Introducing the ARGO Dataset of Post-Ischemic Ventricular Tachycardia Bipolar Electrograms

Marco Orrù[1,2], Giulia Baldazzi[2], Stefano Bandino[3], Davide Zirolia[3], Livio Bertagnolli[4], Graziana Viola[3], Danilo Pani[2]

[1]Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genova, Genova, Italy
[2]Department of Electrical and Electronic Engineering (DIEE), University of Cagliari, Cagliari, Italy
[3]Clinical and Interventional Cardiology Unit, Santissima Annunziata Hospital, Sassari, Italy
[4]Electrophysiology and Cardiac Pacing Unit, San Maurizio Regional Hospital, Bolzano, Italy

Abstract

The development of automatic tools to identify, characterize, and delineate abnormal ventricular potentials (AVPs) in post-ischemic ventricular tachycardia (VT) electrograms (EMGs) is still an open research issue. As a significant amount of annotated data is required for this purpose, the lack of open and annotated datasets hampers the research progress in this field. This work introduces a new open dataset that will be made available on PhysioNet by the end of 2023, which was recorded, annotated, and validated in the framework of the "Ablation Reinforcement by computer-aided Guidance and Optimization: ARGO study" project. The proposed dataset comprises 2034 anonymized recordings from nine post-ischemic VT patients, including 2.5 s bipolar EGMs, unipolar EGMs, the 12-leads surface ECGs, and the electroanatomic maps (voltage and local activation time maps). The dataset annotation was carried out by three expert electrophysiologists, who categorized each bipolar EGM into the classes of Physiological, AVP, or Unknown/Undetermined, and delineated the onset and end of the AVP. The ARGO dataset represents a valuable tool to enable the effective development and benchmarking of AVP detection and delineation algorithms, their characterization, but also to encourage collaboration across the research community.

1. Introduction

The management of post-ischemic ventricular tachycardia (VT) stands as a critical challenge in contemporary interventional cardiology. As a leading cause of sudden death mortality among ventricular arrhythmia patients [1], VT necessitates innovative approaches for permanent treatment and effective intervention. Notably, the recurrence rate of VT after radiofrequency ablation procedures is high, approximately 35% [2]. Indeed, the outcome of radiofrequency ablation guided by electroanatomic (EA) mapping strongly depends on accurately finding and treating the cardiac areas sustaining the arrhythmia. The procedure entails the localization of regions associated with the presence of anomalous electrical conduction patterns, resulting in abnormal ventricular potentials (AVPs) [3], [4] identifiable in the electrograms (EGMs). Currently, the identification of AVPs is a time-consuming practice, requiring visual inspection during the procedure and subsequent manual tagging.

With the final aim of supporting clinicians in identifying AVPs and expediting VT ablation procedures, there is an increasing interest [5], [6] in developing computational approaches leveraging artificial intelligence and signal processing techniques. However, the development and validation of these algorithms critically rely on the availability of comprehensive and annotated datasets. Remarkably, there is a noticeable lack of publicly available datasets dedicated to intracardiac signals, especially to post-ischemic VT. Indeed, even though there is a high number of available labelled datasets for 12-lead ECG, the same is not true for intracardiac signals. Existing works on intracardiac EGMs datasets are rare and mainly focus on annotated QRS complexes for atrial fibrillation and flutter [7], or do not provide free and public access to the signals [8]. This fact represents a significant obstacle to the research in this field.

In this work, we aim at proposing a new open dataset, namely the ARGO dataset, recorded and annotated in the framework of the "Ablation Reinforcement by computer-aided Guidance and Optimization: ARGO study" project. This dataset contributes to solving the lack of open datasets on post-ischemic VT by proposing the first open-access, multi-annotated dataset as a valuable tool to

enable the effective development and benchmarking of AVP detection and delineation algorithms.

2. Materials and Methods

2.1. Population

The recorded data come from nine post-ischemic VT patients who underwent EA mapping and radiofrequency ablation procedure between 2017 and 2018 at the San Francesco Hospital (Nuoro, Italy). All participants who contributed their signals to the study provided signed informed consent. A basic description of the dataset population is summarized in Table 1. Anonymization techniques were employed to remove any identifying information that could link data to individual participants.

The study received ethical approval by the Independent Ethical Committee of the Azienda Tutela Salute, Sardegna (Prot. n. 351/2021/CE, date of approval: 13/07/2021) and has been performed following the principles outlined in the 1975 Helsinki Declaration, as revised in 2000.

