

In-Silico Data Based Machine Learning Technique Predicts Premature Ventricular Contraction Origin Coordinates

Andony Arrieula^{1,2,3}, Hubert Cochet^{2,4,5,6}, Pierre Jaïs^{2,4,5,6},
Michel Haïssaguerre^{2,4,5,6}, Nejib Zemzemi^{1,2,3}, Mark Potse^{3,2,1}

¹CARMEN Research Team, Inria Bordeaux – Sud-Ouest, Talence, France

²IHU Liryc, fondation Bordeaux Université, Pessac, France

³Univ Bordeaux, IMB, UMR 5251, Talence, France

⁴Univ Bordeaux, CRCTB, U1045, Bordeaux, France

⁵INSERM, CRCTB, U1045, Bordeaux, France

⁶Bordeaux University Hospital (CHU), Electrophysiology and Ablation Unit, Pessac, France

Abstract

Premature ventricular contraction (PVC) can induce ventricular tachycardia or ventricular fibrillation. Drug-resistant PVCs can be cured by catheter ablation, but the accurate localization that this requires can be difficult and time-consuming. An accurate pre-procedural estimate of the origin could make the procedure more efficient.

We propose a machine-learning method for accurate pre-procedural origin estimation. It uses a database of paced 12-lead ECGs with known pacing locations and presents its results on an imaging-based model of the patient. The method was tested using 7 realistic heart-torso models with hundreds of PVCs everywhere in the ventricles.

We found that increasing the number of patients in the training database increased the accuracy of the predictions. The optimal number of pacing sites per patient in the training dataset was about 25, resulting in a prediction error around 15 mm.

We conclude that our method gives a good indication to clinicians to efficiently start a pace-mapping during a catheter ablation procedure. It can be complemented with an intra-procedural method that uses the patient's own paced beats to refine the prediction.

1. Introduction

Premature Ventricular Contraction (PVC) is a relatively common event where the heartbeat is initiated in the ventricles rather than by the sinoatrial node. Incidental PVCs are usually harmless, but multiple PVCs can induce a ventricular tachycardia or fibrillation.

To cure very frequent and drug-resistant PVCs, the site

of origin can be ablated with an endocardial catheter. We previously proposed a patient-specific method to localize PVC origin based on iterative pace mapping, using only a 12-lead ECG and no cardiac imaging [1]. However, this method can only work during the intervention, because it needs to be trained with stimulated beats. Cardiologists are often interested in a pre-interventional estimate, and often dispose of cardiac imaging performed before the procedure. To provide such an estimate we propose a machine learning method, based on cardiac imaging and a database of multiple patients.

Recent studies have used machine learning to classify PVC origin by regions, but some work with very large regions [2], are only applicable for endocardial origins [3], or only for left ventricular origins [4]. Our proposed method, in contrast, can predict origins in both ventricles and at any depth, and can also predict 3D coordinates of origin, using only the QRS complex of the 12-lead electrocardiogram (ECG).

The method works as follows: cardiac imaging data are used to create a 3D mesh of the ventricles of the patient. We register this mesh to a reference bi-ventricular mesh. In addition we use a database of 12-lead ECGs previously recorded after pacing at known locations in multiple pa-

Patient	LV endo	LV epi	RV endo	RV epi	All sites
#1 (M,44)	99	94	99	95	387
#2 (F,31)	99	88	99	97	383
#3 (F,20)	99	95	99	99	392
#4 (M,59)	99	95	99	98	391
#5 (F,15)	99	95	99	97	390
#6 (M,35)	99	92	99	96	386
#7 (M,44)	99	95	99	99	392
All patients	693	654	693	681	2721

Table 1. Number of paced beats by patient and tissue.

tients, also registered to the reference mesh. Thus, using features of the 12-lead ECGs and the corresponding origin coordinates, we can train a model to predict the 3D coordinates of a PVC origin on the reference mesh. The last step is to translate the prediction to the patient’s own mesh.

In this study, we are interested in determining the impact of the number of patients and pacing sites in the database on the results.

2. Methods

2.1. Database

To construct the database, we used detailed anatomical models of the heart and torso of 7 cardiac patients. These models were previously created from computed tomography data. Simulations were performed on 3-dimensional computational meshes of normal ventricles with a uniform edge length of 200 μm . The activation of the ventricles was simulated with a monodomain reaction-diffusion equation and the Ten Tusscher – Noble – Noble – Panfilov ionic model [5]. The 12-lead ECG was computed using the lead-fields method [6]. The simulations were performed with a recent version of the propag-5 software [6, 7].

