Topological Analysis and Gaussian Decision Tree: Effective Representation and Classification of Biosignals of Small Sample Size

Zhifei Zhang *Student Member, IEEE*, Yang Song *Student Member, IEEE*, Haochen Cui, Jayne Wu *Senior Member, IEEE*, Fernando Schwartz and Hairong Qi *Senior Member, IEEE*.

***Abstract*—*Goal:* Bucking the trend of big data, in microdevice engineering, small sample size is common, especially when the device is still at the proof-of-concept stage. The small sample size, small inter-class variation, and large intra-class variation, have brought biosignal analysis new challenges. Novel representation and classification approaches need to be developed to effectively recognize targets of interests with the absence of a large training set. *Methods:* Moving away from the traditional signal analysis in the spatiotemporal domain, we exploit the biosignal representation in the topological domain that would reveal the intrinsic structure of point clouds generated from the biosignal. Additionally, we propose a Gaussian-based decision tree (GDT), which can efficiently classify the biosignals even when the sample size is extremely small. *Results:* This study is motivated by the application of mastitis detection using low-voltage alternating current electrokinetics (ACEK) where five categories of bisignals need to be recognized with only 2 samples in each class. Experimental results demonstrate the robustness of the topological features as well as the advantage of GDT over some conventional classifiers in handling small dataset. *Conclusion:* Our method reduces the voltage of ACEK to a safe level and still yields high-fidelity results with a short assay time. *Significance:* This paper makes two distinctive contributions to the field of biosignal analysis, including performing signal processing in the topological domain and handling extremely small dataset. Currently, there have been no related works that can efficiently tackle the dilemma between avoiding electrochemical reaction and accelerating assay process using ACEK.**

***Index Terms*—Biosignal, topology, delay embedding, persistent homology, witness complex, low-voltage ACEK, decision tree.**

I. Introduction

In microdevice engineering, small sample size is common, especially when the device is still at the proof-of-concept stage. In biosensor studies, the repeatability/stability of a device is usually demonstrated by a repetition of 3. Sometimes there is only one data point. This is due to the difficulty in fabricating the device, sample preparation and acquisition/storage, etc., at the laboratory stage. Being a common scenario in the development of novel devices, innovative data processing approaches need to be developed in par with device engineering.

A. Motivating Application – Mastitis Detection by Low-Voltage ACEK

Mastitis is caused by bacterial infection of udder tissues, which causes economic loss because sick cow will eventually reduce milk production and spread to others. Once the infection happens, the immune system of cows will respond and fight the infection with an increase in the number of immune cells, referred to as somatic cells, primarily white blood cells. The number of somatic cells in milk, i.e., somatic cell count (SCC), is an important measure of milk quality used throughout the world [1]. Milk with a high SCC is associated with a higher incidence of antibiotic residues in milk and the presence of pathogenic organisms and toxins in milk. Therefore, high SSC milk is normally rejected by diary processing factory because it can't be used for cheese or others. Annually, mastitis causes approximately $2 billion in losses to the U.S. dairy industry, and 60% of the losses, or $1.2 billion, comes from mastitis.

Cows with *clinical mastitis* are easy to identify with the swollen teats and thick, curdled discharge in the milk, while *subclinical* cases, the most common form of mastitis, are difficult to detect due to no obvious clinical symptoms of the illness and no visible changes to the milk composition. Subclinical mastitis in cows can still lead to abnormally high SCC in milk and can be up to 40 times more than clinical cases of the illness. The primary focus of most subclinical mastitis programs is to reduce the prevalence of the contagious pathogens, S agalactiae and S aureus, as well as other gram-positive cocci. Because mastitis can be caused by many different pathogens, the early availability of diagnostic results, i.e., the pathogens that cause an elevated SCC in a herd as well as their concentration levels, is crucial towards controlling the spread of new infections.

Affinity-based biosensors detect biological particles based on the specific analyte-probe binding between analyte and probe molecules. In recent years, they have been applied to a wide range of applications in many fields, such as immunological study, infectious disease diagnostics, biological warfare, etc. [2]–[4]. Alternating current electrokinetics (ACEK) [5], implemented by microelectrodes immersed in sample fluids, induces directional particle or fluid motion by externally applied AC electric fields over the electrodes. Since its advent in

Manuscript received December 31, 2015; revised September 29, 2016; accepted November 12, 2016.

Zhifei Zhang (e-mail: zzhang61@vols.utk.edu), Yang Song, Haochen Cui, Jayne Wu, and Hairong Qi are with the Department of Electrical Engineering and Computer Science, Fernando Schwartz was with the Department of Mathematics, the University of Tennessee, Knoxville, TN 37996, USA.

1990s, ACEK has been studied and utilized by researchers to accelerate the movement of macromolecules towards sensing areas [6]–[8]. Applying a voltage higher than 1V over biofluids will raise the concerns of electrolysis, biofouling, etc. Using lower voltage, however, cannot efficiently accelerate the assay process, which means we have to wait for tens minutes or even hours before achieving an available result.

Besides the low-voltage challenge, another difficulty of the mastitis detection problem is related to the sample size. During the experimental setup, the samples and electrodes need to be freshly prepared on the day of testing. While the detection process of this method is rapid, low cost and straightforward, the sensor preparation needs to be meticulous and takes long time. In addition, the same as all biomolecular sensors, the sensor can be used only once, so a great amount of work is spent on preparing the sensors manually in a laboratory setting. The number of tests is also limited by the availability of fresh biological samples. Due to these limitations, only two biosignals are obtained from each class. Such a small sample size would fail most conventional classifiers, such as SVM, kNN, decision tree, etc.

B. Experimental Setup

Milk samples with five concentrations of bovine IgG whole molecules (0ng/mL, 1ng/mL, 5ng/mL, 10ng/mL, 100ng/mL) are tested using exactly the same biosensor. The biosignals labeled with the same concentration are considered as the same class. In purpose of demonstrating the voltage effect, the above experiments are repeated under four low-voltage levels (20mV, 40mV, 70mV and 135mV). All data were acquired from measurement of biological samples. The sensing method is ACEK capacitive sensing. The raw data output from the sensor is the sensors serial capacitance. At the end of measurement (i.e., 30 seconds), a microcontroller in the sensor reader converts the capacitance transients into the normalized capacitance change rates, i.e., dlCl/dt, which are the data set used in the manuscript. More response curves from ACEK capacitive sensor can be found in relevant publications [9]–[11]. Those were all taken from measuring clinical samples.

C. Challenges

To learn a classifier that can precisely label biosignals into their corresponding classes, we need to 1) extract representative features from the raw signals and 2) design a suitable classifier to separate the biosignals based on their features.

Since the accuracy of a classifier significantly depends on the features, extracting distinguishable and reliable features from the biosignals is particularly important. The low-voltage ACEK, however, challenges the feature extraction in mainly three aspects: 1) biosignals from different classes may appear similar, 2) biosignals from the same class sometimes are significantly different, and 3) all biosignals show random oscillation with irregular amplitude.

Due to the adoption of the low-voltage ACEK technique to avoid the electrochemical reaction, the biosignals from different concentration levels are more difficult to distinguish. When the voltage is 40mV, for example, the detected biosignals

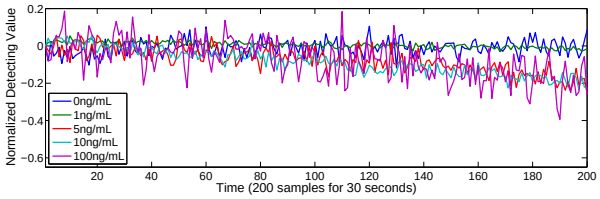


Fig. 1. Biosignals from different concentrations of bovine IgG whole molecules under the voltage of 40mV. The signals are normalized and aligned to start from zero.

