

# Classification of Hand Movements Based on EMG Signals using Topological Features

Jiayang Li<sup>1</sup>, Lei Yang<sup>2</sup>, Yunan He<sup>3\*</sup>, Osamu Fukuda<sup>4</sup>

Mathematical Science Research Center, Chongqing University of Technology, Chongqing 400054, China<sup>1, 2, 3</sup>  
Graduate School of Science and Engineering, Saga University, Saga 840-8502, Japan<sup>4</sup>

**Abstract**—Hand movement classification based on Electromyography (EMG) signals has been extensively investigated in the past decades as a promising approach used for controlling upper prosthetics or robotics. Topological data analysis is a relatively new and increasingly popular tool in data science that uses mathematical techniques from topology to analyze and understand complex data sets. This paper proposes a method for classifying hand movements based on EMG signals using topological features crafted with the tools of TDA. The main findings of this work on hand movement EMG classification are as follows: (1) topological features are effective in classifying EMG signals and outperform other time domain features tested in the experiments; (2) the 0-th Betti numbers are more effective than the 1-st Betti numbers; (3) Betti amplitude is a more stable and powerful feature than other topological features discussed in this paper. Additionally, Betti curves were used to visualize topological patterns for hand movement EMG.

**Keywords**—EMG classification; persistent homology; topological features; betti curve

## I. INTRODUCTION

Electromyography (EMG) is a bioelectrical signal generated by muscle cells, providing valuable information about muscle activity that can be used to recognize hand movements [1]. Over the past few decades, various techniques have been developed to discriminate hand movements from EMG signals [2], [3], [4], [5], and most follow a unified analysis pipeline that involves preprocessing, feature extraction, and classification. Feature extraction involves transforming raw data into a feature vector that is later fed into a classifier. The dimension of the feature vector is usually much smaller than the raw data, which helps to reduce redundant data and prevent learning from the curse of dimensionality [6], thus accelerating learning speed and generalization steps. Therefore, choosing the right features has a significant impact on classification performance.

Previous studies have evaluated the ability of various EMG features to recognize hand movements, which can be categorized into time and frequency domains. Time-domain features analyze EMG signals over time, and common ones include mean absolute value (MAV) and root mean square (RMS) [7]. Frequency-domain analysis involves measurements that describe specific aspects of the signal's frequency spectrum, with mean frequency and median frequency being two useful features [8]. In addition, there are also end-to-end learning models that try to learn a feature representation of raw signals instead of using hand-crafted features [9], [10].

These models map the initial input to the final outputs directly, making it difficult to classify features into a particular domain.

In this work, we introduce a new type of feature for hand movement classification called topological features. The primary method used in this study to extract topological features falls in the field of topological data analysis (TDA). TDA is an approach that extracts topological features and describes the geometric shapes of datasets using techniques from topology [11], [12], [13]. Topological invariants are flexible and independent of metrics and coordinates, which allows us to compare EMG data collected from various wearing manners of EMG sensors. Additionally, topological invariants characterize the overall features of datasets, making topology-based solutions less sensitive to noise and enabling the identification of EMG signal shapes despite countless deformations [14]. For a point set in a Euclidean space, we can obtain a filtration of simplicial complexes as the spatial scale changes. The filtration of simplicial complexes provides a multi-scale perspective to understand data. The most common constructions of the filtration of simplicial complexes are the Vietoris-Rips complex and the Čech complex [15], [16]. Persistent homology is proposed to measure the topological features of data that persist across all scales. Since topological features are detected over a wide range of spatial scales, it is more likely to find true patterns of shape behind the data rather than noise.