For each model, a few hundreds of paced beats were simulated. Their origins were chosen randomly but well distributed to cover all ventricular regions. The details of the database are shown in Table 1.

2.2. Reference Mesh

For each patient, a 3D mesh of the ventricles was created. The mesh of patient 1 was chosen as the reference mesh. The other meshes and the pacing sites defined on them were scaled to the reference mesh using the iterative closest point method.

2.3. Machine Learning Algorithm

As an input of our machine learning algorithm, we used the features proposed by Giffard-Roisin et al. [8]. For each QRS complex, the following features were extracted: 1) QRS integral, 2) absolute potential of the global extremum, 3) timing of the global extremum, 4) sign of the potential of the global extremum, 5) number of zero crossing, 6) number of local extrema, 7) sign of the potential of the first extremum. Thus, for each paced beat we had $12 \times 7 = 84$ features.

We chose the *Gradient Boosting* method, using the implementation in the *scikit-learn* [9] package. We performed 500 boosting stages, with a maximum depth of 5 for the trees. At each step, we considered only the square root of the number of features when looking for the best

split, which was performed only if at least 1 % of the samples would be in the new nodes. The learning rate to shrink the contribution of each tree was set to 0.05. A random subsample of 80 % of the training dataset was selected at each tree, so the method becomes *Stochastic Gradient Boosting*.

To assess the effect of the mesh variability on the database, we gradually increased the database by including a new patient geometry at each step. We then had 6 cases: In the first case we learned on simulations performed on a single geometry, in the second case we used two geometries, and so on. For each case, we evaluated prediction errors. For the i -th case, we had C_6^i possible combinations to construct the learning database, built using i geometries. For a given combination, we randomly chose 50 pacing sites per patient, so the size of the database is $i \times 50$ simulations. We then predicted the stimulation sites for the remaining geometries. In order to avoid bias in the choice of the stimulation sites, this step was repeated 100 times and the error between the predicted sites and the real location of the stimulation sites was computed.

In order to reduce the number of stimulation sites used in the training phase, in the 6th case, we gradually decreased the number of selected pacing sites per patient from 50 to 4, and we computed the corresponding prediction error.

3. Results

Fig. 1 shows the effect of geometric variability in the database. The i -th bar group corresponds to the prediction error in the i -th case described above. For instance, the first bar group provides the mean and the standard deviation (20.7 ± 9.5) of the prediction error when learning on 50 simulations performed on a single geometry, while the last group provides the mean and the standard deviation (14.8 ± 7.1) of the prediction error when learning on 6 different geometries. The mean error of the prediction decreases as the number of patients in the database increases.

Fig. 2 shows the sensitivity of the method to different numbers of pacing sites per patient in the training database. The figure shows that the prediction decreases with increasing number of pacing sites per patient in the training database. However, the improvement in the prediction accuracy was negligible when more than 30 pacing sites were used.

4. Discussion

In this study, we proposed a machine learning method to localize PVC origins, using the 12-lead ECG and cardiac imaging data, before the clinical intervention.

For the preparation of an ablation procedure it is important to know in advance in which ventricle the origin lies. With 15 mm accuracy, our method will in most cases be