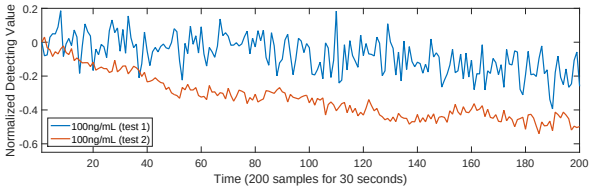


Fig. 2. Biosignals from the same concentration (100ng/mL) of bovine IgG whole molecules under the voltage of 40mV.

from different concentrations/classes present severe overlap, as shown in Fig. 1. This phenomenon is referred to as *small inter-class variation*. The signals are normalized and aligned to start from zero for comparison purpose. In the time domain, it is quite challenging to identity their corresponding concentration levels due to the overlap along the time axis. Statistical methods, such as calculating mean, variance, skewness or kurtosis [12], would yield indistinguishable features because the overlap will be directly passed onto the feature space. The model-based algorithms, like polynomial fitting [13] and logistic regression [14], would cause large error because none of the signals from certain class can be well represented by a fixed model.

On the other hand, biosignals of the same concentration/class but acquired from different experiments might exhibit completely different characteristics, as shown in Fig. 2. This phenomenon is referred to as the *large intra-class variation*. Recent years, dictionary learning and sparse coding [15], [16] have been deeply researched to effectively decompose a mixed signal into a linear combination of multiple source patterns where the sparse coefficients can be used for detection and identification purpose. However, the small inter-class variation may cause different classes to share the same source pattern, and the large intra-class variation may cause the same class to be represented by different source patterns.

The random oscillation refers to the random change in both period and amplitude as shown in Fig. 1. No fixed pattern can be extracted from the oscillation. Therefore, it is unreasonable to apply a filter to smooth the raw signal, or to extract separable features from spectral analysis, such as fast Fourier transform (FFT) or discrete wavelet transform (DWT) [17].

This paper presents a novel signal representation approach based on feature extraction through topological analysis (TA). Conventional signal processing methods, like statistical analysis in the time domain and spectral analysis in the frequency or wavelet domain, fail to extract representative and stable features because the signals in the time domain do not tightly

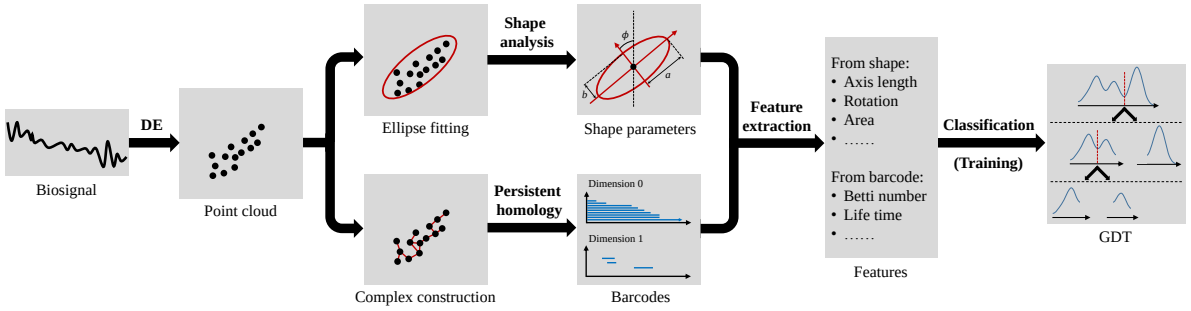


Fig. 3. Flow of the proposed topological signal representation and classification. The raw biosignal in the time domain is converted to two dimensional point cloud in the topological space through delay embedding (DE). Then the flow divides into two branches: upper branch (shape analysis) fits the point cloud with an optimal ellipse, and then extracts parameters of the ellipse; lower branch (persistent homology) constructs witness complex and compute barcodes. Finally, features from upper and lower branches are merged to train the classifier – Gaussian-based Decision Tree (GDT).

associate with their labels (concentration levels), and their spectra are cluttered by random oscillation. Instead, topological method, including delay embedding [18] and persistent homology [19], converts the raw signal into a higher dimensional space (the topological space), which will yield extra features that may effectively separate those originally indistinguishable signals. In addition, topological analysis investigates intrinsic structure of the signals in the topological space, and features extracted from the intrinsic structure are more robust. The second contribution of the proposed work is a Gaussian-based decision tree (GDT) that effectively handles the small sample size problem. The proposed TA-GDT framework allows much lower voltage for ACEK and significantly reduces the risk of electrochemical reaction, while maintaining high fidelity within a short assay time (30 seconds).

The rest of this paper is organized as follows. Section II describes the topological-based feature extraction from the biosignals. Section III elaborates on the Gaussian-based decision tree classifier. Section IV provides experimental study. Section VI concludes the paper. Section V discusses the potential limitations of the TA-GDT framework.

II. Topological Signal Representation

Due to the challenges associated with the biosignals acquired through low-voltage ACEK, conventional approaches would fail to extract distinguishable and reliable features for classification purpose. In this section, we investigate the topological method to analyze the intrinsic structure of the data. There have been existing works adopting topological analysis for biosignal processing. For example, [20] used delay embedding to convert the 1-D ECG signal into a 2-D point cloud, whose structure shows strong pattern. [21] analyzed human speech through its topological structure converted through delay embedding. [22], [23] combined delay embedding and persistent homology to detect wheeze. [24] used the similar way to explore a robust detection of periodic patterns in gene. However, existing works only focus on the signals with strong periodicity. For example, ECG signals possess periodic peaks, and wheeze signals oscillate with almost fixed period. Because our signals do not show discriminative patterns/periodicity, we need to not only extract features through persistent homology, but also analyze the shape of the point cloud, which yields robust feature regardless of periodicity.

The main flow of the proposed topological signal representation and classification is shown in Fig. 3. The raw biosignal is first converted to a 2-D point cloud by delay embedding (Sec. II-A), which has been widely adopted as a powerful tool for geometric time series analysis [25]. Then, shape features are extracted from the ellipse best fitting the point cloud (Sec. II-B). Meanwhile, intrinsic structure of the point cloud is represented through persistent homology (Sec. II-C). Finally, classification performs on these features that will be summarized in Sec. II-D. The classifier, Gaussian-based decision tree (GDT), will be further discussed in Sec. III.

A. Delay Embedding

Delay embedding (DE) transforms a 1-D signal in the time domain to a higher dimensional point cloud in the topological space. Suppose a signal sequence can be represented by a discrete function $f(x),\ x \in \mathbb{Z}^+$ corresponding to sampling index. Choosing a delay step $s \in \mathbb{Z}^+$ and a target dimension of the topological space $d \in \mathbb{Z}^+$, the DE of $f(x)$ at $t \in \mathbb{Z}^+$ can be expressed in Eq. 1.

$$DE(f,t;s,d) = \begin{bmatrix} f(t) \\ f(t+s) \\ \vdots \\ f\left(t+(d-1)s\right) \end{bmatrix} \tag{1}$$

$DE(f,t;s,d)$ transforms a 1-D point $f(t)$ in the time domain to a d-D vector in the topological space. Assume there are n sampling points in a signal sequence. Given s and d, if sliding t from the first to last available point for DE, we will obtain $m = n-(d-1)s$ d-D points, all of which form a point cloud $\mathbf{C}$ as expressed in Eq. 2.

$$\mathbf{C}_{s,d} = \{DE(f,t_1), DE(f,t_2), \cdots, DE(f,t_m)\} \tag{2}$$

Fig. 4 explains the procedure of generating point cloud through delay embedding from a sample time sequence. In the left figure, the red dots denote the sampling points in the time domain. DE parameters are set as follows: the number of sampling points $n = 5$, delay step $s = 1$, target dimension

$d = 2$. When $t = 1$, DE yields a 2-D point $(f(1), f(2))$ according to Eq. 1. The dots with green circles indicate the points used in DE ($t = 1$), and the corresponding 2-D point is shown in the same color in the right figure. When $t = 2$, the points with orange circles are collected to generate the next 2-D point $(f(2), f(3))$, which is shown as an orange dot in the right figure. Iterating t, we will get a point cloud with $m = 4$ points in the topological space.

