A unified framework for multi-lead ECG characterization using Laplacian Eigenmaps

Amalia Villa, Sebastian Ingelaere, Ben Jacobs, Bert Vandenberk, Sabine Van Huffel, Rik Willems, Carolina Varon

***Abstract*—The analysis of multi-lead electrocardiographic (ECG) signals requires integrating the information derived from each lead to reach clinically relevant conclusions. Hence, this analysis could benefit from methods compacting the information in those leads into lower-dimensional representations (i.e., 2 or 3 dimensions instead of 12). In this work, we explore the value of Laplacian Eigenmaps (LE) to find structural and arrhythmogenic changes in the ECG.**

We conceive a normal reference ECG space based on LE, calculated using signals of healthy subjects in sinus rhythm. Signals from new subjects can be mapped onto this reference space creating a loop per heartbeat that captures ECG abnormalities. A set of parameters, based on distance metrics and features to describe the shape of loops, are proposed to quantify the differences between subjects .

This methodology consistently captured the characteristics of healthy ECGs, confirming that normal signals behaved similarly in the LE space. This approach detected significant differences between normal signals, and those from patients with ischemic heart disease or dilated cardiomyopathy. In contrast, the exploration of arrhythmia prediction LE biomarkers did not identify differences between patients with cardiomyopathy and a history of ventricular arrhythmia and their matched controls.

This LE unified framework allows reducing the dimensionality of multi-lead signals while enhancing their abnormalities, enabling the comparison of signals from different subjects in a common low-dimensional space.

***Index Terms*—Laplacian Eigenmaps, ECG signal processing, dimensionality reduction, manifold learning.**

I. Introduction

The electrocardiogram (ECG) is a graphical representation of the electrical activity of the heart and an important diagnostic tool in cardiology. The standard ECG consists of 12 leads containing information of different regions in the heart. The analysis of these multi-dimensional (i.e. multi-lead) signals can be challenging, since the integration of the information from the different leads is not straightforward. It is done by visually exploring the ECG taking into account the spatial information associated to the different leads [1].

AV, SVH and CV received funding from the Flemish Government (AI Research Program) and are affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium. RW is supported as postdoctoral clinical researcher by the Fund for Scientific Research Flanders (FWO Vlaanderen).

AV, SVH and CV are with Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium (e-mail: amalia.villagomez@kuleuven.be).

SI, BV and RW are with Department of Cardiovascular Diseases, Experimental Cardiology, KU Leuven, Leuven, Belgium.

BJ is with Quality by Design NV (QbD), Wilrijk, Belgium.

BV is also with Department of Cardiac Sciences, Libin Cardiovascular Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

CV is also with Service de Chimie-Physique E.P., Université Libre de Bruxelles, Brussels, Belgium.

Laplacian eigenmaps (LE) are a data-driven dimensionality reduction technique that allows identifying low-dimensional structures in high-dimensional data [2]. This algorithm captures non-linear characteristics of the data, while preserving locality-information. This is done by building a similarity graph, where the relations between the data points are emphasized. Using singular value decomposition (SVD), this new representation of the data is reduced to a lower-dimensional space that allows reconstructing the main characteristics of the data. This low-dimensional version of the data is referred to as a Laplacian Eigenmaps space . This approach was first used for the analysis of dynamical changes in multi-lead electrogram signals in time by Erem et al., 2012 [3]. Each heartbeat is represented by a trajectory in the LE space, and authors quantified the changes between consecutive trajectories, which were then related to the effects of ischemia on the electrical properties of the heart. Erem et al. proposed an out-of-sample extension for the LE algorithm, which allowed to first build a LE space in a training phase, and later map other signals onto it [4].

One of the most relevant aspects of LEs is the enhancement of imperceptible changes in the signals. For example, Good et al. showed that LEs capture changes caused by ischemia more accurately than traditional signal-derived features [5], [6]. Moreover, these findings are highlighted by LE already during ventricular depolarization, while traditional metrics can only identify this in the ST segment representing ventricular repolarization, which is vulnerable to noise and artefacts [7], [8].

While these publications have explored dynamic changes within the data of a single subject, the comparison of multiple subjects is not straight-forward. The LE algorithm derives a space from the signals, and the out-of-sample extensions allow to map back future instances of these signals to this original map [4]–[6]. Nevertheless, this LE space is signal-dependent, and hence the morphologies of the trajectories of different subjects in their corresponding spaces cannot be compared.

The goal of this work is to propose a new method that enables the characterization of different subjects in a unified LE space. The main novelty is the definition of a reference LE space onto which new subjects can be mapped, therefore generalizing the out-of-sample extension properties of the algorithm. This space is created using ECG signals recorded during sinus rhythm from presumed healthy subjects with comparable demographics. Our hypothesis is that patterns in multi-lead ECG signals of patients with cardiac disease can be detected in this reference space. The ECGs of patients with cardiac disease are hypothesized to translate in different,

more complex patterns than those of healthy subjects. These differences could be detected by analyzing the trajectories related to each type of subject, or even those of a particular subject at different moments in time. We propose a set of distance metrics to compare different trajectories, and other features to characterize their morphology.

This method was evaluated on the detection of structural and arrhythmogenic signs on the QRS complex of the ECG. First, the framework was applied to discriminate ECGs of healthy subjects from ECGs of patients with ischemic or non-ischemic cardiomyopathy. Additionally, we explored the utility of this approach for the discrimination between cardiomyopathy-patients with ventricular arrhythmia and patients without ventricular arrhythmia. The goal was to capture the differences between these groups in the reference space. This framework is publicly available in https://github.com/avillago/LEnormalSpace[1], where the codes to generate the reference space are available, together with examples of how to map new data onto it and how to derive the distance metrics.

II. Methods

A. Data

This work used ECG signals from three groups. The first population includes healthy subjects, the second gathers patients with ischemic and non-ischemic cardiomyopathy; and the last one includes cardiomyopathy-patients either with or without ventricular arrhythmia. The different populations were matched for age and female to male ratio to the smallest group, i.e the cardiomyopathy-patients with ventricular arrhythmia, more specifically, with a history of electrical storm (ES).