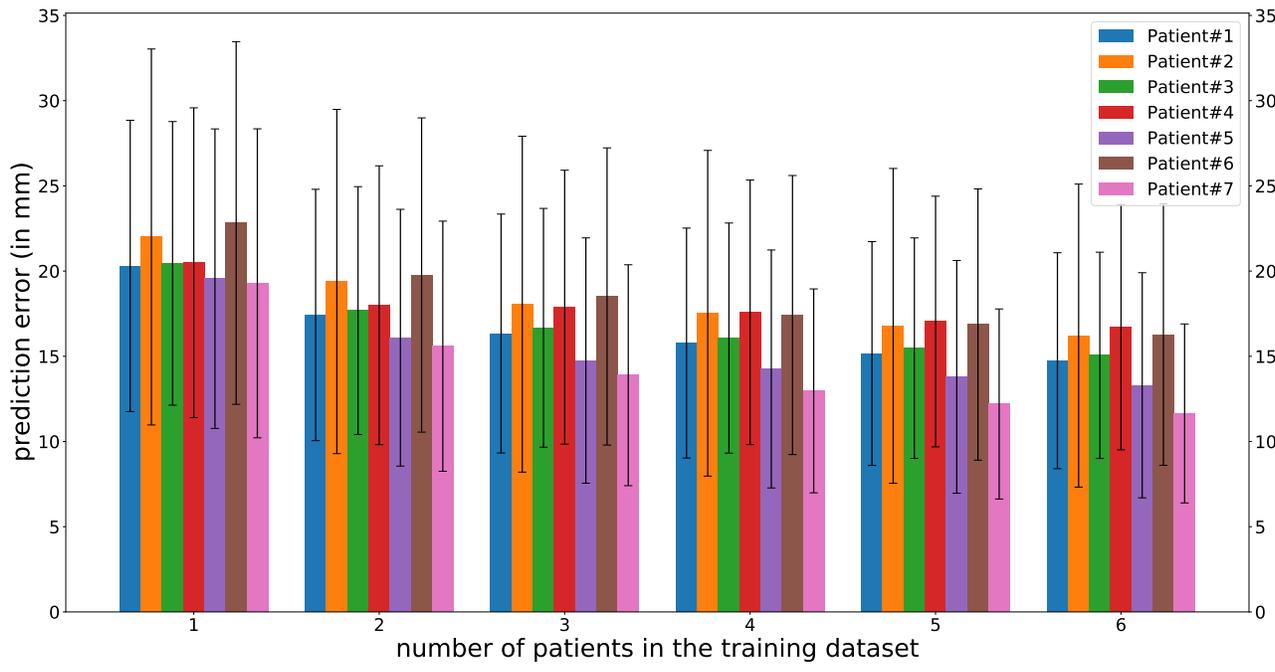


Figure 1. Mean prediction error over the pacing sites of the considered patient as a function of the number of patients in the training dataset.

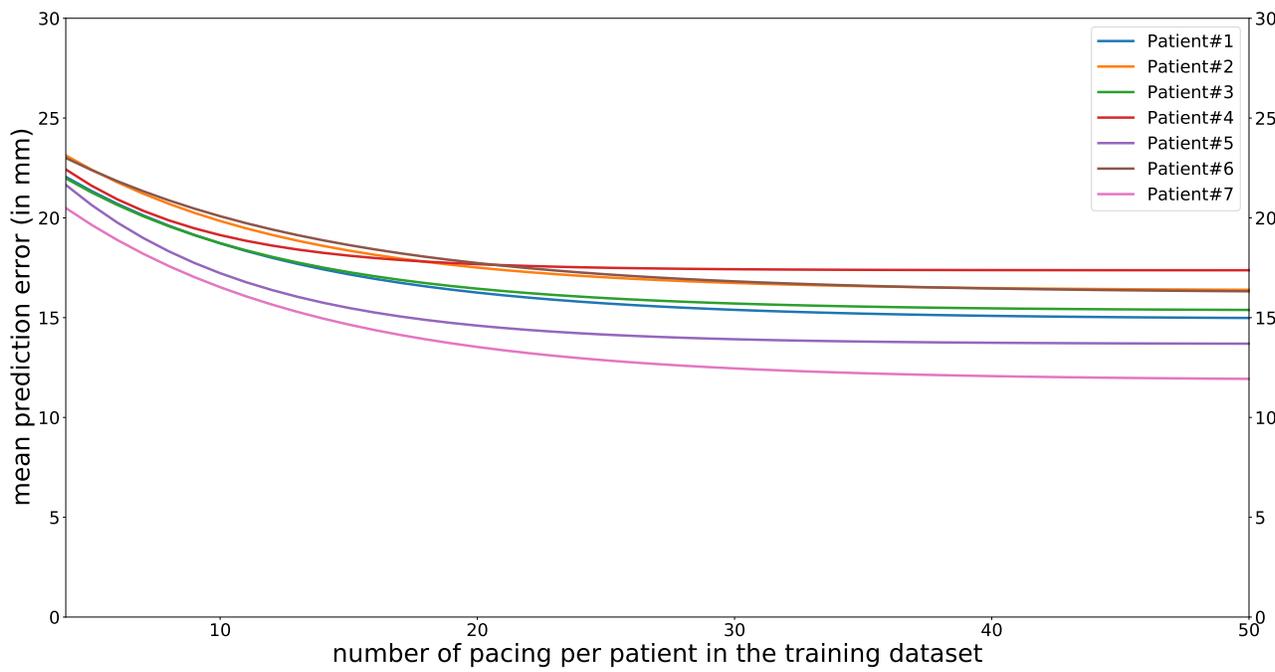


Figure 2. Fitted curves showing the mean prediction error over the pacing sites of the considered patient as a function of the number of pacing per patient in the training dataset.