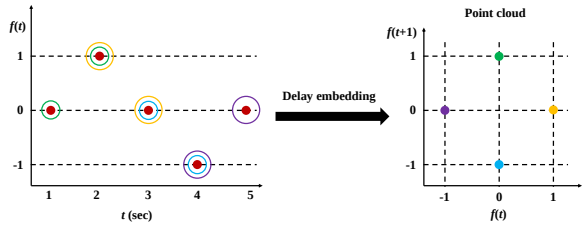
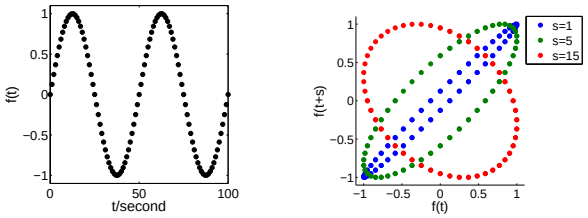


Fig. 4. Delay embedding of a signal ($s = 1$, $d = 2$). Left: red dots denote sampling points in time domain, and circles in the same color indicate the set of points used in DE. Right: point cloud in 2-D topological space, in which each dot is generated through DE from the sample points marked by circles with the same color in the left figure.

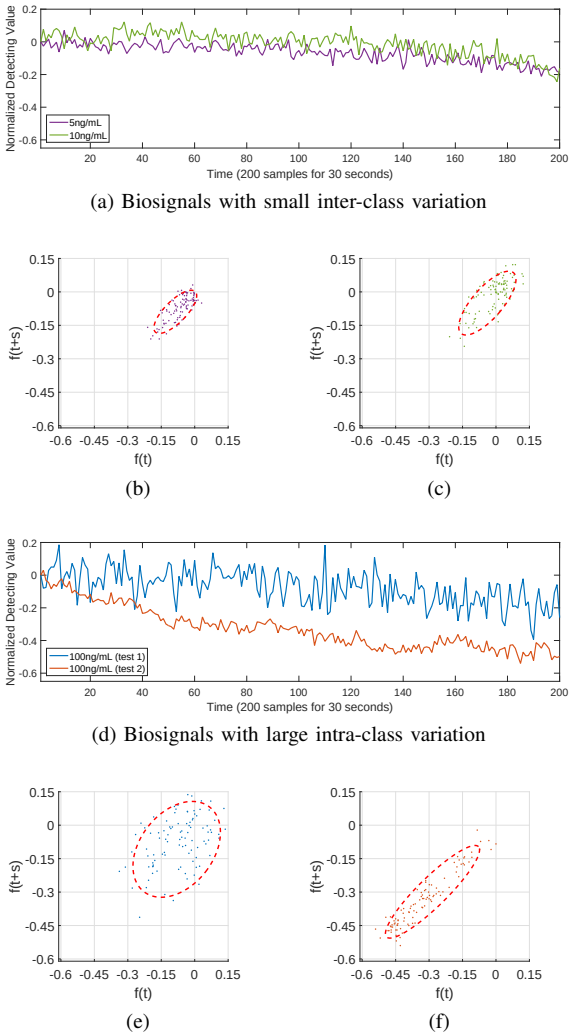
For simplicity, we set d to be 2. Empirically, larger d does not necessarily increase the classification accuracy but certainly reduces the computational efficiency. Then, the only free parameter of DE is the delay step s. Given the same signal, different s will generate different shapes of point cloud as illustrated in Fig. 5, where Fig. 5(a) samples from a synthetic function $f(t) = \cos(2\pi t/T)$ with $T = 50$ sec (sampling frequency $f_s = 1$ Hz). Intuitively, let $d = 2$, Fig. 5(b) displays the corresponding point cloud when $s = 1,\ 5,\ 15$.



(a) Signal (b) Point cloud

Fig. 5. Delay embedding ($d = 2$) on a synthetic periodic signal with different delay step s. (a) The raw signal, and (b) the point cloud corresponding to different s in different colors.

Empirically, a larger point cloud area implies better representation. For the above example, the optimal $s = T \times f_s/2d \approx 13$ [26], which theoretically achieves the largest area of the point cloud. Usually, the point cloud has a regular pattern when the original signal oscillates periodically. However, the real biosignals studied in this paper show random oscillation, whose point cloud does not present like ellipse. Instead, the point cloud forms a solid cluster whose boundary approximates an ellipse, as shown in Fig. 6. Empirical experiments show that modifying s does not make significant difference to the elliptic boundary of the point cloud, so a constant delay step (e.g., $s = 5$) is set, and the shape of the elliptic boundary is analyzed as robust features.



(a) Biosignals with small inter-class variation

(b) (c)

(d) Biosignals with large intra-class variation

(e) (f)

Fig. 6. Shape analysis of biosignals. Point cloud (colored dots) are generated from the biosignals in corresponding colors. Red dash curves display fitted ellipses. Biosignals with small inter-class variation show relatively large variation in the shape of ellipses (length of the major axis). Biosignals with large intra-class variation share similar length of the major axis.

B. Shape Analysis

As shown in Fig. 6, we fit the point cloud to an ellipse using the least square criterion [27] and extract parameters of the ellipse as shape features of the point cloud. Fig. 6 shows two groups of point cloud and the fitted ellipses, as well as their corresponding raw signals. Fig. 6(a) displays biosignals from two different classes, which demonstrate the small inter-class variation. Signals in Fig. 6(d) are from the same classes that exhibit large intra-class variation. Fig. 6(b) and (c) show point clouds and fitted ellipses of the signals in Fig. 6(a). The colors of point clouds correspond to the colors of biosignals. By the same token, Fig. 6(e) and (f) are point clouds and fitted ellipses of the signals in Fig. 6(d). We observe that biosignals with small inter-class variation deviate from each other in the shape of point cloud, especially the length of the major axis. Meanwhile, biosignals with large intra-class variation share similar length of the major axis. This example shows the great potential of shape analysis to yield separable and robust features.

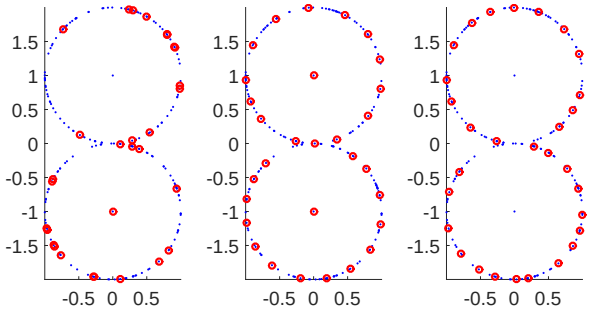
Actually, the shape features reflect important time-domain characteristics of the raw signal. For example, the length of the major axis associates with the amplitude range of the raw signal, the length of the minor axis implies the strength of oscillation, and the center location of the ellipse indicates the slope of the line that is fitted to the raw signal. Comparing Fig. 6(e) and (f), their raw signal should share similar amplitude range, but significantly differ in the strength of oscillation. If the raw signals of Fig. 6(e) and (f) are fitted to straight lines, respectively, the latter should have smaller slope. Beside the shape features, the intrinsic structure of the point cloud also potentially provides distinguishable features, which will be explored through the persistent homology.

C. Persistent Homology

Persistent homology is a method for computing topological features of the point cloud at different spatial resolutions [28], [29]. This section investigates the intrinsic structure of the point cloud through persistent homology (Sec. II-C3), before which a connected graph of the point cloud has to be constructed via certain criterion (Sec. II-C2). Because the computation of persistent homology is very time-consuming, we propose a cluster-based downsampling method to reduce the size of the point cloud while preserving its geometric structure, which will be first discussed in Sec. II-C1.