Table 1. Basic description of the dataset population. In the last two rows, the number of EGMs per patient (p) is reported.

Attribute					Value			
Number of patients					9			
Age [y.o.]					66 ± 10			
Sex [%]					M: 78, F: 22			
Ejection Fraction [%]					29 ± 6			
Length of records [s]					2.5			
Total number of records					2034			
Number of records per patient								
p1	p2	p3	p4	p5	p6	p7	p8	p9
96	54	112	48	489	129	161	867	78

2.2. Signal characteristics

All the recordings and left ventricle EA mapping were performed by an expert electrophysiologist during routine electrophysiological studies and ablation procedures, by using the CARTO®3 system (Biosense Webster, Inc., Diamond Bar, California). Specifically, EGMs and EA data were recorded in sinus rhythm by employing PentaRay, ThermoCool SmartTouch, and ThermoCool SmartTouch SF catheters by Biosense Webster, Inc. The acquired electrophysiological data include unipolar and bipolar EGMs, 12-lead surface ECGs, and EA maps, i.e., voltage maps and local activation time (LAT) maps.

Each recording featuring 12-lead ECGs and EGMs was sampled at a rate of 1 kHz and lasted 2.5 seconds, with a resolution of 0.003 mV. All recordings analyzed in the study underwent band-pass filtering by the CARTO®3 system, designed differently according to the kind of signal. Specifically, bipolar EGMs were filtered between 16 and 500 Hz, unipolar EGMs between 2 and 240 Hz, and 12-lead ECG traces between 0.5 and 120 Hz.

During the annotation procedure, only the segment of the EGM around the last cardiac cycle, as pinpointed by a reference fiducial point, was considered for each recording. This choice ensured the inclusion in the annotated dataset of beats acquired exclusively while the catheter was in perfect contact with the endocardium.

2.3. Annotation procedure

All the recorded bipolar EGM segments were manually annotated by three experienced electrophysiologists (namely, Annotator1, Annotator2 and Annotator3) blinded to the case using an ad-hoc graphical user interface (GUI) entirely developed in MATLAB [9]. The annotations were divided into three classes based on the physiological or pathological nature of the signal. The first class is called *"Physiological"* and is related to EGMs with physiological patterns, even if acquired from damaged myocardium. The second main class is *"AVP"* and pertains to EGMs associated with abnormal ventricular activations, thus in principle related to myocardial areas triggering or sustaining the VT. The last class, namely *"Unknown"*, was conceived as a class including all EGMs with doubtful or noisy morphologies or recorded in deep scar substrates, without any recognizable pathological behavior. Remarkably, if a signal was annotated as AVP, the annotators were additionally asked to delineate the onset and the end of the abnormal potential.

The set of data to be annotated and included in the dataset consists of 2034 EGMs with an unequal distribution among the nine patients. Specifically, among the data extracted from the nine EA mapping procedures, only recordings with reliable representation onto the EA maps, i.e., only those corresponding to EA points spatially projected in the reconstruction of the LV ventricle with a projection distance less than 8 mm, were considered. Moreover, the EGMs projected onto non-myocardial tissue areas, such as the aortic and mitral valve regions, were excluded.

In order to ensure comprehensiveness and robustness for the whole data labelling process, a final majority voting of the annotations provided by the three annotators was performed. This approach enhances the reliability and accuracy of the dataset annotations, mitigating potential biases. As a result, for each EGM, a final annotation derived from the voting process is made available, along with the three individual annotations of the initial annotation phase. As regards the delineation data, since it constitutes a quantitative measure, voting was not applicable. Therefore, only the onset and end delineations performed by individual annotators are provided.

2.4. Dataset description and organization

The dataset is organized in a collection of patient-specific folders. A comprehensive set of electrophysiological traces is stored in each folder, including a bipolar EGM, the two corresponding unipolar EGMs, the 12-lead ECGs, and EA map data. The electrophysiological traces and the related annotation data were formatted using the standard WFDB software package format, which can be imported and analyzed using functions such as `rdsamp` and `rdann` from the WFDB Software Package. The WFDB-compatible signals are reported as `Pn.dat` files associated with a corresponding `Pn.hea` header file, where `P` stands for "point", and `n` is an index assigned by the CARTO system during the export of the procedures. The `Pn.dat` files are structured as matrices with dimensions 2500x15, wherein rows denote signal samples, and columns correspond to distinct leads. The arrangement of the 15 columns adheres to a specific sequence, starting with the bipolar channel. Subsequently, the two corresponding unipolar EGMs are reported, followed by the 12-lead ECG. For clarity and precision, all recordings are stored as dimensionless integer values, with a conversion factor of 0.003 to be employed to establish the corresponding millivolt (mV) measurements. An example of the electrophysiological recordings contained in the proposed dataset is shown in Fig.1.