ES is defined as 3 or more appropriate interventions within 24 hours for life threatening ventricular arrhythmia by an implantable cardioverter-defibrillator (ICD) [11]. The ES patients were recruited from the database of patients with an ICD in the University Hospitals Leuven (UZL). This is a single-center registry containing information of all patients with ischemic heart disease (IHD) or dilated cardiomyopathy (DCM) implanted with an ICD between 01/01/1996 and 31/12/2018. The Ethical Committee of UZ Leuven approved retrospective analysis on this database. This database contains 12-lead ECG signals of 10 seconds, sampled at 250 or 500 Hz. Two different subsets of patients were extracted from this database according to their pathology and history of appropriate ICD interventions as surrogate for lethal ventricular arrhythmias. The first subset were patients with ES. Exclusion criteria were age at implantation $\leq$ 18 years, less than 6 months between ICD implantation and ES, ES induced by sepsis, and permanent ventricular pacing. In total, 32 ES patients were identified and matched for age, sex, underlying cardiomyopathy (including both IHD and DCM), and type of prevention with 32 ICD patients without any appropriate ICD intervention (non-ES). Baseline demographics are shown in Table I, under the columns ES and non-ES. The ECG signals considered for the ES patients correspond to the last visit prior to the ES event, in which the signal is in sinus rhythm or atrial fibrillation (AF). The average time to ES of the selected signals is 4 months, with a standard deviation of 2.5 months.

In order to evaluate the framework on structural ECG changes, a subset of signals from ICD patients without any appropriate ICD intervention at the moment of database closure is selected. To exclude QRS-width as confounding factor, patients with a narrow QRS (QRS duration $\leq$ 120ms) were selected from the UZL registry. A total of 64 ECG signals on sinus rhythm or atrial fibrillation (AF) were selected, 32 from IHD patients and 32 from DCM patients. IHD patients have a documented history of myocardial ischemia, while DCM patients were patients with cardiac dysfunction in absence of significant coronary lesions. Baseline demographics of these subjects are presented in Table I.

TABLE I
Baseline demographics of the ICD database from UZ Leuven and the matched normal ECGs from PTB-XL.

mean (SD)	**H**	**IHD**	**DCM**	**ES**	**non-ES**
Age(years)	63.5 (7.9)	65.7 (7.8)	62.2 (5.7)	64.0 (9.5)	63.0 (9.9)
LVEF(%)	NA	35.4 (13.3)†	27.9 (6.8)	38.4 (13.6)	37.8 (11.2)
HR(bpm)	70 (12)	68 (16)	71 (16)	65 (10)	65 (8)
QRS(ms)	78 (14.8)	87.7 (12.7)*	89.4 (12.1)*	120.0 (32.5)*	111.4 (32.9)*
number (%)					
Male	29 (91)	28 (88)	22 (69)	29 (91)	29 (91)
IHD	NA	32 (100)	0 (0)	28 (88)	28 (88)
Sec.Prev.	NA	11 (34)	5 (16)	25 (78)	25 (78)

* Significant differences to the H group † Sig. Diff. between IHD and DCM
LVEF Left ventricular ejection fraction, **HR** Heart Rate, **QRS** QRS duration, **Sec. Prev.** Secondary Prevention

The definition of the normal space required ECG signals from healthy subjects. These signals were extracted from the PTB-XL database [9], which contains 21837 12-lead ECG recordings of 10 seconds sampled at 500 Hz. These recordings include annotations according to the SCP-ECG standards about diagnostic and rhythm; as well as a small description of abnormalities in the signals [10]. For our aim, we considered a subset of 865 signals from different subjects matching age and male to female ratio to the ES population, including only signals belonging to the normal class, sinus rhythms and without abnormalities. Out of these signals, 32 were selected to build the training space. The details of this group are also shown in Table I, under the column H. Albeit there is no detailed information about the comorbidities of the patients without abnormalities on their ECG, they were considered to be healthy subjects [9]. In the remaining of this paper, the ECG signals of this group are referred to as normal ECGs. Despite the selection of patients with a QRS-width $\leq$ 120 ms for comparison, as it can be seen in Table I, the QRS duration of the different groups was significantly different to that of the H group. Therefore, we performed an additional analysis to address this.

B. Preprocessing

The preprocessing of the signals comprises all the steps applied to the ECGs prior to the use of the LE algorithm. The different steps are illustrated in Figure 1.

First, the heartbeats were detected and delineated. In this work, we focus on changes in ventricular depolarization, hence

[1]The codes and the reference space will be made public upon publication.

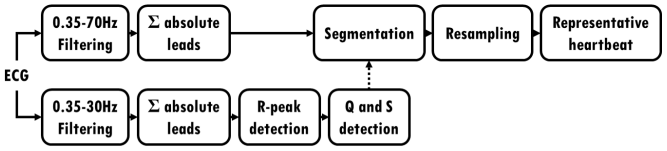

Fig. 1. Preprocessing steps of the ECG signals prior to applying the LE algorithm.

only the QRS complex was considered in the analysis. To segment the QRS complexes, the signals were first bandpass-filtered between 0.35 and 30 Hz, to remove baseline wander and reduce the impact of high-frequency noise. Then the absolute value of each lead was calculated, and the addition of these 12 absolute leads was used to delineate the QRS complexes. First, the R-peaks were identified using the algorithm proposed in the R-DECO software [12], and the onset and offset of the QRS complexes were detected using the ECGkit [13].

Even though high frequencies were removed for the QRS delineation process, the information of those frequency bands is relevant in our analysis. The presence of high-frequency notches during the QRS complex can be associated to fragmentation (fQRS), which is an indicator of high risk of arrhythmia [14]. Therefore, the raw 12-lead ECG signals were bandpass-filtered between 0.35 and 70 Hz, and the 12 absolute filtered leads where added to obtain a single-dimensional signal. The location of the Q and the S waves detected before were imported to this signal, as indicated by the dotted arrow in Figure 1. Once the QRS complexes were segmented, they were resampled to a common length of N samples, here fixed to 200. After that, a representative heartbeat per signal is selected, as the one with the highest correlation with all the other QRS complexes of that recording. This heartbeat in the 12-lead format was further included in the LE analysis.

C. Definition of the reference LE space

One of the main contributions of this work is the definition of a reference space onto which the ECGs of different patients can be mapped and compared in a unified framework. Given that the goal is to detect abnormalities in the QRS complexes, normal ECGs in sinus rhythm, were used to build this space. To exclude age- and gender- dependent changes, a subset of 32 normal ECGs were selected from PTB-XL, belonging to subjects that were matched based for their age and sex to the patient databases described in Section II-A.