1) Downsampling the Point Cloud: To investigate the intrinsic structure of the point cloud, persistent homology analysis needs to be applied on these points. Because the computational complexity increases exponentially with the number of points, we select some representative points from the point cloud to approximate its structure. These selected points are referred to as the *landmarks*. There have been two common methods for landmark selection, namely, random selection [30] and sequential maxmin [31], [32]. As shown in Fig. 7, the random selection may lose the intrinsic structure of the point cloud, and sequential maxmin tends to spread every corner, involving outliers (the two blue dots at the center of the "8"-shaped point cloud). We propose a cluster-based landmark selection method, which selects landmarks according to the density distribution of point cloud. Points in high-density areas are more likely to be selected as landmarks, and outliers tend to be ignored. Specifically, the mean-shift clustering method [33] would first group the point cloud into multiple clusters, and the cluster centers are considered as candidates of landmarks. Then, those candidates with small cluster size (low density) are discarded, and the remaining candidates are selected as landmarks. Fig. 7(c) demonstrates the better performance of cluster-based landmark selection.

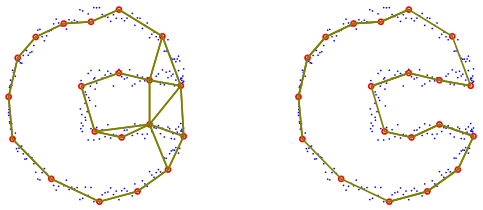
2) Criterion of Building Connected Graph: To prepare for the persistent homology, the landmarks need to form a connected graph through certain connection criterion, just like the complex construction in Fig. 3. A common way of building a connected graph is to connect point pairs whose Euclidean distance is less than or equal to a threshold ϵ. The simplicial complex constructed by this manner is called Rips complex [34]. Rips complex builds a sort of neighborhood graph, which tends to approximate the original structure of



(a) Random selection (b) Sequential maxmin (c) Cluster-based

Fig. 7. Different methods of landmark selection based on a synthetic "8"-shaped point cloud, adding two outliers at the centers. Blue dots are points, and red circles indicate the points selected as landmarks. (a) random selection, from which it is difficult to recover the "8"-shaped structure. (b) sequential maxmin that uniformly chooses landmarks which well represent the intrinsic structure of point cloud, but the outliers are also selected. (c) the cluster-based selection that can kick out outliers and preserve density information.

point cloud. However, it may cause "over connection" as shown in Fig. 8(a). Based on the landmarks (red circles), Rips complex connects all point pairs whose distances are no larger than certain ϵ. Obviously, some connections are unreasonable because they destroy the original structure of point cloud (blue dots). By contrast, the complex in Fig. 8(b) better represents the intrinsic structure by removing those connections surrounded by only a few points. The simplicial complex shown in Fig. 8(b) is referred to as the witness complex [35], [36], which utilizes the witnesses (non-landmark points) to evaluate each simplex (connection). If the number of witnesses for a simplex is lower than certain threshold (e.g., 2), it will be removed from the constructed simplicial complex.



(a) Rips complex (b) Witness complex

Fig. 8. Comparison of Rips and witness complex with the same landmarks and ϵ. Blue dots are synthetic point cloud and red circles are landmarks. (a) constructs the complex only using landmarks regardless of blue dots, so it cannot exactly reflect the structure of the point cloud; (b) utilizes the rest blue dots to estimate whether a simplex (connection) is reasonable to be there. A simplex without enough witnesses will be removed.

3) Persistent Homology: Given a point cloud and the corresponding landmarks, persistent homology tries to find persistent features from a group of witness complex with increasing threshold ϵ. Fig. 9 illustrates the basic idea of persistent homology. In general, persistent homology tracks a series of witness complex, whose variant may follow certain patterns. With the increase of ϵ, as shown in Fig. 9, more point pairs whose Euclidean distance is less than or equal to ϵ are connected (red lines) by following the construction rule of witness complex. At the same time, any convex hull constructed by points whose pair-wise distances are not larger

than ϵ is filled as shown by the shaded areas. Note that a hole forms at $\epsilon = \epsilon_3$, persists to $\epsilon = \epsilon_4$ and is filled (disappears) at $\epsilon = \epsilon_5$. In the view of independent connected component, there are five independent connected components (black dots) at $\epsilon = \epsilon_0$, then the number reduces to three ($\epsilon = \epsilon_1$), two ($\epsilon = \epsilon_2$), and one for $\epsilon \geq \epsilon_3$.

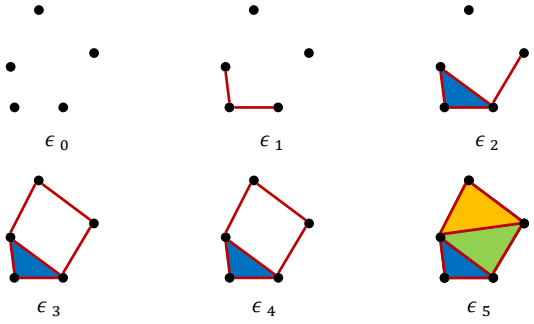
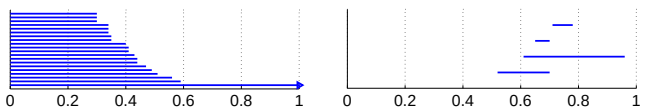


Fig. 9. Persistent homology on a 2-D point cloud. With the increase of ϵ, more points are connected, and more simplices are constructed. At ϵ_3, a hole is formed, then the hole is filled (disappears) at ϵ_5. Thus, $\epsilon_5 - \epsilon_3$ indicates the lifetime of the hole.

As described above, persistent homology tracks two dynamic processes: 1) how fast those independent connected components merge to a single component and 2) when a hole appears and disappears. Specifically, the holes are of paramount importance for discovering the intrinsic structure of a point cloud [19], [37]–[39]. As the "8"-shaped point cloud shown in Fig. 7, two holes are supposed to appear regardless of variants of scale, location, rotation, noise, or geometric distortion. Therefore, what we are interested are at what value of ϵ a hole appears and how long it persists until being filled (disappear). In the 2-D space, assuming a hole appears at ϵ_{birth} and dies at ϵ_{death}, the length it persists is $\epsilon_{death} - \epsilon_{birth}$ that is referred to as the lifetime of the hole. For any independent connected component, it is born at ϵ_0 and dies when it is merged. In the computational topology, the birth and death points, as well as lifetime, are illustrated through barcodes as displayed in Fig. 10, which is computed from Fig. 8(a). The barcodes of dimension 0 indicate the lifetime of independent connected components, and dimension 1 shows the lifetime of holes. The start and end points of each bar indicate the birth and death points. The horizontal axis indicates the ϵ. As ϵ increases and approaches infinity, more and more components merge together and finally combine into a single component which is denoted by the longest bar with an arrow. Because the lifetime of a hole quantifies its significance [40], longer bars are expected. In this paper, the lifetime of holes and independent connected components are extracted as features.



(a) Dimension 0 (b) Dimension 1

Fig. 10. Barcodes of Fig. 8(a). The horizontal axis of the barcodes denotes the value of ϵ. The length of bars indicates the lifetime of each (a) independent connected component in dimension 0 or (b) hole in dimension 1.

D. Summary of Topological Features

First, all raw biosignals are rescaled and aligned to start from zero. Then, each signal is converted to 2-D point cloud through DE. We do not convert the signal to higher dimension for two reasons: 1) higher dimension results in higher computational complexity which increases exponentially with the number of dimensions, and 2) empirical experiment shows that higher dimension cannot provide more representative features. Based on the 2-D point cloud, features are extracted from both shape analysis and persistent homology. From shape analysis, we extract geometric features from the fitted ellipse, namely length and ratio of the major and minor axes, center location, orientation, area, and maximum offset from origin of the topological space. From persistent homology, lifetime of holes and independent connected components are obtained from barcodes. Table I lists all topological features in detail. Note that these features are extracted from the point cloud rather than from the raw signals, thus the raw data could be any length.