TDA has found applications across diverse fields, such as material science [17], [18], biomolecules [19], [20], oncology [21], sensor networks [22], and data science [23]. Numerous TDA-based methods have been proposed for time series analysis [24], [25], [26]. For example, Pereira and de Mello developed a time series clustering approach using topological features computed by persistent homology [27]. Khasawneh and Munch tracked the stability of stochastic dynamical systems with TDA [28]. Gidea and Katz detected market crashes with financial time series analysis using topological features [29]. Emrani et al. applied TDA for wheeze detection in breathing sounds [30]. TDA has also been employed in ECG signal analysis, with Ignacio et al. identifying Atrial Fibrillation using topological features [31] and Dlugas detecting arrhythmias with topological methods [32].

This study presents a novel and effective method for visualizing and recognizing hand movement EMG signals and provides guidance for selecting hyper-parameters. The main innovations of this work are as follows. First, our approach does not rely on a specific embedding dimension and delay to transform time series signals into point clouds for analysis. Instead, we explored various combinations of embedding

\*Corresponding authors. email ID: heyunan@cqut.edu.cn

dimensions and delays to determine the optimal values and understand the relationship between classification accuracy and the embedding parameters. Secondly, a wide range of topological features (Wasserstein amplitude, Betti amplitude, landscape amplitude, and persistent entropy) were considered from the perspective of entropy and amplitude to classify EMG signals. The topological features derived from different dimensions of homology groups were considered separately and jointly. Finally, we pioneered the use of Betti curves to visualize the topological structure of EMG signals. Betti amplitude can be used as a metric to evaluate the difficulty in distinguishing two hand movements using EMG signals.

To the best of our knowledge, this is the first work that employs topological features for classifying hand movements based on EMG signals. Our experimental results show that the proposed topological features achieve higher accuracy than other time-domain features for EMG signal classification. The main contributions of this study are: (1) The information conveyed by the 0-th Betti numbers is found to be more effective than that conveyed by the 1-st Betti numbers in classifying EMG signals. Combining the 0-th and 1-st Betti numbers does not result in higher classification accuracy, but instead yields a lower accuracy than using only the 0-th Betti numbers. (2) Among the four topological features explored, Betti amplitude, which is the  $L_2$ -metric between Betti curves, is stable and effective in classifying EMG signals, with its accuracy almost unaffected by the embedding dimension and embedding delay. (3) Inspired by the effectiveness of the 0-th Betti numbers in classifying EMG signals, we visualized EMG signals of 53 types of hand movements using 0-th Betti curves and clearly observed differences in the topological structure.

The rest of this paper is organized as follows: Section II introduces the general approach for classifying EMG signals using topological features, along with related concepts and tools. Section III delves into the experimental aspects, addressing topics such as hyperparameter selection, topological feature selection, and the effectiveness of topological features. Finally, Section IV summarizes the study and highlights potential directions for future research.

## II. METHODS

This section will begin with a review of the fundamental concepts and methods of TDA that are relevant to this work. It will cover the concepts of simplex, simplicial complex, and persistent homology. Additionally, it will introduce the notion of a persistence diagram, which is one type of representation used in persistent homology. The latter part of this section will focus on the pipeline used to classify EMG signals. The pipeline involves transforming the time-series EMG signal into a metric space, followed by converting the metric space into a topological space. From this topological space, topological features are extracted and fed into classifiers.

### A. Homology and Persistent Homology

The triangle is known to be one of the simplest geometric shapes in the plane, and we can combine triangles into a more complex shape. The simplex can be viewed as a generalization of the notion of the triangle in any dimension. Specifically, a  $p$ -simplex is a  $p$ -dimensional polytope which is the convex

hull of its  $p + 1$  geometrically independent vertices in the Euclidean space  $\mathbb{R}^n$ . For example, a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex means a tetrahedron, etc. A simplicial complex  $K$  in the Euclidean space is a collection of simplices such that (1) each face of a simplex of  $K$  is a simplex; (2) the intersection of any two simplices of  $K$  is either empty or a common face of them. Simplicial complexes provide discrete representations for topological spaces.