Using the signals of these subjects, a low-dimensional space was defined, where the underlying manifold structure of the data is captured. This means that the most relevant characteristics of the data are emphasized in this alternative space, while reducing the number of dimensions. The training and explicit mapping stages defined in [4] can be used here to calculate the space, and further used to map new data onto it.

The first step is to concatenate the representative heartbeats of each of the 32 normal ECGs. This data can be referred to as a set of points $X = \{x_i\}_{i=1}^{L} \in \mathbb{R}^{L\times d}$, with L the number of concatenated samples, i.e., 200 samples per heartbeat, 32 times; and d the number of dimensions (i.e., 12 channels) in the data. Both d and L can take different values according to the application. For the case of 12-lead ECG signals, the limiting factor in terms of computation is the number of samples L, since it determines the size of the matrix that needs to be stored in the memory and manipulated to derive the SVD. The structure of the data is shown in Figure 2. To emphasize the non-linear relations between the points X, a similarity matrix W is built by calculating the pairwise kernel similarity between the points. Each entry of this matrix $W \in \mathbb{R}^{L\times L}$ is calculated as:

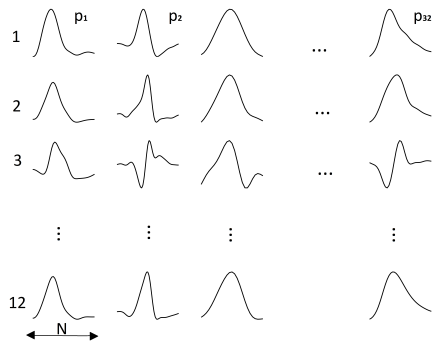

Fig. 2. Structure of the data used to build the matrix required for the LE algorithm. The dimension L corresponds to N samples per each reference heartbeat per subject (here N fixed to 200 and the number of patients to 32). The number d of dimensions is set to 12, the number of leads.

$$W_{ij} = \frac{-exp(||x_i - x_j||_2^2)}{\sigma^2}, \tag{1}$$

where i and j take values between $1,\ldots,L$; $W_{i,i} = 1$, $W_{i,j} = W_{j,i}$, and σ is the radial basis function (RBF) kernel parameter. In matrix W, those pairs of points that are closer in the high-dimensional space receive a higher value $W_{i,j}$ than those that are very distant. The value of σ is set to the maximum $||x_i - x_j||_2^2$ value as suggested in [4]. Then the degree matrix D is computed as $D = diag(\sum_i W_i)$. This diagonal matrix characterizes the relevance of each point x_i in the manifold.

These matrices allow to formulate the singular value decomposition problem, defined as:

$$D^{-1}W = USV^T, \tag{2}$$

which allows to calculate the matrix V, where each column corresponds to one dimension of the Laplacian Eigenmaps space. These columns are sorted according to their associated singular values contained in S, from high to low values according to the amount of information they capture of the high-dimensional space. Since the first column is constant, the next ones are to be considered when constructing the LE space. In line with previous work, the next three dimensions are used to define the space [2]. These three dimensions capture $>$90% of the information contained in the data, according to the singular values, and they allow intuitive visualization of the data [2]. The LE space is then defined by 3 vectors:

$$LE = [V_2, V_3, V_4]. \tag{3}$$

Additionally, the LE space is characterized not only by its vectors V_2, V_3, V_4, but also by a template trajectory or loop.

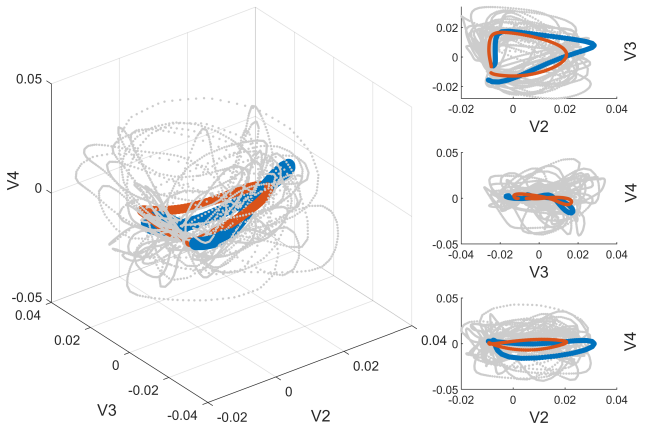

Fig. 3. Trajectories of the ECGs used to build the reference space (grey). The template is shown in blue, and it corresponds to the trajectory that is closest to the average of all the training loops (in orange).

The representative heartbeats of each of the normal ECGs representing this LE space can be mapped onto it, creating a trajectory, as those observed in grey in Figure 3. All the loops fall within a manifold that could be represented by a template loop. This template is selected as the normal loop out of the 32 closer to the mean of all those used to create the space according to the euclidean distance. This mean is calculated point by point, obtaining an average loop of the same length as those of the different subjects. Figure 3 presents in grey the loops of the normal ECGs used to build the LE space, the template loop in blue, and the average used to specify the template, in orange.

D. Explicit mapping

As suggested in [4], the decomposition formulated in Equation 2 can be used to project multi-lead signals of new patients onto this manifold. In our case, the signal to be projected is the representative heartbeat of a new patient, extracted as explained in Section II-B.

The explicit mapping function is defined as

$$f(y) = S^{\dagger}U^{T}D^{-1}g(y), \tag{4}$$

where

$$g(y) = [exp(-\frac{||y-x_1||_2^2}{\sigma^2})...exp(-\frac{||y-x_N||_2^2}{\sigma^2})]^T, \tag{5}$$

and $S^{\dagger}$ is the Moore-Penrose pseudo-inverse of matrix S. This results in the projection of any new point y onto the reference LE space.

The new heartbeats mapped onto this space also form a trajectory that can be compared to that one of the normal template heartbeat to identify differences. Figure 4 shows the examples of projected reference heartbeats for 10 ES and non-ES patients respectively. Additionally, the differentiable characteristics of this formulation allow mapping points back from the LE space to the original multi-lead signal.