TABLE I
FEATURES FROM TOPOLOGICAL SIGNAL REPRESENTATION

Denotation	Description
shape-a	Length of the major axis of the ellipse.
shape-b	Length of the minor axis of the ellipse.
shape-x	Center location of the ellipse along the x axis.
shape-y	Center location of the ellipse along the y axis.
shape-ϕ	Orientation of the ellipse.
shape-A	Area of the ellipse.
shape-L	Maximum offset. The longest distance between any point on the ellipse and the origin of the topological space.
shape-R	Ratio of the major and minor axes of the ellipse.
homo-0	Accumulated lifetime of independent connected components. Integration of bar length in the barcode (dimension 0).
homo-1	Accumulated lifetime of holes. Integration of bar length in the barcode (dimension 1).

III. GAUSSIAN-BASED DECISION TREE

Topological signal representation provides more robust features, based on which a so-called Gaussian-based decision tree (GDT) is designed as the classifier. The GDT is motivated by the scenario that sometimes we cannot obtain enough samples to train a classifier. When the sample size is too small (e.g., only 2 samples for each class in our case), most conventional classifiers, such as SVM [41], decision tree, kNN, etc., would fail to handle this dataset. Besides better performance on small dataset, the GDT also learns faster than decision tree.

Decision tree is very effective because of its non-parametric and naturally non-linear properties, and it does not need to adjust many parameters or select kernels like neural network or SVM. In addition, decision tree potentially achieves better performance when the training dataset is small [42], while neural network (e.g., deep learning) requires a large training dataset. Conventional decision tree splits a node based on impurity or information gain, which are both entropy-based methods. It has to search the whole feature space for the optimal splitting boundary, and the computational complexity will increase when the features overlap among different

classes [42]. If decision tree is applied directly on a small dataset, its performance will degrade drastically.

Due to above shortcomings, we propose a Gaussian-based decision tree that mainly offers two improvements: 1) speed up node splitting by assuming Gaussian distribution on the features, and 2) better handling of small dataset by splitting nodes based on both class distance and entropy. Class distance refers to the Euclidean distance between the means of two classes in the feature space, and the entropy of a node given a splitting boundary x is expressed in Eq. 3.

$$E(x) = \begin{cases} -\frac{1}{N}\sum_{i=1}^{c}\sum_{l=1}^{2} N_l^i \ln \frac{N_l^i}{N^i} & , \ N_{l\in\{1,2\}} > 0 \\ \\ \infty & , \text{ otherwise} \end{cases} \tag{3}$$

where i denotes the class label, and there are N samples divided into c classes in total. l indicates the descendant node. There are two descendant nodes (left or right) since we are using a binary tree. The number of samples belonging to class i is N^i, and N_l^i denotes the number of class i samples that are split to the lth descendant node. N_l indicates the number of samples split to the lth descendant node. Given a feature that consists of tens or hundreds of observations, the N^i is obtained directly from corresponding ground truth, and the N_l^i is decided by the splitting boundary x. $E(x) = 0$ if all samples belonging to the same class are completely split into the same descendant node, although a descendant node may contain samples from several classes.

The Gaussian assumption is motivated by the fact that biosignals of the same class are supposed to vary within a finite interval. Specifically, a feature extracted from biosignals of the same class will locate within a fixed interval. If sample size is large enough, the feature will show a Gaussian distribution within the interval. Based on this assumption, the optimal splitting boundary can be obtained through Eq. 4. Assume there are $n \geq 2$ classes on certain node, and the feature $\mathbf{v}$ is selected, using which the means of each class are calculated and sorted in ascending order in $\boldsymbol{\mu} \in \mathbb{R}^n$, and the corresponding standard deviation is $\boldsymbol{\sigma} \in \mathbb{R}^n$. In the extreme case, there is only one sample in a class, whose variance results in zero that fails to approximate the true distribution. Because variance of classes is supposed to change smoothly, we replace the zero by a value interpolated on $\boldsymbol{\sigma}$. In other words, variance of a single-sample class is estimated from its neighbor classes in the feature space.

$$\begin{aligned} \mathcal{F}(\mathbf{v}) = \min_x & \left\{ \frac{G_{\mu_i,\sigma_i}(x)}{\mu_{i+1} - \mu_i} + \gamma E(x) \right\} \\ \text{s.t. } & \mu_i < x < \mu_{i+1}, \ i = 1, 2, \cdots, c-1 \\ & G_{\mu_i,\sigma_i}(x) = G_{\mu_{i+1},\sigma_{i+1}}(x) \\ & \mu_i \in \boldsymbol{\mu}, \sigma_i \in \boldsymbol{\sigma} \end{aligned} \tag{4}$$

where $G_{\mu_i,\sigma_i}(\cdot)$ denotes the Gaussian function whose mean and variance are μ_i and σ_i^2, respectively. γ is the penalty coefficient to adjust the effect of entropy that efficiently avoids empty nodes and ensures the most pure splitting. Iterating on each feature, the global optimal splitting boundary can be obtained from Eq. 5, where x_j denotes the optimal splitting boundary on the jth feature $\mathbf{v}_j$. Assume there are m features.

$$\arg\min_{x_j} \mathcal{F}(\mathbf{v}_j), \ j = 1, 2, \cdots, m \tag{5}$$

Actually, Eq. 5 minimizes the weighted sum of class distance and entropy on the intersections of Gaussian distributions. This dramatically reduces the number of points that need to search for optimal splitting.

IV. Experimental Results

This section implements the topological signal representation and classification (TA-GDT) on a small dataset, as well as a larger dataset. Section IV-A introduces the datasets. Then, demo results of TA and GDT on the small dataset are illustrated in sections IV-B and IV-C, respectively. Section IV-D compares the topological features to some common statistical and spectral features. Finally, the propose TA-GDT is validated on a much larger dataset to demonstrate its generality and robustness to sample size in section IV-E.

A. Dataset and Experiment Setup

To verify the effectiveness of our method, five solutions with different concentration of bovine IgG whole molecules are adopted as testing samples. Their concentrations are 0ng/mL, 1ng/mL, 5ng/mL, 10ng/mL and 100ng/mL. Using the same type of biosensor whose sampling period is 0.15 sec, the solutions are tested for 30 sec and yield biosignals with 300 sampling points. Considering that the voltage level of ACEK may significantly affect the performance of classification, we obtain biosignals respectively under 20mV, 40mV, 70mV and 135mV in order to test the voltage effect. For reasons discussed in section I-A, only 2 samples can be obtained for each concentration under certain voltage.

We also adopt an open-source ECG dataset, the Fantasia database [43], to illustrate the generality of the proposed method, as well as the robustness to data size. This database collected continuous electrocardiographic (ECG) signal for 120 minutes from two groups of healthy subjects – twenty young (21-34 years old) and twenty elderly (68-85 years old). There are large variation within each group. Because the ECG signal for each subject consists of 1 billion sampling points (lasting 2 hours), we evenly divide it into 100 segments with the length of 10000 points. Therefore, there are $20 \times 100 = 2000$ samples in each group (class).