Homology is one of the essential topological invariants for describing the intrinsic properties of spaces. In this work, we focus on the simplicial homology on simplicial complexes. Let  $K$  be a simplicial complex. Let  $\mathbb{F}$  be a field. Denote  $C_p(K; \mathbb{F})$  the  $\mathbb{F}$ -linear space generated by the  $p$ -simplices of  $K$ . Then  $C_*(K; \mathbb{F})$  is a chain complex with the boundary operator  $\partial_p : C_p(K; \mathbb{F}) \rightarrow C_{p-1}(K; \mathbb{F})$  given by

$$\partial_p[v_0, v_1, \dots, v_p] = \sum_{i=0}^p (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_p], \quad p \geq 1$$

for any simplex  $[v_0, v_1, \dots, v_p]$  of  $K$ , where  $\hat{v}_i$  means omission of the term  $v_i$ . For  $p = 0$ , we denote  $\partial_0 = 0$ . Then the  $p$ -th homology group of  $K$  is defined by

$$H_p(K; \mathbb{F}) := \frac{\ker \partial_p}{\text{Im } \partial_{p+1}}, \quad p \geq 0.$$

The homology groups reflect the topological features of simplicial complexes. The 0-dimensional homology group detects the connected components of simplicial complexes, the 1-dimensional homology group detects the loops while higher dimensional homology groups detect higher dimensional voids or cavities. In essence, homology detects “holes” in a simplicial complex. Betti number, defined by  $\beta_p = \dim H_p(K; \mathbb{F})$  is the usual topological invariant to describe the information of “holes”. The advantage of using Betti number to represent the topology pattern of data is that it is more intrinsic and more resistant to noise.

Persistent homology is the central method to detect the topological features and describe the geometric shapes of high-dimensional data in topological data analysis. The persistence is intended to focus on the multi-scale information of data sets. We build the persistent homology on data sets by a filtration of simplicial complexes. Given a collection of points in Euclidean space  $\mathbb{R}^n$ , the Vietoris-Rips complex  $\mathcal{R}_\epsilon$  is the abstract simplicial complex whose  $p$ -simplices are the sets of  $p + 1$  points which are pairwise within distance  $\epsilon$  [15]. The Vietoris-Rips complex is the most frequently used filtration of complexes constructed from a point set. Let  $K$  be a simplicial complex equipped with a real-valued function  $f : K \rightarrow \mathbb{R}$ . Then we have a filtration of simplicial complex  $\{K_\epsilon\}_\epsilon$  given by  $K_\epsilon = \{\sigma \in K | f(\sigma) \leq \epsilon\}$ . The  $(a, b)$ -persistent homology of  $K$  with respect to  $f$  is defined by

$$H_p^{a,b}(K; \mathbb{F}) := \text{Im}(H_p^a(K; \mathbb{F}) \rightarrow H_p^b(K; \mathbb{F})), \quad p \geq 0.$$

The  $(a, b)$ -persistent Betti number is given by  $\beta_p^{a,b} = \dim H_p^{a,b}(K; \mathbb{F})$ . There are two typical representations of persistent Betti numbers, the barcode, and the persistence diagram. The barcode and the persistence provide the visualization of the persistent homology. See Fig. 1 as an example. Consider a collection of points in a Euclidean plane  $\mathbb{R}^n$ . We