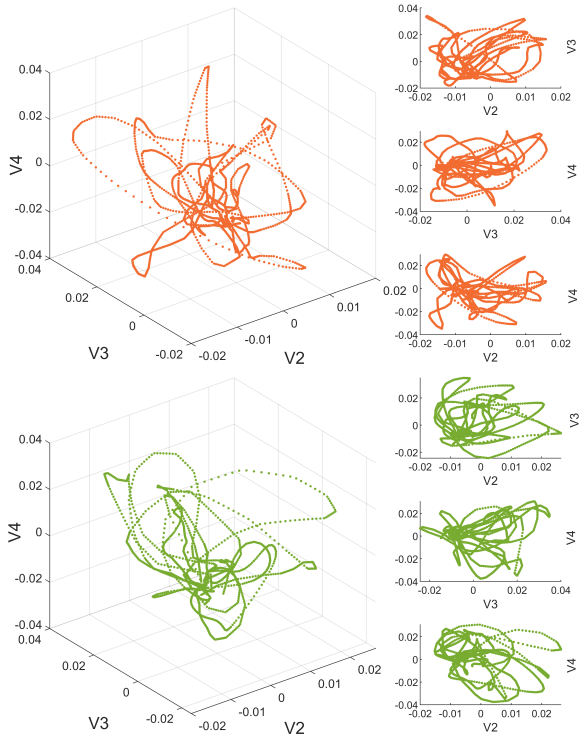

Fig. 4. Examples of reference heartbeats explicitly mapped on the LE space for ES (orange) and non-ES patients (green).

E. Feature extraction

By creating the reference space and applying explicit mapping to the signals of interest, each of these is then characterized by a loop in the LE space. The goal is to identify specific characteristics of the patients that capture the complexity of their signals. Here, this complexity is represented by the differences between the trajectories of normal ECGs and those of the different patient groups.

Two types of features are proposed for this aim, namely distance metrics and features that intrinsically characterize the trajectories. They are summarized in Table II.

TABLE II
FEATURES DERIVED FROM THE LE TRAJECTORIES

Features	Notation
Distance metrics	
Point-to-point	d_{P2P}
Dynamic Time Warping	d_{DTW}
Hausdorff	d_{Haus}
Loop characterization	
Range V2	r_{V2}
Range V3	r_{V3}
Range V4	r_{V4}
SD V2	σ_{V2}
SD V3	σ_{V3}
SD V4	σ_{V4}
Mean curvature	$\overline{curv}$
SD Curvature	σ_{curv}

1) Distance metrics: These features characterize differences as the distance between two trajectories. Following our previous work [15], three of the metrics proposed are further explored in this study. The first is the point-to-point

distance which, given that all loops are resampled to the same length, calculates the euclidean distance between each pair of points. The final distance is the average of all the distances d_i, with $i = 1,...,L$. The second distance is based on Dynamic Time Warping (DTW) [16]. This allows a more flexible comparison of the loops by first finding their optimal match and then calculate the euclidean distance between the aligned time series. Similarly to the first metric, the final distance is calculated as the average of all the distances along the loop. The last metric is the Hausdorff distance, which is widely used to calculate the similarity between two sets of points [17]. This metric is calculated by finding, for each point x_i in trajectory A, its closest point in trajectory B, where trajectories A and B are the representative loops of two different patients A and B. The same is done for each point in B to trajectory A. The final Hausdorff distance is the maximum among all these shortest distances.

2) Loop characterization: The trajectories in the LE space can also be independently characterized, i.e. without comparing them to any other. Given the formulation of the LE problem and its relation with kernel PCA, the values of the trajectories in each dimension are related to their corresponding singular value and hence, to the amount of information that they capture. Therefore, the hypothesis here is that the simpler loops, corresponding to normal ECGs, should contain most of their information in $V2$. On the other hand, complex loops are those that differ from a normal sinus rhythm, and they shall distribute their information along the other dimensions. To capture this, the metrics proposed are the range and the standard deviation of the loops in each of the 3 LE dimensions. The range r_{Vi} represents the difference between the maximum and minimum value of the points in each of the LE dimensions, and the standard deviation σ_{Vi} represents the dispersion of those values from the average. Additionally, the mean and the standard deviation of the curvature measured along the loop are also proposed to characterize the morphology of the loops mapped. The curvature for each point x_i of the loop is the inverse of the radius circumscribing points x_{i-1}, x_i and x_{i+1} [18]. Therefore, the average and the standard deviation of these values along the vector characterize the roundness of the loops in the space.

F. Experiments

In order to explore the potential of the proposed reference LE space to characterize different groups of patients, three experiments were performed. Firstly, we evaluated if the reference space is capturing normal properties of the signals. The second goal was to test the ability of the reference space to distinguish between ECGs of patients with different underlying pathologies, namely IHD and DCM. Finally, the last experiment evaluates the capability of the methodology to identify risk signs of arrhythmia. The reference space for all these experiments is the same, since all the groups of patients were matched on a population level.

1) Capturing characteristics of normal ECGs: The first step to assess the potential of this relative space was to verify if it was able to capture characteristics of normal ECGs. The hypothesis was that all normal sinus rhythm ECGs from healthy subjects of comparable age should present similar properties in this reference space.

To prove this, 10 separate groups of 32 normal ECGs, referred to as H_1,H_2,... H_{10}, were mapped onto the reference space. There was no repetition of subjects between the different groups, neither with those included in the creation of the reference space. For each of these groups, the features proposed were calculated and evaluated to check if all the 10 groups belonged to the same population.

2) Identifying cardiovascular diseases: differences between IHD and DCM: The second experiment aimed to verify if the LE reference space could distinguish between ECGs from healthy subjects and ECGs from patients of comparable age and gender with underlying cardiovascular diseases and a narrow QRS. To this end, the group of patients with no history of shocks from the UZL was mapped onto the reference space, including 32 IHD patients and 32 DCM patients.

The IHD patients considered have a history of myocardial infarction. Impaired coronary flow, mostly due to coronary atherosclerosis, can lead to hypoxia and subsequent cell death of cardiomyocytes in the region of the affected coronary artery [19]. This results in localized scar formation, which affects the conduction characteristics of the heart. Pathological Q-waves are a well-known ECG biomarker of previous infarction [20]. The second group of patients contains DCM patients. While in IHD the problem arises at the level of the coronaries, DCM primarily affects the heart muscle. There is a variety of causes that can underlie DCM, but they all lead to left or biventricular dilatation with systolic dysfunction [21]. Imaging is necessary for diagnosis, but some ECG biomarkers have been proposed to aid in the diagnostic work-up, for example low QRS voltages due to loss of cardiomyocytes, presence of left bundle branch block and T-wave inversion [22].