B. Feature Extraction

This section demonstrate the feature extraction process on the solutions with different concentrations. First, DE transforms biosignals into topological space. The parameters of DE are set as follows: target dimension $d = 2$ and step size $s = 5$. As explained in section II-D, higher dimension cannot provide more representative feature, instead it slows down computation speed. Thus, we set $d = 2$ through empirical studies. Because of the random oscillation, it is difficult to find an s to make

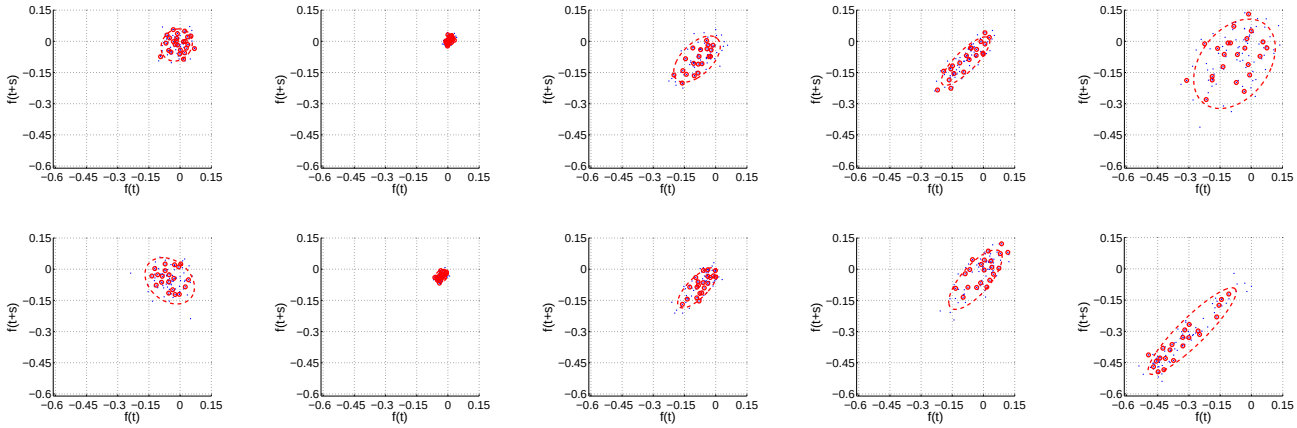


Fig. 11. Point cloud (blue dots), fitted ellipse (red dash curve) and landmarks (red circle) of the biosignals under 40mV. Each column displays two repeated detection on the same solution sample. From left to right, the concentrations are 0ng/mL, 1ng/mL, 5ng/mL, 10ng/mL, 100ng/mL.

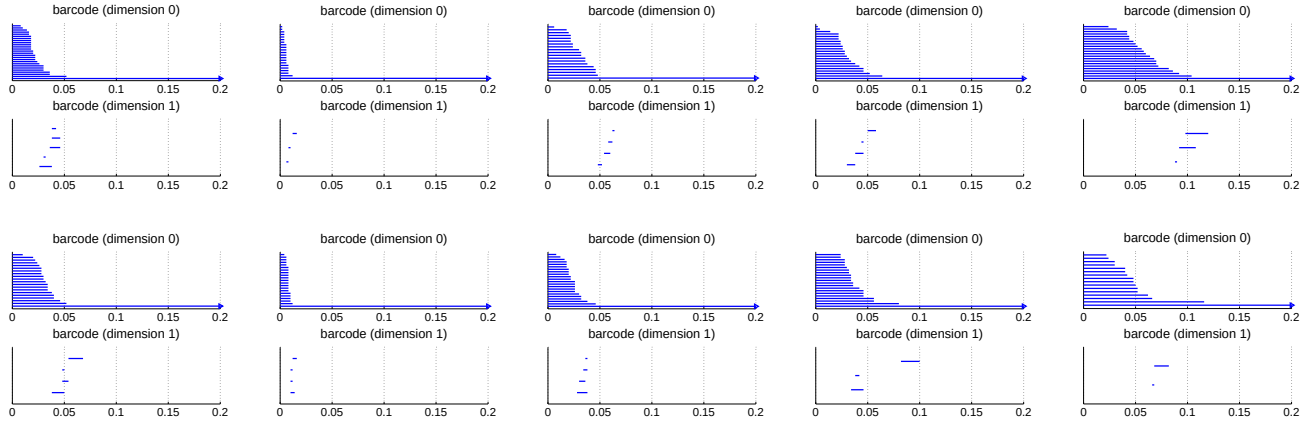


Fig. 12. The barcodes of dimensions 0 and 1 corresponding to the point cloud and landmarks in Fig. 11. The horizontal axis denotes the change of ϵ to construct the witness complex; the vertical axis stacks bars and has no specific physical meaning.

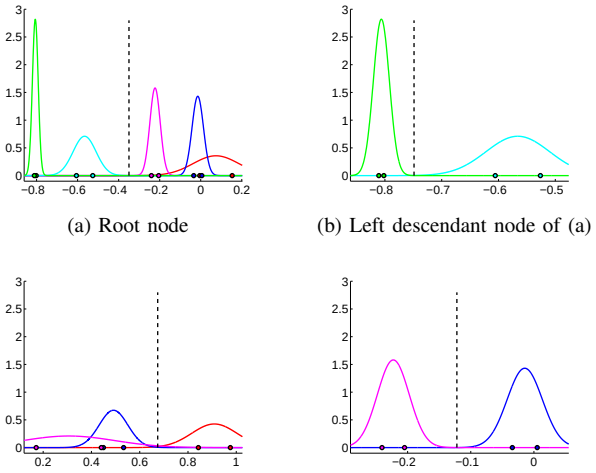
the point cloud form a large hole. To capture sufficient details, we set s to be a relatively small value, i.e., $s = 5$. Based on the five biosignals of different concentrations under 40mV in Fig. 1, their point cloud from DE and fitted ellipse are shown in Fig. 11 (the first row), and the second row displays the results of another five biosignals from repeated detections. Robustness of the shape feature can be intuitively realized — the shape of point cloud changes in almost the same trend with the increment of the concentration level. Shape analysis transforms overlapped raw signals into distinguishable feature space. The extracted shape features have been listed in Table I with a prefix "shape-".

In persistent homology, the barcodes are illustrate the persistent feature of a point cloud. Fig. 12 displays the barcodes of dimension 0 (independent connected components) and dimension 1 (holes) corresponding to the point cloud and landmarks in Fig. 11. The features extracted from persistent homology are listed in Table I with a prefix "homo-". Integration of the 0-dimension barcodes represents the area of the point cloud, and the length of bars (lifetime) generally reflects how the point cloud scatters. Longer bars indicate wider spread. Roughly, the holes under lower concentration have relatively shorter lifetime and earlier birth point.

C. Classification Using GDT

To better illustrate the construction procedure of GDT, ten samples from five concentrations under the highest voltage (135mV) are selected because the biosignals are more separable. Applying Eq. 5, the feature and corresponding splitting boundary are selected at the root node. The samples of different classes, as well as their Gaussian curves, are plotted with different colors in Fig. 13(a). The horizontal axis denotes values of the selected feature. The data on root node is split into left and right descendant nodes as shown in Fig. 13(b) and (c). By the same token, Eq. 5 is employed on each descendant node (Fig. 13(b) and (c)), and the splitting boundaries are plotted with dash lines. Fig. 13(d) shows the splitting boundary of the left descendant of Fig. 13(c).

Compared to conventional decision tree, GDT makes a strong assumption that all classes follow the Gaussian distribution and avoid splitting within a class. Therefore, it can efficiently decrease the depth of the tree and speed up the con-

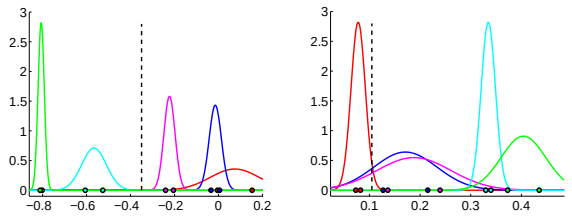


(a) Root node (b) Left descendant node of (a)

(c) Right descendant node of (a) (d) Left descendant node of (c)

Fig. 13. An example of constructing GDT. (a) is the root node. Dots in different colors denote the samples of different classes, and the corresponding curves represent their Gaussian distributions. The vertical dash line shows the splitting boundaries. (b) and (c) are descendant nodes of (a), and (d) is one of the descendant nodes of (c). Horizontal axis represents values of the selected feature, which is not necessary to be the same on each node.

struction procedure. Conventional decision tree uses entropy-based rule to search for the optimal feature and splitting boundary in a greedy manner. So it could draw a splitting boundary anywhere, which will definitely make the tree deeper and less computationally efficient. When playing with a small dataset, such as the dataset used in Fig. 13, the decision tree fails to find the optimal feature at the root node, let alone the later splittings at the descendant nodes. Fig. 14 compares GDT and decision tree on the root node using the same dataset. Obviously, GDT achieves a better splitting boundary on a more separable feature, on which different classes depart further from each other and the splitting boundary presents lower probability of misclassification.



