.523	0.576	0.312	0.688	0.805	0.464
5	synthetic	synthetic	CNN (#395)	0.713 ± 0.197	0.946	0.928	0.649	0.825	0.986	0.955
6	synthetic	curated	CNN (#395)	0.270 ± 0.015	0.421	0.386	0.202	0.411	0.672	0.426
7	synthetic	curated	CNN (#1109)	0.415 ± 0.018	0.402	0.513	0.278	0.813	0.796	0.258

ML methods (lines 1 to 3) and two CNNs (lines 4 to 7) tested in this work.

From Table 2 we can note the challenging scenario presented by the curated dataset for all employed methodologies of classification. It is important to mention that when testing with the synthetic data, we resampled the test set to have the same class imbalance of the real dataset. Also, the synthetic /curated scenario demanded normalizing the lead signals to unit L2 norm, since the two dataset have different amplitude values.
When comparing lines 2 and 5, it is possible to note that the CNN take better advantage of the higher amount of data provided by the synthetic dataset. This is especially true when a bigger model is employed in the synthetic / curated scenario (lines 6 and 7). This opens the possibility of a better tuning of parameters when using convolutional and deep approaches with the proposed synthetic data. Finally, from lines 1, 3,4 and 7 we can notice that the synthetic data as a training dataset was able to provide results in pair with those obtained with the raw dataset. This indicates the synthetic data can reasonably represent real data, which supports its use with better-tuned parameters.

6. Conclusion

In this work, we presented different artificial intelligence based algorithms designed for identification of STD patterns from real and synthetic multipolar EGMs. This study brings us to a new understanding of the challenging scenario of the STD pattern classification. Besides, it indicates that the exploitation of synthetic data can be useful in the present classification problem. In future work, we will consider better tuned convolutional and deep models. Additionally, we will extend the synthetic data to a more realistic 3D model of the heart.

Acknowledgments

This work has been supported in part by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA 0002. V. Zarzoso holds the 3IA chair "IAblation".

References

[1] Morillo CA, Banerjee A, Perel P, Wood D, Jouven X. Atrial fibrillation: the current epidemic. Journal of geriatric cardiology: JGC. 2017;14(3):195.

[2] January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. Journal of the American College of Cardiology. 2014;64(21):e1-e76.

[3] Seitz J, Durdez TM, Albenque JP, Pisapia A, Gitenay E, Durand C, et al. Artificial intelligence software standardizes electrogram-based ablation outcome for persistent atrial fibrillation. Journal of Cardiovascular Electrophysiology. 2022;33(11):2250-60.

[4] Seitz J, Bars C, Théodore G, Beurtheret S, Lellouche N, Bremondy M, et al. AF ablation guided by spatiotemporal electrogram dispersion without pulmonary vein isolation: a wholly patient-tailored approach. Journal of the American College of Cardiology. 2017;69(3):303-21.

[5] Chen RJ, Lu MY, Chen TY, Williamson DF, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nature Biomedical Engineering. 2021;5(6):493-7.

[6] Courtemanche M, Ramirez RJ, Nattel S. Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. American Journal of Physiology-Heart and Circulatory Physiology. 1998;275(1):H301-21.

Address for correspondence:

Sara Frusone
I3S Laboratory, Euclide B, 2000 Route des Lucioles, 06900 Sophia Antipolis, France
sara.frusone@etu.univ-cotedazur.fr