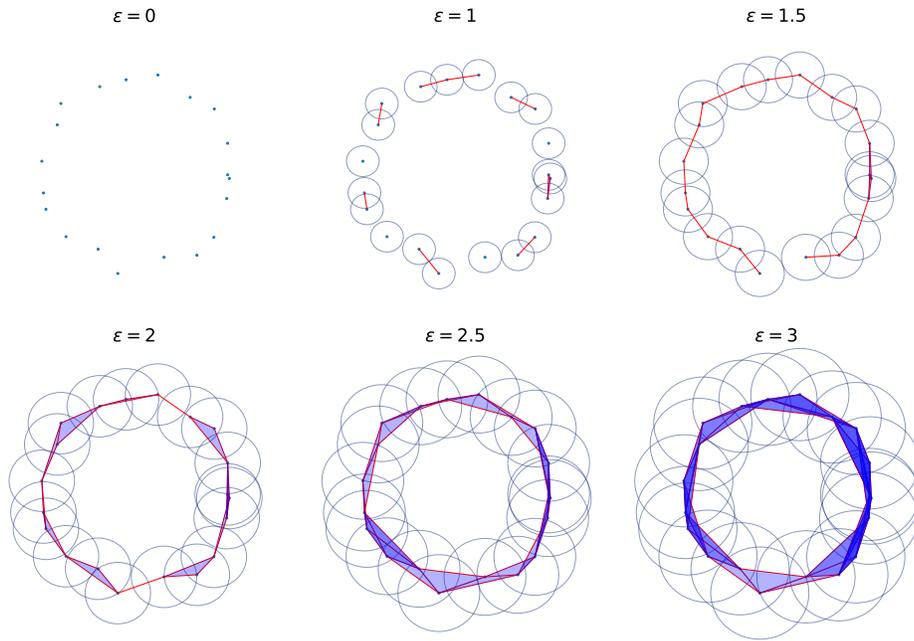


Fig. 1. A nested sequence of simplicial complexes as the radius increases.

obtain a filtration of simplicial complexes as the parameter  $\epsilon$  grows. When  $\epsilon = 0$ , the Vietoris-Rips complex  $\mathcal{R}_0$  is the simplicial complex of discrete points. By setting the sample points as the circle center and setting the scale parameter  $\epsilon$  as the radius of a circle, the filtration of simplicial complexes can be constructed as follows: if two circles have an intersection, then add a line segment between the two points ( $\epsilon = 1$ ). If three circles have intersections with each other, then add a triangle among the three points ( $\epsilon = 2$ ). Thus simplices are added to the complex step by step, which generates a nested sequence of simplicial complexes as the radius increases. As the scale parameter grows, some generators come into existence at some parameters while some generators disappear at some parameters. The parameter that a generator  $\alpha$  appear is called the birth time of  $\alpha$ , and the parameter that  $\alpha$  disappears is called the death time. The (birth, death) pairs are plotted in a diagram called persistence diagram. Persistence diagram provides a concise description of the topological changes over all scale parameters.

### B. Topological Features

Although persistence diagrams are useful descriptors of data, they may not be suitable as input data for most machine learning models due to the absence of a natural linear structure. However, this limitation has been addressed through vectorizing persistence diagrams or using kernel methods. In this study, we extract four types of topological features by vectorizing persistence diagrams. These features have been successfully employed in other work for time series analysis. The four features can be categorized into two groups: entropy and amplitude. Entropy refers to persistent entropy, while amplitude includes Wasserstein amplitude, landscape amplitude, and Betti amplitude. The followings are the details of the topological features.

- (a) *Persistent entropy*: Persistent entropy is defined as the Shannon entropy of the persistence diagram. In general, the lower the persistent entropy is, the simpler the shape of the data will be. Let  $D$  be a persistence diagram and  $\alpha = (b_\alpha, d_\alpha)$  be a point in the diagram. Then, the persistent entropy of  $D$  is defined as:

$$E(D) := - \sum_{\alpha \in D} p_\alpha \log p_\alpha \quad (1)$$

where

$$p_\alpha = \frac{d_\alpha - b_\alpha}{\sum_{\alpha \in D} (d_\alpha - b_\alpha)} \quad (2)$$

- (b) *Amplitude*: The main idea behind amplitude is to partition a diagram into sub-diagrams based on homology dimension, and use a metric to measure the distance of each sub-diagram (or the derivative of each sub-diagram) with respect to the diagonal diagram. The diagonal diagram consists of only the diagonal line. Wasserstein amplitude computes the  $p$ -Wasserstein distance between sub-diagrams and the diagonal diagram, and it is defined as:

$$A_w = \frac{\sqrt{2}}{2} \sum_{\alpha \in D} ((d_\alpha - b_\alpha)^p)^{\frac{1}{p}} \quad (3)$$

Landscape amplitude computes the  $L_p$  distance between persistence landscapes derived from sub-diagrams and the persistence landscape derived from the diagonal diagram. Betti amplitude, on the other hand, computes the  $L_p$  distance between Betti curves derived from sub-diagrams and Betti curve derived from the diagonal diagram. Persistence landscapes and Betti curves are two other representations of topological signatures. In this study, we only consider the case of  $p = 2$ .