(a) GDT (b) Decision tree

Fig. 14. The selected features and splitting boundaries of GDT and conventional decision tree on the root node using the same dataset in Fig. 13.

More comparison between GDT and some other common classifiers are displayed in Table II. The comparison is performed on the samples under the highest voltage (135mV) – the same dataset used in Fig. 13. To avoid over-fitting, the leave-one-out cross validation is applied for each classifier. GDT achieves a perfect 100% accuracy under 135mV. The accuracy decreases as the voltage decreases. The other classifiers, although using the best set of parameters, still cannot outperform GDT.

TABLE II
ACCURACY OF DIFFERENT CLASSIFIERS ON SMALL DATASET

Classifier	20mV	40mV	70mV	135mV
GDT	<70%	80%	90%	100%
decision tree	—	—	—	70%
random forest	—	—	—	60%
SVM	—	—	—	60%
kNN	—	—	—	30%
sparse coding	—	—	—	20%

D. Comparison to Statistical and Spectral Features

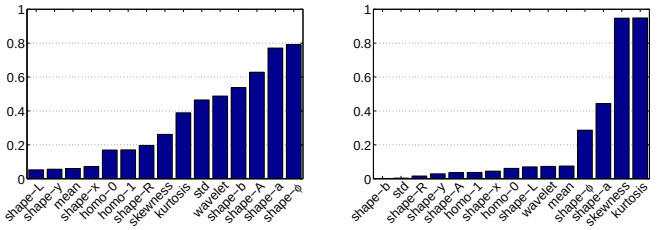
To demonstrate the advantage of topological features, we compare them to statistical and spectral features listed in Table III.

TABLE III
STATISTICAL AND SPECTRAL FEATURES INVOLVED IN COMPARISON

Denotation	**Description**
mean	Mean of a biosignal.
std	Standard deviation of a biosignal.
skewness	Asymmetry of the probability distribution of the values of all sampling points in a biosignal.
kurtosis	Peakedness of the probability distribution of the values of all sampling points in a biosignal [44].
wavelet	Average of the coefficients of the lowest frequency level from discrete wavelet transform (DWT), which captures not only frequency content like Fourier transform, but also temporal content [45].

In feature comparison, the correlation of a feature to the ground truth is concerned, which evaluates the separability of a feature. Higher separability implies a more reliable and predictable feature. In correlation estimation, each feature is associated with a P-value (between 0 and 1), which implies the confidence on the null hypothesis that the feature is independent to the ground truth. Thus, lower P-value indicates better feature which is more correlated to the ground truth. Fig. 15 shows the P-value of each feature in ascending order. To make the comparison persuasive, features under the two lowest voltages (20mV and 40mV) are adopted because biosignals tend to be more difficult to be classified as the voltage decreases. Comparing the two sub figures in Fig. 15, higher voltage makes features more separable (lower P-value) in general. Usually, a P-value less than 0.05 is considered as a strong evidence of high correlation (good feature). The spectral feature – wavelet – is always above this threshold because of the random oscillation. The statistical features perform better, but they still cannot compete with topological features.

At last, we need to investigate whether the topological features really promote the performance of classification. Table IV shows the accuracy of different classifiers using topological and statistical features, respectively. Here, the spectral feature (wavelet) is ignored because of its bad performance as analyzed above. The samples under 40mV and 135mV are adopted in comparison. As the voltage decreases, the accuracy of statistical features degrades dramatically. However, the topological features still preserve relatively high accuracy.



(a) 20mV (b) 40mV

Fig. 15. P-value of each feature in correlation estimation under (a) 20mV and (b) 40mV.

TABLE IV
ACCURACY OF DIFFERENT CLASSIFIERS USING DIFFERENT FEATURES

Classifier	Statistical features		Topological features	
	40mV	135mV	40mV	135mV
GDT	60%	90%	80%	100%
decision tree	30%	70%	80%	70%
random forest	30%	70%	40%	70%
SVM	50%	50%	50%	60%

E. Application on the ECG Dataset

We validate the proposed TA-GDT on the ECG dataset to demonstrate its generality and effectiveness in handling small dataset. Applying random leave-50%-out cross validation, Table V displays the classification accuracy of different algorithms on the ECG dataset, as well as the standard deviation obtained by iterating the validation 100 times. We did not compare to the state-of-the-art classifier – deep learning – because it requires much larger dataset in the training, which is not available for our scenario.

TABLE V
ACCURACY OF DIFFERENT ALGORITHMS ON THE ECG DATASET

	kNN	Random forest	SVM	TA-GDT
Accuracy (%)	72.7±.04	81.6±.05	78.1±.07	82.5±.05

To further validate the effectiveness of TA-GDT in handling data with small sample size, we shrink the ECG dataset incrementally from 100% (2000 samples per class) to 0.1% (2 samples per class), and randomly selected samples are removed from the training set at each shrinking step. The classification accuracy at each shrinking step is listed in Table VI which is visualized in Fig. 16. Compared to the other algorithms, the TA-GDT preserves relatively higher classification accuracy as the size of dataset shrinks. The accuracy of SVM and kNN decreases fast as the data size shrinks. The random forest performs competitively to the proposed method, however, TA-GDT still achieves the highest accuracies in all tests. The reason that TA-GDT does not show as big performance improvement as in the mastitis detection application is due to the less challenging problem at hand, where ECG signals possess periodic peaks which are easier distinguish as compared to the biosignal from mastitis detection.

TABLE VI
ACCURACY OF DIFFERENT ALGORITHMS ON ECG SUBSET

Percentage of the raw ECG dataset	kNN	Random forest	SVM	TA-GDT
0.1%	0.4000	0.6000	0.5000	0.6000
0.5%	0.5000	0.6400	0.5000	0.6527
1%	0.5050	0.6700	0.5000	0.6981
10%	0.6100	0.7600	0.5800	0.7821
20%	0.6150	0.7675	0.5775	0.7883
30%	0.6116	0.7708	0.5864	0.7901
40%	0.6275	0.7850	0.5888	0.7956
50%	0.6280	0.7992	0.6220	0.8002
60%	0.6367	0.7978	0.6633	0.8105
70%	0.6414	0.7987	0.6777	0.8191
80%	0.6631	0.8110	0.6944	0.8201
90%	0.7167	0.8150	0.7500	0.8243
100%	0.7270	0.8160	0.7810	0.8250

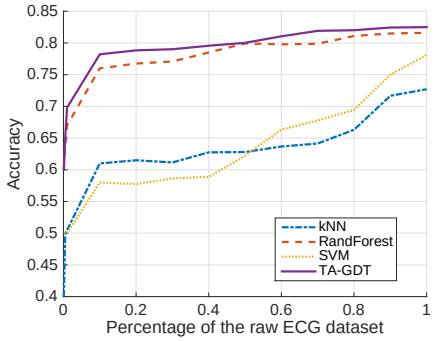


Fig. 16. Accuracy of different algorithms on ECG subset when reducing the size of training set.

V. WEAKNESS OF THE PROPOSED METHOD

Although effective in handling challenging biosignal analysis problems, the limitation of TA-GDT is three-fold:

- The signals only different in phase cannot be distinguished because DE is robust to phase shift.
- The high computational complexity of persistent homology forces us to restrict the number of landmarks to a small value (e.g., 50 or less), which may cause loosing the intrinsic structure of the raw point cloud when the point cloud is large and widespread.
- It is worth noting that GDT is relatively more effective when handling smaller sample sizes. For larger datasets, GDT may not be able to fully characterize the probability distribution of the data unless if the data distribution is indeed Gaussian.