The header files contain annotation data generated through the majority voting process and individually defined by each annotator. Additionally, if the `n`-th signal has been labelled as AVP, delineation files compatible with WFDB are supplied independently for each annotator. As a result, if some annotators have not labeled the `n`-th signal as AVP, the corresponding delineation file for that annotator is absent. Finally, the EA map reconstruction data are reported in text (.txt) format files, which supply the necessary data for mesh reconstruction. These details are presented in a text-based format, ensuring accessibility and ease of reference for subsequent analyses. An example of the EA maps reconstruction contained in the proposed dataset is shown in Fig.2.

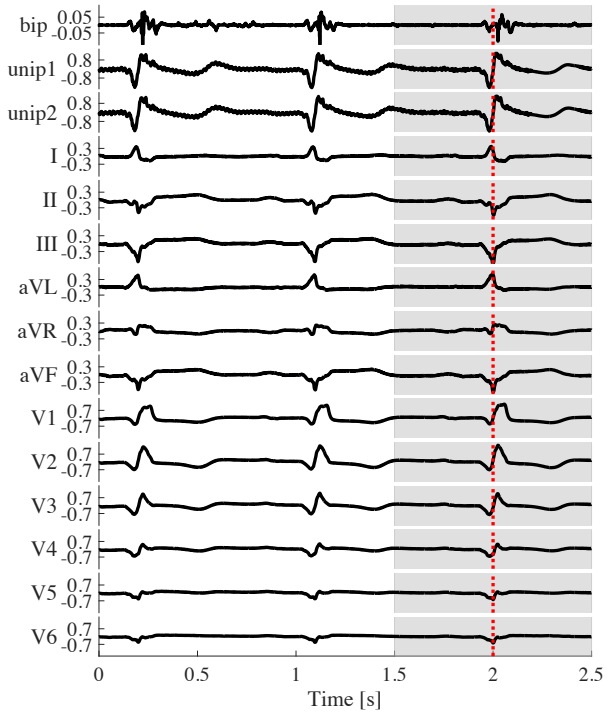

Figure 1. An example of raw bipolar EGM, unipolar EGMs and 12-lead ECGs (in mV) from the proposed dataset, with the indication of the fiducial point (red dotted line) pinpointing the last cardiac cycle (grey area).

3. Results and Discussion

After the independent annotation process involving the three annotators, followed by a majority voting procedure, the final class annotation of the dataset resulted in 1022 EGMs (50.25%) categorized as AVP, 708 EGMs (34.8%) as Physiological, 247 EGMs (12.1%) as Unknown, and 57 EGMs (2.8%) as Undetermined. In this phase, the *"Undetermined"* category was introduced to collect signals for which the three annotators provided different annotations, making it impossible to assign a

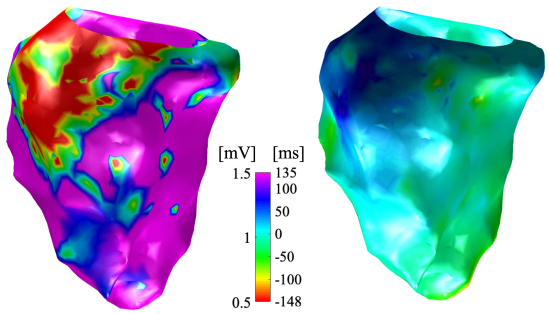

Figure 2. An example of EA maps which can be obtained from the mesh reconstruction data. On the left (a), the Voltage map, and on the right (b) the LAT map of the same ventricle.

final class through majority voting. By assessing the agreement between the annotators, the annotators reached full agreement in their classification for 1338 (65. 8%) out of 2034 analyzed EGMs, while for 639 EGMs (31.4%), only two out of three annotators agreed in giving the same annotation. Furthermore, to give an overview of each annotator performance, the accuracy of each annotator in correctly determining the EGM class has been estimated by considering the final voted labels as the gold standard. Specifically, the accuracy of each annotator was assessed by calculating the percentage of instances where the annotator's labelling was aligned with the labels derived from the majority voting (Table 2).