In this experiment, the goal was to evaluate if these two groups present different properties in the LE reference space, as well as if these are different than those of healthy subjects. The hypothesis is that IHD ECGs should be associated to the most complex loops in the LE space, in line with their morphology in the 12-lead format. On the contrary, DCM trajectories are hypothesized to differ from healthy subjects more subtle.

3) LE biomarkers for arrhythmia prediction: electrical storm: The last experiment explored the use of the LE reference space to identify patients at risk of electrical storm. For this aim, the 64 patients included in the ES study were mapped onto the reference space. Half of these patients had a history of ES and the other half did not. Both groups included IHD and DCM patients, in a similar proportion. Signals from ES patients were hypothesized to present possible signs associated with arrhythmia, such as fragmentation [23], [24].

G. Statistical tests

The LE features were calculated for the different groups of patients. Significant differences between these groups were evaluated using the Kruskal-Wallis test ($p<0.05$) with Bonferroni correction for multiple comparisons. Correlation tests between features were evaluated using Pearson correlation.

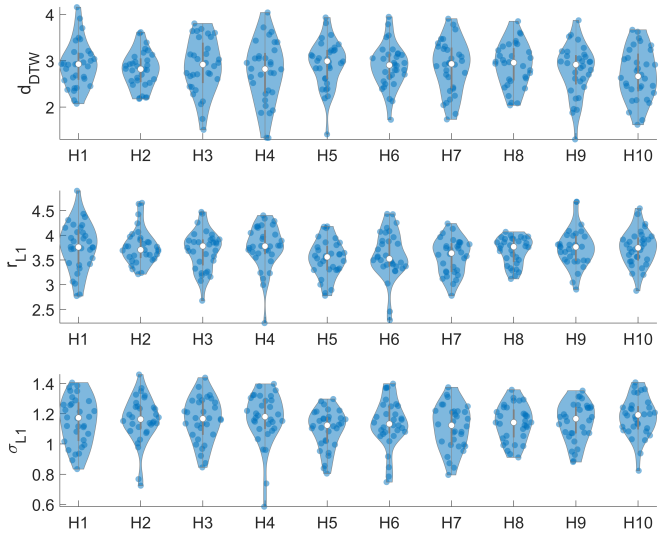

Fig. 5. Violin plots for d_{DTW}, r_{V2} and σ_{V2} for the 10 groups of normal ECGs.

TABLE III
RESULTS FOR H_1, IHD, DCM, ES AND NON-ES. THE AVERAGE AND STANDARD DEVIATION VALUES OF THE FEATURES ARE PRESENTED, TOGETHER WITH THE STATISTICAL COMPARISONS PER EXPERIMENT (I.E. STRUCTURAL OR ARRHYHMOGENIC DIFFERENCES COMPARED TO H_1). DIFFERENCES ARE SIGNIFICANT FOR P < 0.05.

Features	H_1	IHD	DCM	ES	non-ES
d_{P2P}	3.4 ± 0.8	3.9 ± 0.8*,†	3.4 ± 0.8	4.3 ± 0.6*	3.9 ± 0.6
d_{Haus}	2.7 ± 0.5	3.0 ± 0.7	2.8 ± 0.6	3.2 ± 0.6*	2.9 ± 0.5
d_{DTW}	2.9 ± 0.5	3.5 ± 0.8*,†	3.0 ± 0.7	3.8 ± 0.7*	3.6 ± 0.6*
r_{V2}	3.8 ± 0.5	2.8 ± 0.8*	3.0 ± 0.7*	2.8 ± 0.8*	2.7 ± 0.8*
r_{V3}	4.0 ± 1.2	4.0 ± 0.9	4.5 ± 0.8	3.8 ± 1.2	3.7 ± 0.9
r_{V4}	4.7 ± 0.2	5.0 ± 1.1	4.8 ± 1.2	4.7 ± 1.1	4.6 ± 1.4
σ_{V2}	1.1 ± 0.6	0.7 ± 0.3*	0.8 ± 0.2*	0.7 ± 0.3*	0.7 ± 0.3*
σ_{V3}	1.1 ± 0.4	1.1 ± 0.3	1.2 ± 0.3	1.0 ± 0.3	1.0 ± 0.2
σ_{V4}	1.1 ± 0.1	1.2 ± 0.3	1.1 ± 0.3	1.2 ± 0.3	1.1 ± 0.3
$\overline{curv}$	27.0 ± 7.3	29.8 ± 10†	34.6 ± 7.7*	30.0 ± 14.6	29.9 ± 12.0
σ_{curv}	29.7 ± 9.3	28.9 ± 9.3	31.2 ± 9.6	33.5 ± 27.8	36.2 ± 30.8

* Significant differences to the H group † Sig. Diff. between IHD and DCM

III. RESULTS

A. Capturing characteristics of normal ECGs

The first experiment comparing 10 different groups of normal ECGs in the reference space did not find any significant difference between the 10 groups. Figure 5 shows the violin plots of the different normal ECG groups for some of these features. This confirmed that the reference space was able to reproducible characterize normal ECGs. Therefore, in the following experiments the groups of patients are compared to the group H_1 of normal signals, that can be considered as a representation of the PTB-XL database in the LE space. The average and standard deviation of the LE features for the H1 group are shown in Table III.

B. Identifying cardiovascular diseases: differences between IHD and DCM

The features obtained for the IHD and DCM group are shown in Table III. The symbol $*$ indicates significant differences (p-value < 0.05) between IHD or DCM patients and normal ECGs; and † between the IHD and DCM groups. The violin plots related to the features showing significant differences are presented in Figure 6, where these differences are indicated with $*$. As can be observed, the distance metrics identified significant differences between IHD and the other two groups, while the mean curvature $\overline{curv}$ of the trajectories allowed to distinguish between all the three groups. Additionally, features related to the first dimension of the LE space (r_{V2} and σ_{V2}) were significantly different between normal ECGs and the two groups of patients. However, this last distinction was also reflected in the significant differences between QRS duration of normal ECGs and patients shown in Table I.