VI. CONCLUSION

This paper presented an innovative representation and classification framework for biosignal analysis with very small sample size, referred to as TA-GDT. In the topological space, robust and representative features were extracted from the biosignals which appear indistinguishable in the time domain. The features include both shape characteristics of the point cloud generated from delay embedding and the intrinsic structure of the point cloud revealed through persistent homology. To efficiently classify the biosignals with small sample size,

we also proposed the Gaussian-based decision tree (GDT), which outperformed most existing classifiers.

We applied TA-GDT on a challenging problem – mastitis detection where biosignals with five different concentration levels are separated at high accuracy although the sample size is extremely small. The TA-GDT has the potential to analyze other biosignals that are not easy to be distinguished in the time or spectral domain, due to its effectiveness in handling challenging issues like significantly random oscillation, small inter-class variation, large intra-class variation, and extremely small sample size. The proposed TA-GDT has shown itself to be an effective solution to time-series biosignal analysis.

REFERENCES

[1] B. Harmon, "Somatic cell counts: a primer," in *Annual Meeting–National Mastitis Council Incorporated*, vol. 40. National Mastitis Council; 1999, 2001, pp. 3–9.

[2] H. Jiang, X. Weng, and D. Li, "Microfluidic whole-blood immunoassays," *Microfluidics and Nanofluidics*, vol. 10, no. 5, pp. 941–964, 2011.

[3] M.-I. Mohammed and M. P. Desmulliez, "Lab-on-a-chip based immunosensor principles and technologies for the detection of cardiac biomarkers: a review," *Lab on A Chip*, vol. 11, no. 4, pp. 569–595, 2011.

[4] X. Pei, B. Zhang, J. Tang, B. Liu, W. Lai, and D. Tang, "Sandwich-type immunosensors and immunoassays exploiting nanostructure labels: A review," *Analytica Chimica Acta*, vol. 758, pp. 1–18, 2013.

[5] H. Morgan and N. G. Green, *AC electrokinetics: colloids and nanoparticles*. Research Studies Press, 2003, no. 2.

[6] M. L. Sin, T. Liu, J. D. Pyne, V. Gau, J. C. Liao, and P. K. Wong, "In situ electrokinetic enhancement for self-assembled-monolayer-based electrochemical biosensing," *Analytical Chemistry*, vol. 84, no. 6, pp. 2702–2707, 2012.

[7] J. Wu, "Biased AC electro-osmosis for on-chip bioparticle processing," *IEEE Transactions on Nanotechnology*, vol. 5, no. 2, pp. 84–89, 2006.

[8] K. Yang and J. Wu, "Numerical study of in situ preconcentration for rapid and sensitive nanoparticle detection," *Biomicrofluidics*, vol. 4, no. 3, p. 034106, 2010.

[9] S. Li, H. Cui, Q. Yuan, J. Wu, A. Wadhwa, S. Eda, and H. Jiang, "AC electrokinetics-enhanced capacitive immunosensor for point-of-care serodiagnosis of infectious diseases," *Biosensors and Bioelectronics*, vol. 51, pp. 437–443, 2014.

[10] S. Li, Y. Ren, H. Cui, Q. Yuan, J. Wu, S. Eda, and H. Jiang, "Alternating current electrokinetics enhanced in situ capacitive immunoassay," *Electrophoresis*, vol. 36, no. 3, pp. 471–474, 2015.

[11] H. Cui, C. Cheng, X. Lin, J. Wu, J. Chen, S. Eda, and Q. Yuan, "Rapid and sensitive detection of small biomolecule by capacitive sensing and low field AC electrothermal effect," *Sensors and Actuators B: Chemical*, vol. 226, pp. 245–253, 2016.

[12] R. Guo, S. Li, L. He, W. Gao, H. Qi, and G. Owens, "Pervasive and unobtrusive emotion sensing for human mental health," in *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, 2013, pp. 436–439.

[13] R. Archibald, A. Gelb, and J. Yoon, "Polynomial fitting for edge detection in irregularly sampled signals and images," *SIAM Journal on Numerical Analysis*, vol. 43, no. 1, pp. 259–279, 2005.

[14] D. W. Hosmer Jr and S. Lemeshow, *Applied logistic regression*. John Wiley & Sons, 2004.

[15] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[16] Y. Song, W. Wang, Z. Zhang, H. Qi, and Y. Liu, "Multiple event analysis for large-scale power systems through cluster-based sparse coding," in *IEEE International Conference on Smart Grid Communications (SmartGridComm): Cyber Security and Privacy (IEEE SmartGridComm'15 Symposium - Security and Privacy)*, Miami, USA, Nov. 2015.

[17] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. Wells Jr, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

[18] H. D. Abarbanel, T. Carroll, L. Pecora, J. Sidorowich, and L. S. Tsimring, "Predicting physical variables in time-delay embedding," *Physical Review E*, vol. 49, no. 3, p. 1840, 1994.

[19] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.

[20] M. Richter and T. Schreiber, "Phase space embedding of electrocardiograms," *Physical Review E*, vol. 58, no. 5, p. 6392, 1998.

[21] D. Sciamarella and G. Mindlin, "Topological structure of chaotic flows from human speech data," *Physical Review Letters*, vol. 82, no. 7, p. 1450, 1999.

[22] S. Emrani, T. Gentimis, and H. Krim, "Persistent homology of delay embeddings and its application to wheeze detection," *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 459–463, 2014.

[23] S. Emrani, H. Chintakunta, and H. Krim, "Real time detection of harmonic structure: A case for topological signal analysis," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3445–3449.

[24] S. Emrani and H. Krim, "Robust detection of periodic patterns in gene expression microarray data using topological signal analysis," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 1406–1409.

[25] M. R. Muldoon, D. S. Broomhead, J. P. Huke, and R. Hegger, "Delay embedding in the presence of dynamical noise," *Dynamics and Stability of Systems*, vol. 13, no. 2, pp. 175–186, 1998.

[26] J. A. Perea and J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, pp. 1–40, 2013.

[27] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, 1999.

[28] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

[29] H. Edelsbrunner and J. Harer, *Computational topology: an introduction*. American Mathematical Soc., 2010.

[30] L. Tang and M. Crovella, "Geometric exploration of the landmark selection problem," in *Passive and Active Network Measurement*. Springer, 2004, pp. 63–72.

[31] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian, "On the local behavior of spaces of natural images," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 1–12, 2008.

[32] H. Adams and G. Carlsson, "On the nonlinear statistics of range image patches," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 110–117, 2009.

[33] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[34] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.

[35] V. De Silva and G. Carlsson, "Topological estimation using witness complexes," in *Proceedings of the First Eurographics Conference on Point-Based Graphics*. Eurographics Association, 2004, pp. 157–166.

[36] L. J. Guibas and S. Y. Oudot, "Reconstruction using witness complexes," *Discrete & Computational Geometry*, vol. 40, no. 3, pp. 325–356, 2008.

[37] H. Edelsbrunner, "Persistent homology: theory and practice," 2014.

[38] X. Zhu, "Persistent homology: An introduction and a new text representation for natural language processing," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*. AAAI Press, 2013, pp. 1953–1959.

[39] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," *Discrete and Computational Geometry*, vol. 28, no. 4, pp. 511–533, 2002.

[40] D. Freedman and C. Chen, "Algebraic topology for computer vision," *Computer Vision*, pp. 239–268, 2009.

[41] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass svm method? an empirical study," in *Multiple Classifier Systems*. Springer, 2005, pp. 278–285.

[42] H. Deng, G. Runger, and E. Tuv, "Bias of importance measures for multi-valued attributes and solutions," in *Artificial Neural Networks and Machine Learning–ICANN*. Springer, 2011, pp. 293–300.

[43] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[44] T.-H. Kim and H. White, "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, vol. 1, no. 1, pp. 56–73, 2004.

[45] M. Shensa, "The discrete wavelet transform: wedding the a trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, 1992.