### C. Time Delay Embedding

To compute persistent homology and extract topological features from the 1-dimensional time series EMG signals, they must first be represented in the form of a point cloud. This can be achieved using the method of time delay embedding, also known as Taken's embedding. Given a 1-dimensional time series  $X(t)$ , its time delay embedding can be described as a sequence of vectors in the following form:

$$X_i = [X(t_i), X(t_i + \tau), \dots, X(t_i + (d-1)\tau)] \in \mathbb{R}^d \quad (4)$$

where  $d$  is the embedding dimension and  $\tau$  is the time delay. Each vector is treated as a point in a  $d$ -dimensional Euclidean space, and all vectors together constitute a point cloud. There are two main parameters that need to be set in time delay embedding: embedding dimension and time delay. The embedding dimension determines the dimension of the Euclidean space. A higher dimension means more information is embedded in one single point, but it is harder to find patterns from these points in a higher space. The embedding delay determines how long a term of memory is embedded in one single point. A longer term of memory means longer memory, but the resulting data points lose the short-term memory and local features. The selection of embedding dimension and delay determines the topology of the embedded point clouds and thus affects the classification results.

In most cases, EMG signals are collected from multiple electrodes, making them multivariate time series. To apply time delay embedding on multivariate signals, it is applied to each channel separately, and each channel of an EMG sequence corresponds to a point cloud. In our proposed method, we use the same embedding dimension and time delay for all channels. Considering a window of an  $n$ -channel EMG signal with the shape of  $(m \times n)$ , where  $m$  is the length of the window, we obtain  $n$  point clouds.

## III. EXPERIMENTS

### A. Dataset and Preprocessing Policy

The EMG dataset employed in our experiments is a publicly accessible dataset called NinaPro DB5 [33], which records muscle activity using two Thalmic Myo armbands. Each Myo armband is equipped with 8 electrodes, yielding a total of 16 channels of EMG signals collected. The dataset comprises 6 repetitions of 53 distinct movements (including rest) performed by 10 intact subjects. The Thalmic Myo already incorporates a 50 Hz notch filter, eliminating the need for additional filtering [34].

Before feeding the EMG signals into classification models, data preprocessing is necessary. To compare the results of our proposed method with other benchmarks, we must maintain consistency in data preprocessing. Thus, we followed the exact procedure outlined in [34]. This process involves dividing each detected repetition into 200-sample windows with an overlap of 100 samples. Subsequently, each window is labeled with its corresponding movement number. For training and testing dataset splitting, repetitions 1, 3, 4, and 6 were used for training, while repetitions 2 and 5 were designated for validation. Classification was performed on all 53 movements (including

rest). It is important to note that the rest movement's sample size is significantly larger than that of other movements, so it was reduced to avoid dataset imbalance. Python served as the programming language for conducting the experiment. We employed various Python packages, including Risper [35] for computing persistent homology and scikit-learn for building classifiers.

### B. Selecting the Best Embedding Delay and Dimension

As outlined in Section II-C, we employed time delay embedding to transform time-series EMG signals into point clouds in Euclidean space. Time delay embedding involves two primary parameters: embedding dimension and delay. The selection of these parameters defines the topology of the embedded point clouds, which in turn influences the classification accuracy. To determine the optimal values for embedding dimension and delay in EMG classification, we generated point clouds for each channel by applying the same dimension and delay parameters, ranging from 2 to 10 and 1 to 9, respectively.