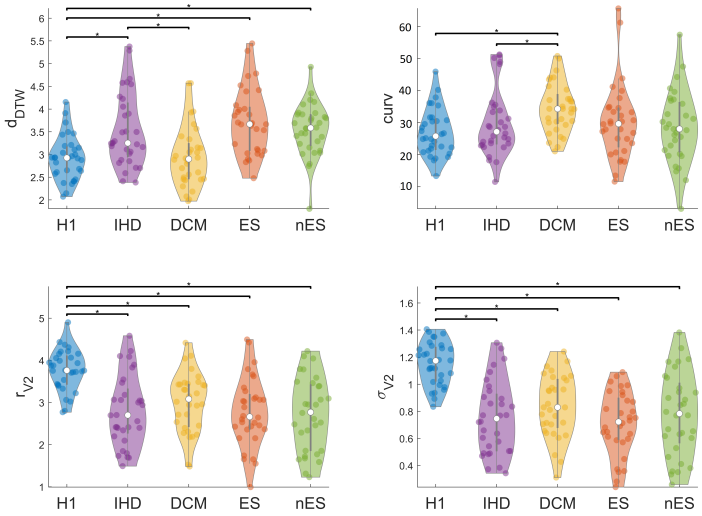

Fig. 6. Violin plots for the features presenting significant differences for experiments 2 and 3. H1 presents the values for normal ECGs, compared to those of IHD and DCM patients, and to ES and non-ES. Significant differences are indicated as $*$ for p-value < 0.05.

In view of the significant differences in QRS duration between the groups, the correlation between QRS duration and each feature was calculated, obtaining the highest values for the standard deviation of the values in the first LE dimension σ_{V2} (-0.27), and for the average curvature (-0.29). The upper plots of Figure 7 show the relation between these features and QRS duration. The shaded areas give a visual intuition of the separation that could be done between normal trajectories (blue points) and the patient groups (purple and yellow symbols).

C. LE biomarkers for arrhythmia prediction: electrical storm

The last experiment evaluated the use of the LE methodology for the characterization of patients developing ES. The average values of the features for both ES patients and their control group are shown in the last columns of Table III. The LE features could not identify significant differences between the ES and non-ES group. Nevertheless, similarly to the results of the experiment described above, the space was able to distinguish between normal ECGs and these patients. The most relevant features for this distinction are the distance features and those related to V_2, both the range and the standard deviation. However, in this case the correlation of these features with QRS duration is stronger than in the previous experiment, since these patients have much wider QRS complexes than healthy subjects. The correlations with

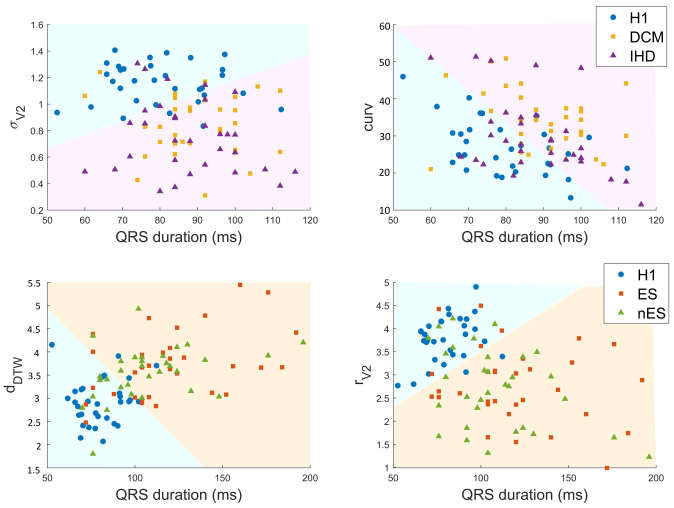

Fig. 7. Relation between QRS duration and the features for the cases of higher correlations. In the top row, σ_{V2} and $\overline{curv}$ for Experiment 2 exploring DCM and IHD patients, and on the lower row d_{DTW} and r_{V2} for Experiment 3, exploring ES and non-ES patients.

QRS duration for d_{P2P}, d_{DTW}, r_{V2} and σ_{V2} ranged between 0.44 and 0.57. Figure 7 shows the relation of d_{DTW} and r_{V2} with the QRS duration. Despite this weak correlation, the combination of QRS duration and these features contributes to differentiate between normal ECGs (blue circles) and both ES and non-ES signals (orange and green signs), especially for QRS duration below 100ms.

IV. Discussion

This study explored the potential of building a reference Laplacian Eigenmaps space based on normal multi-lead ECG signals. ECGs of other subjects can be projected onto this 3-dimensional space creating a trajectory per heartbeat. These trajectories capture the information of 12-lead ECGs and link it to the manifold information, representing the ECG in a lower-dimensional space.

The first experiment allowed to confirm that the LE reference space captures the similarities between normal ECGs. This was tested by projecting 10 groups of signals from the PTB-XL database, and comparing the features obtained for each of these groups. No significant differences were found between them, confirming that the method was able to detect that they all belonged to the same population. Thus, the LE features and the reference space proved to represent the characteristics of normal ECG signals.

Secondly, the methodology was applied to identify differences between three groups of ECG signals: those belonging to patients with IHD, DCM, and normal ECGs. The distances between the derived trajectories and the normal template were found to be significantly larger for IHD ECGs than for normal ECGs or those recorded from DCM patients. This indicates that the trajectories of IHD patients are more different from those derived from normal ECGs and DCM patients. While IHD and DCM are distinguishable using gadolinium enhanced cardiac MRI, where DCM often presents with patchy scars, this distinction is not yet possible using the QRS complexes of the ECG [25]. In IHD the electrical propagation over the heart is disturbed by scar tissue that affects the conduction through the ventricles, which leads to changes in the ECG [19], [20], [24]. Nevertheless, most of the well-known characteristics of IHD ECGs involve irregularities at the end of the QRS complex, which are difficult to detect, often presenting high false-positive rates [26]. Since we only considered QRS complexes in the analysis, the proposed methodology is able to identify differences in ventricular depolarization. This is in line with the results reported in [5], where LE features were able to detect anomalies in ischemic ECG signals during the QRS complex, while traditional metrics required the analysis of longer segments including the T wave. Nevertheless, [5] and [6] study the evolution of patients over time, without comparing the loops of different subjects. Our methodology enables the analysis of different subjects in a unified framework, as well as the progress of these signals over time in a similar trend to previous works [5], [6]. This detection of signs of ischemia in the analysis of the QRS complex may suggest a big advantage of LE methods especially in noisy signals, since the low amplitude of the T-wave makes it more sensitive to artefacts [7], [8].