Remarkably, the results reveal a mean annotation accuracy of 87.7%, as well as a consistent association between annotators' accuracy and their respective levels of expertise. This coherence underscores, even more, the strong influence of annotator experience on the quality and reliability of annotations in this kind of exam and possibly on the EA procedures outcome.

Table 2. Accuracy with respect to the final voting and years of experience of each annotator.

Annotator	Accuracy (%)	Experience (years)
Annotator1	80.5	2
Annotator2	90.9	12
Annotator3	91.7	12

4. Conclusion

This work presented the development of the first open-access dataset of annotated post-ischemic VT intracardiac bipolar EGMs, including unipolar EMGs, simultaneous 12-lead surface ECGs and information for the reconstruction of EA maps. Particular attention was paid to meticulously categorizing each EGM to create a multi-expert annotated dataset. Specifically, the current dataset includes a total of 2034 EGMs, distributed into four distinct categories, i.e., AVPs (1022), Physiological (708), Unknown (247), and Undetermined (57). Furthermore, the dataset currently includes an annotation ground truth obtained through majority voting for each EGM class. In future works, we expect to conduct further analysis to assess the labelling performance of each annotator and convene annotators in consensus meetings to establish an agreed-upon ground truth for both annotation and delineation. This extension will offer the possibility of reliably investigating other EGM characteristics, such as AVP temporal localization or specific morphology features associated with AVPs.

Acknowledgments

The authors wish to thank Elisa Pitzalis for her important contribution to the preliminary development of the MATLAB graphical user interface.

This work was conducted within the "ARGO study: Ablation Reinforcement by computer-aided Guidance and Optimization" project, approved by Azienda Tutela Salute Sardegna (ATSSardegna).

The research leading to these results has received funding from the European Union - NextGenerationEU through the Italian Ministry of University and Research under PNRR - M4C2-I1.3 Project PE_00000019 "HEAL ITALIA" to G. Baldazzi, CUP F53C22000750006. The views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

[1] R. Lo, K. K. M. Chia, and H. H. Hsia, "Ventricular tachycardia in ischemic heart disease," *Card Electrophysiol Clin*, vol. 9, no. 1, pp. 25–46, 2017.

[2] A. Radinovic *et al.*, "Matching ablation endpoints to long-term outcome: The prospective multicenter Italian ventricular tachycardia ablation registry," *JACC Clin Electrophysiol*, 2023.

[3] P. Jaïs *et al.*, "Elimination of local abnormal ventricular activities : A new end point for substrate modification in patients with scar-related ventricular tachycardia," *Circulation*, vol. 125, no. 18, pp. 2184–2196, May 2012.

[4] S. K. S. Huang and J. M. Miller, *Catheter ablation of cardiac arrhythmias*. Elsevier Health Sciences, 2019.

[5] G. Baldazzi, M. Orrù, G. Viola, and D. Pani, "Computer-aided detection of arrhythmogenic sites in post-ischemic ventricular tachycardia," *Sci Rep*, vol. 13, no. 1, p. 6906, 2023.

[6] A. Pitzus *et al.*, "Exploring Transfer Learning for Ventricular Tachycardia Electrophysiology Studies," in *Computing in Cardiology (CinC)*, Sep. 2022.

[7] J. Peterson and D. Kynor, "Intracardiac Atrial Fibrillation Database," *PhysioNet*, 2003. https://physionet.org/content/iafdb/1.0.0/ (accessed Jul. 11, 2023).

[8] R. Redina, J. Hejc, D. Pospisil, M. Ronzhina, P. Novotna, and Z. Starek, "Arrhythmia Database with Annotated Intracardial Atrial Signals from Pediatric Patients Undergoing Catheter Ablation," in *2022 Computing in Cardiology (CinC)*, IEEE, 2022, pp. 1–4.

[9] M. Orrù, E. Pitzalis, G. Baldazzi, G. Viola, and D. Pani, "A novel tool for the annotation of bipolar electrograms for cardiac electrophysiology," in *Convegno Nazionale di Bioingegneria*, 2023.

Address for correspondence:

Prof. Danilo Pani
Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy.
danilo.pani@unica.it