Subsequently, we calculated persistent homology on these point clouds and extracted persistence diagrams. The 0-th and 1-st Betti numbers were utilized to detect the number of connected components and the number of independent loops of simplicial complexes. This approach was chosen as computing higher-dimensional Betti numbers can be highly complex. Persistent entropy was then extracted as a feature descriptor from the persistence diagram. For a single channel, persistent entropy is a 2-dimensional vector in a plane, with coordinates corresponding to the 0 and 1-dimensional homology groups. As illustrated in Fig. 2, the two hypotheses being compared are labeled as  $H_0$  and  $H_1$ . The dataset used in this experiment contained recordings from 16 channels. When persistent entropy was computed for each channel, both  $H_0$  and  $H_1$  were produced, each 16-dimensional. Concatenating  $H_0$  and  $H_1$  resulted in a single 32-dimensional feature vector.

The EMG data was classified using persistent entropy of  $H_0$ ,  $H_1$ , and their concatenation. Each feature was input into a random forest classifier separately. The classifiers' performance was then evaluated to identify which feature yielded the best results. Fig. 2 displays the classification outcomes. The heatmap reveals the following findings: (1) The persistent entropy of  $H_1$  does not contribute any valuable information for EMG signal classification and exhibits poor performance. (2) Concatenating  $H_0$  and  $H_1$  does not significantly enhance classification accuracy. The highest classification accuracy achieved in our experiment was 71.86%, obtained with an embedding delay of 1 and a dimension of 6.

The optimal combination of embedding delay and dimension discovered in our previous experiment using persistent entropy and the random forest classifier may not be entirely convincing. Moreover, while the highest accuracy was attained at dimension 6, it is unclear whether this dimension is significantly better than dimensions 5 or 4, as the classification accuracies are only marginally different. To gain further insight into the effect of embedding delay and dimension on classification performance, we conducted a second experiment. Similar to the first experiment, we transformed the raw time-series data into point clouds for each channel using dimension

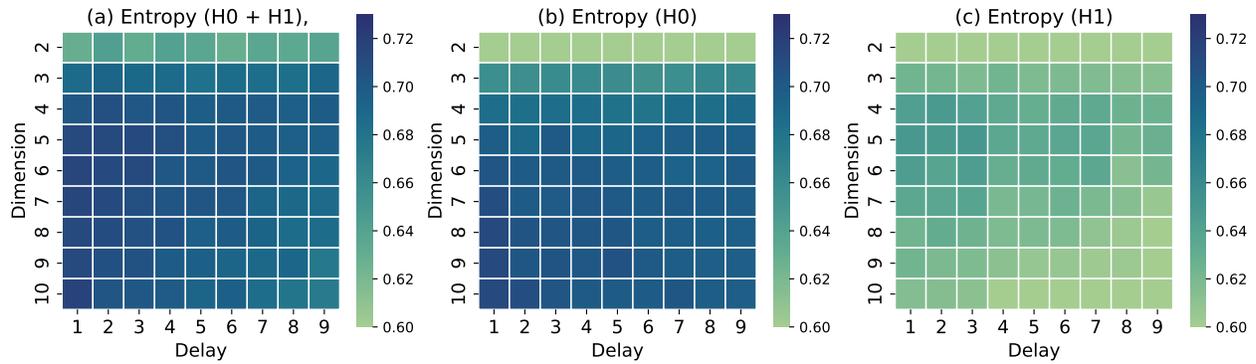


Fig. 2. Accuracy heatmap for various combinations of embedding delay and dimension.

and delay parameters selected from the range of 2 to 10 and 1 to 9, respectively, and computed persistent homology on these point clouds. We then extracted four types of topological features, namely persistent entropy, Wasserstein amplitude, landscape amplitude, and Betti amplitude. For each type of feature, we considered both  $H_0$ ,  $H_1$ , and their concatenation. These features were then fed into both SVM and random forest classifiers for further analysis.