Additionally, the analysis of the mean curvature of the trajectories of the three groups showed significant intergroup differences, with higher values for DCM signals, followed by IHD and the lowest values obtained for normal ECGs. This metric can be associated with the complexity of the ECG signal: higher values of curvature describe more convex curves, while a regular QRS complex is expected to be characterized by a smooth rounded trajectory. Therefore, the most complex loops should be those of cardiomyopathy patients, and the simplest those related to normal ECGs. The discovery of significant differences in this aspect between healthy and DCM ECGs is nevertheless remarkable, since the characteristics of DCM ECGs are not evident by visual inspection [27]. Nevertheless, the presence of diseased cardiomyocytes and patchy scar in the heart of patients with DCM can be the cause of micro-fragmentation in the QRS complex [28], which could be captured by the LE algorithm.

These differences were also captured by features r_{V2} and σ_{V2}, which characterize the trajectories in the first dimension of the LE space. This first dimension, due to the properties of the singular value decomposition, is the one that contains most of the information when the trajectories are simple, i.e. more similar to a normal regular sinus rhythm. More complex heartbeats, which may include fragmentation or conduction delay will additionally contain more information in the second and third dimension. This rationale has been considered in the decomposition of the ECG using SVD to characterize micro-fragmentation [29], and it can be applied to the Laplacian considered here. This intuition is shown in the results for r_{V2} and σ_{V2}, which are higher for normal ECGs than for DCM and IHD patients. While these features were slightly correlated to the duration of the QRS complex, as it can be seen in Figure 7, both aspects are complementary for the distinction between normal and abnormal ECGs.

Finally, the value of this methodology to detect features related to the future occurrence of arrhythmia, was tested. The

goal was to detect differences between the ECG signals of patients who presented with electrical storm, and those who did not. However, none of the described features could detect differences between these two groups of patients. Currently no clear ECG biomarkers for risk prediction of this serious arrhythmic event are known. While previous work linked the manifestation of ES to the presence of QRS fragmentation [23], this is not visible in this limited database. Further studies should explore comparable times between recordings, to reduce the impact of the time to event, which was variable and long in our series. Additionally, our dataset mixed patients with IHD and DCM, which have different LE trajectories. Therefore, future work in this field should try to study the ECGs of both populations separately to discard confounding effects, similarly to how it was done with other non-ECG biomarkers [30], [31].

Similarly to the IHD-DCM study, the ECGs of the patients included in the ES study were found to be significantly different from those of healthy subjects. Nevertheless, these findings were moderately correlated to the duration of the QRS, which was much larger for the group of patients than for the normal ECGs as reported in the literature [32]. Figure 8 shows the representative QRS complexes and the trajectories of the template loop (blue), its closest non-ES (green) and the farthest narrow ES (orange). The ES QRS complex in this case is fragmented in leads II and aVF, while the template and the non-ES signals are not. This could be the reason why the ES trajectory in the LE space contains more information in the third dimension, while the closest cases lay mostly in the horizontal plane (V2 and V3). This case enhances the relevance of the LE space to emphasize these differences, regardless of the QRS duration of the signals.

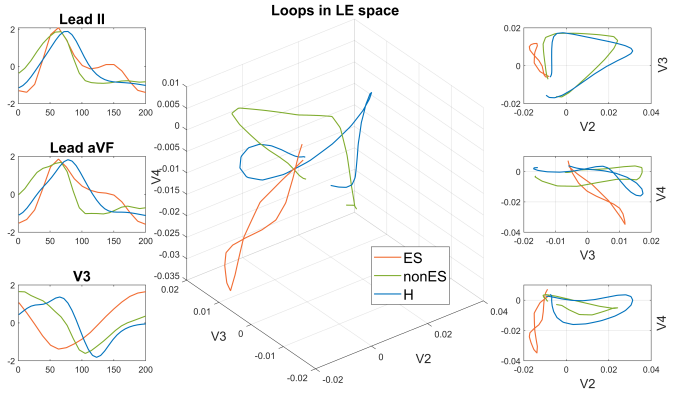

Fig. 8. Example of 3 QRS complexes in the LE space and their corresponding ECGs. The ES signal (orange) corresponds to the farthest trajectory to the template, among the narrow QRS complexes. The green one is the closest narrow trajectory to the template (blue).

The aforementioned experiments confirmed the applicability of the reference LE space and the associated features for the analysis of signals of different subjects in a common framework. Nevertheless, some research questions remain opened. Firstly, the current work exclusively explores changes in ventricular depolarization, since only the QRS complex is included in the analysis. The inclusion of the T wave in the LE algorithm should be considered for further studies. Secondly, the proposed reference space was built using signals of subjects only with ages between 50 and 75, to match these to the age of the ICD population. The behavior of these maps may differ for signals related to different ages, sex and other clinical characteristics, which deserve further investigation. Additionally, when applied to ECG signals, this approach has similarities with the vectorcardiogram, which offers a 3-D representation of 12-lead ECGs [33]. While LE offer the added value of capturing non-linearities and perform a data-driven transformation, exploiting the information contained in the data, future work should explore its relation with VCG. Finally, the proposed approach could be applied to other signals, such as electrograms or electroencephalograms by defining a common reference space built with normal signals in this context.

V. Conclusion

This work proposes a novel approach for the analysis of high-dimensional signals in an alternative space derived using the Laplacian Eigenmaps. This data-driven methodology allows reducing the dimensionality of the space spanned by multi-lead signals while emphasizing abnormalities and differences between subjects.

We evaluated this methodology in the context of 12-lead ECG signals, in the analysis of QRS complexes of signals recorded during sinus rhythm and AF. The methodology was able to capture characteristics of normal ECGs, and to distinguish these from those of patients with ischemic heart disease and patients with dilated cardiomyopathy implanted with a defibrillator in the prevention of sudden death by cardiac arrhythmia. In particular, the identification of differences between normal and DCM ECGs highlights the possible relevance of the proposed approach to identify differences not directly visible in the ECG. This novel methodology opens the path for future analysis of multi-lead ECGs and other high-dimensional signals in a common low-dimensional space, providing a new field of exploration in which subtle ECG abnormalities can be emphasized.

References

[1] Garcia, T. B. (2013). 12-lead ECG: The art of interpretation. Jones Bartlett Publishers.

[2] Belkin, M., Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6), 1373-1396.

[3] Erem, B., Stovícek, P., Brooks, D. H. (2012). Manifold learning for analysis of low-order nonlinear dynamics in high-dimensional electrocardiographic signals. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 844-847). IEEE.