able to indicate the correct ventricle, and if the prediction is close to the interventricular septum the cardiologist will be warned that access to both ventricles may be needed during the procedure. Our results show that the accuracy of the model can be improved by adding more patients to the training database. It also shows that we don't need more than 30 pacing sites per patient in order to be efficient, which is clinically feasible.

However, this method requires cardiac imaging, which is not always available for catheter ablation of PVCs. Since they are generated randomly, some pacing sites in the current database cannot be reached by a catheter. This suggests to construct a model that learns only on accessible pacing sites, which has not been done in the current work. The current accuracy of the model, around 15 mm of error, is not enough for direct ablation, but useful for pre-interventional planning. Moreover, the method can be complemented with an intra-procedural method that uses the patient's own paced beats to refine the prediction, as we proposed in our previous work [1]. Also, this test was performed with only 6 patients in the training database, and our data suggest that additional patients can increase the accuracy of the predictions. This method can already be validated with cathlab data after procedure, by comparing our predictions and the ablation sites.

Acknowledgements

This work was supported by the Région Nouvelle-Aquitaine, grant nr. 2017 – 1R50109 – 00013434; the European Research Council (H2020 grant agreement number 715093, ECSTATIC); and the French National Research Agency, grant references ANR-10-IAHU04-LIRYC and ANR-11-EQPX-0030. This work was granted access to the HPC resources of CINES under the allocation 2021-A0090307379 made by GENCI.

References

- [1] Arrieula A, Cochet H, Jaïs P, Haïssaguerre M, Potse M. An improved iterative pace-mapping algorithm to detect the origin of premature ventricular contractions. In *Computing in Cardiology*, volume 47. 2020; 62.
- [2] Nakamura T, Nagata Y, Nitta G, Okata S, Nagase M, Mitsui

- K, Watanabe K, Miyazaki R, Kaneko M, Nagamine S, Hara N, Lee T, Nozato T, Ashikaga T, Goya M, Sasano T. Prediction of premature ventricular complex origins using artificial intelligence-enabled algorithms. *Digital Health Journal* 2021;2:76–83.
- [3] He K, Nie Z, Zhong G, Yang C, Sun J. Localization of origins of premature ventricular contraction in the whole ventricle based on machine learning and automatic beat recognition from 12-lead ECG. *Physiol Meas* 2020;41:055007.
- [4] Sapp JL, Bar-Tal M, Howes AJ, Toma JE, El-Damaty A, Warren JW, McInnis PJ, Zhou S, Horáček BM. Real-time localization of ventricular tachycardia origin from the 12-Lead electrocardiogram. *JACC Clin Electrophysiol* 2017;3:687–699.
- [5] ten Tusscher KHWJ, Noble D, Noble PJ, Panfilov AV. A model for human ventricular tissue. *Am J Physiol Heart Circ Physiol* 2004;286:H1573–H1589.
- [6] Potse M. Scalable and accurate ECG simulation for reaction-diffusion models of the human heart. *Front Physiol* 2018; 9:370.
- [7] Krause D, Potse M, Dickopf T, Krause R, Auricchio A, Prinzen FW. Hybrid parallelization of a large-scale heart model. In Keller R, Kramer D, Weiss JP (eds.), *Facing the Multicore-Challenge II*, volume 7174 of *Lecture Notes in Computer Science*. Berlin: Springer, 2012; 120–132.
- [8] Giffard-Roisin S, Jackson T, Fovargue L, Lee J, Delingette H, Razavi R, Ayache N, Sermesant M. Noninvasive personalization of a cardiac electrophysiology model from body surface potential mapping. *IEEE Trans Biomed Eng* 2017; 64(9):2206–2218.
- [9] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 2011;12:2825–2830.

Address for correspondence:

Andony Arrieula
CARMEN research team
Inria Bordeaux – Sud-Ouest
200 Avenue de la vieille tour
33450 Talence, France
andony.arrieula@inria.fr