[4] Erem, B., et al. (2016). Extensions to a manifold learning framework for time-series analysis on dynamic manifolds in bioelectric signals. *Physical Review E*, 93(4).

[5] Good, W. W., Erem, B., Zenger, B., Coll-Font, J., Brooks, D. H., MacLeod, R. S. (2018). Temporal performance of laplacian eigenmaps and 3d conduction velocity in detecting ischemic stress. *Journal of electrocardiology*, 51(6), S116-S120.

[6] Good, W. W., et al. (2020). Characterizing the transient electrocardiographic signature of ischemic stress using Laplacian Eigenmaps for dimensionality reduction. *Computers in Biology and Medicine*, 127.

[7] Bortolan, G., Christov, I., Simova, I., & Dotsinsky, I. (2015). Noise processing in exercise ECG stress test for the analysis and the clinical characterization of QRS and T wave alternans. *Biomedical Signal Processing and Control*, 18, 378-385.

[8] Clifford, G. D., Azuaje, F., McSharry, P. (2006). Advanced methods and tools for ECG data analysis (p. 12). Boston: Artech house.

[9] Wagner, P., et al. (2020). PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1), 1-15.

[10] ISO Central Secretary. Health informatics – Standard communication protocol – Part 91064: Computer-assisted electrocardiography. Standard ISO 11073-91064:2009, International Organization for Standardization, Geneva, CH (2009).

[11] Eifling, M., Razavi, M., Massumi, A. (2011). The evaluation and management of electrical storm. *Texas Heart Institute Journal*, 38(2), 111.

[12] Moeyersons, J., Amoni, M., Van Huffel, S., Willems, R., Varon, C. (2019). R-DECO: An open-source Matlab based graphical user interface for the detection and correction of R-peaks. *Peerj computer science*, 5, e226.

[13] Martínez, J. P., Almeida, R., Olmos, S., Rocha, A. P., Laguna, P. (2004). A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Transactions on biomedical engineering*, 51(4), 570-581.

[14] Vandenberk, B., et al. (2017). Inferior and anterior QRS fragmentation have different prognostic value in patients who received an implantable defibrillator in primary prevention of sudden cardiac death. International journal of cardiology, 243, 223-228.

[15] Jacobs, B., Villa, A., Moeyersons, J., Van Huffel, S., Willems, R., Varon, C. (2020, September). Can Laplacian Eigenmaps be used for differentiation between healthy subjects and patients with corrected Tetralogy of Fallot?. In *2020 Computing in Cardiology* (pp. 1-4).

[16] Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, 69-84.

[17] Zhang, D., He, F., Han, S., Zou, L., Wu, Y., Chen, Y. (2017). An efficient approach to directly compute the exact Hausdorff distance for 3D point sets. *Integrated Computer-Aided Engineering*, 24(3), 261-277.

[18] Are Mjaavatten (2021). Curvature of a 1D curve in a 2D or 3D space (https://www.mathworks.com/matlabcentral/fileexchange/69452-curvature-of-a-1d-curve-in-a-2d-or-3d-space), MATLAB Central File Exchange. Retrieved December 9, 2021.

[19] Institute of Medicine (US) Committee on Social Security Cardiovascular Disability Criteria. Cardiovascular Disability: Updating the Social Security Listings. Washington (DC): *National Academies Press (US)*; 2010. 7, Ischemic Heart Disease. Available from: https://www.ncbi.nlm.nih.gov/books/NBK209964/

[20] de Luna, A. B., & Fiol-Sala, M. (2008). Electrocardiography in ischemic heart disease: clinical and imaging correlations and prognostic implications. John Wiley Sons.

[21] Jefferies, J. L., & Towbin, J. A. (2010). Dilated cardiomyopathy. *The Lancet*, 375(9716), 752-762.

[22] Finocchiaro, et al. (2020). The electrocardiogram in the diagnosis and management of patients with dilated cardiomyopathy. *European journal of heart failure*, 22(7), 1097-1107.

[23] Villa, A., Ingelaere, S., Van Huffel, S., Willems, R., Varon, C. (2020, September). Fragmented QRS Dynamics Towards Electrical Storm in ICD Patients. In *2020 Computing in Cardiology* (pp. 1-4). IEEE.

[24] Das, M. K., et al. (2010). Fragmented QRS on twelve-lead electrocardiogram predicts arrhythmic events in patients with ischemic and nonischemic cardiomyopathy. Heart rhythm, 7(1), 74-80.

[25] McCrohon, J. A., et al. (2003). Differentiation of heart failure related to dilated cardiomyopathy and coronary artery disease using gadolinium-enhanced cardiovascular magnetic resonance. Circulation, 108(1), 54-59.

[26] Nadour, W., et al. (2014). Does the presence of Q waves on the EKG accurately predict prior myocardial infarction when compared to cardiac magnetic resonance using late gadolinium enhancement? A cross-population study of noninfarct vs infarct patients. Heart Rhythm, 11(11), 2018-2026.

[27] Schultheiss, H. P., et al.(2019). Dilated cardiomyopathy. *Nature reviews Disease primers*, 5(1), 1-19.

[28] Sha, J., Zhang, S., Tang, M., Chen, K., Zhao, X., Wang, F. (2011). Fragmented QRS is associated with all-cause mortality and ventricular arrhythmias in patient with idiopathic dilated cardiomyopathy. Annals of Noninvasive Electrocardiology, 16(3), 270-275.

[29] Hnatkova, K., et al. (2021). Spatial distribution of physiologic 12-lead QRS complex. Scientific reports, 11(1), 1-20.

[30] Arya, A., et al. (2006). Prevalence and predictors of electrical storm in patients with implantable cardioverter-defibrillator. The American journal of cardiology, 97(3), 389-392.

[31] Takigawa, M., et al. (2010). Predictors of Electrical Storm in Patients With Idiopathic Dilated Cardiomyopathy–How to Stratify the Risk of Electrical Storm–. Circulation Journal, 74(9), 1822-1829.

[32] Chen, X., Hansson, P. O., Thunström, E., Mandalenakis, Z., Caidahl, K., Fu, M. (2021). Incremental changes in QRS duration as predictor for cardiovascular disease: a 21-year follow-up of a randomly selected general population. Scientific Reports, 11(1), 1-6.

[33] Vozda, M., & Cerny, M. (2015). Methods for derivation of orthogonal leads from 12-lead electrocardiogram: A review. *Biomedical signal processing and control*, 19, 23-34.