To explore the impact of embedding dimension on classification accuracy, we first calculated the mean accuracy for each embedding dimension across all delays and plotted the mean accuracy curve in Fig. 3. In the legend of the first subfigure, “entropy\_h0\_rf” denotes using persistent entropy of  $H_0$  with a random forest classifier, while “entropy\_rf” denotes using the concatenation of persistent entropy of  $H_0$  and  $H_1$  with a random forest classifier.

The figure reveals that: (1) The topological features of  $H_1$  provide no meaningful information for EMG signal classification and perform poorly, consistent with the conclusion of the first experiment. The classification accuracy using the topological features of  $H_0$  does not show a decreasing trend with any classifier, and even continues to increase beyond a dimension of 10 in the case of persistent entropy and Wasserstein amplitude. (2) Concatenating topological features of  $H_0$  and  $H_1$  generally does not result in a positive effect. The accuracy curve initially exhibits a rising trend, reaches a peak, and then declines. This occurs because  $H_1$  carries less information and negatively impacts the concatenation of  $H_0$  and  $H_1$ . (3) Classification using Betti amplitude of  $H_0$  proves effective and stable across all dimensions, while landscape amplitude is not a useful feature, as the landscape amplitude of  $H_0$ ,  $H_1$ , and their concatenation did not yield acceptable accuracy.

To examine the effect of embedding delay on classification accuracy, we calculated the mean accuracy for each delay across all dimensions and plotted the results in Fig. 4. Our analysis reveals that: (1) For EMG signals, the classification accuracy using any topological feature (except landscape amplitude, which is not effective) of  $H_0$  remains stable across all delays, indicating that embedding delay has only a minor influence on EMG signal classification. (2) When using the concatenation of topological features of  $H_0$  and  $H_1$ , accuracy decreases as delays increase. This can be attributed to the limited usefulness of the information carried by  $H_1$ , which

can negatively impact  $H_0$ . However, classification using Betti amplitude of  $H_0$  remains stable across all delays.

Fig. 3 shows that the accuracy using topological features of  $H_0$  has a trend of increasing even beyond a dimension of 10. To further investigate this trend, we examined dimensions ranging from 2 to 20, with an embedding delay of 1, as delay has only a minor influence on classification results. The classification results, shown in Fig. 5, reveal that: (1) Betti amplitude remains stable even at higher dimensions, without any significant decrease in accuracy. (2) For persistent entropy and Wasserstein amplitude, the increasing trend becomes less pronounced after dimension 10, and their accuracy converges to a maximum value around dimension 12. (3) The highest classification accuracy achieved is 73.93%, using Betti amplitude of  $H_0$  at an embedding dimension of 3 and an embedding delay of 1 with SVM.

In summary, choosing the appropriate delay and dimension in time delay embedding and using Betti amplitude of  $H_0$  as a topological feature can significantly enhance the accuracy of EMG signal classification. The main findings of the experiments are as follows:

- (a) As the embedding dimension increases, the accuracy improves, eventually converging to a specific value. Delay has only minor effects on movement classification accuracy, so it can be simply set to 1. An exception is the case of Betti amplitude, where the classification model using Betti amplitude is robust to changes in dimension and delay. The optimal choices for delay and dimension for Betti amplitude are 1 and 3, respectively.
- (b)  $H_0$  contains more meaningful information than  $H_1$  for EMG hand movement classification, regardless of the topological feature used. Combining the topological features of  $H_0$  and  $H_1$  has negative effects on  $H_0$  and does not enhance accuracy.
- (c) Betti amplitude of  $H_0$  is the most effective and stable topological feature for classifying EMG signals, as it is robust to changes in dimension and delay and achieves the highest accuracy. Persistent entropy and Wasserstein amplitude are also good choices but require appropriate selection of embedding dimension and delay. Landscape amplitude is not an effective feature for EMG classification.

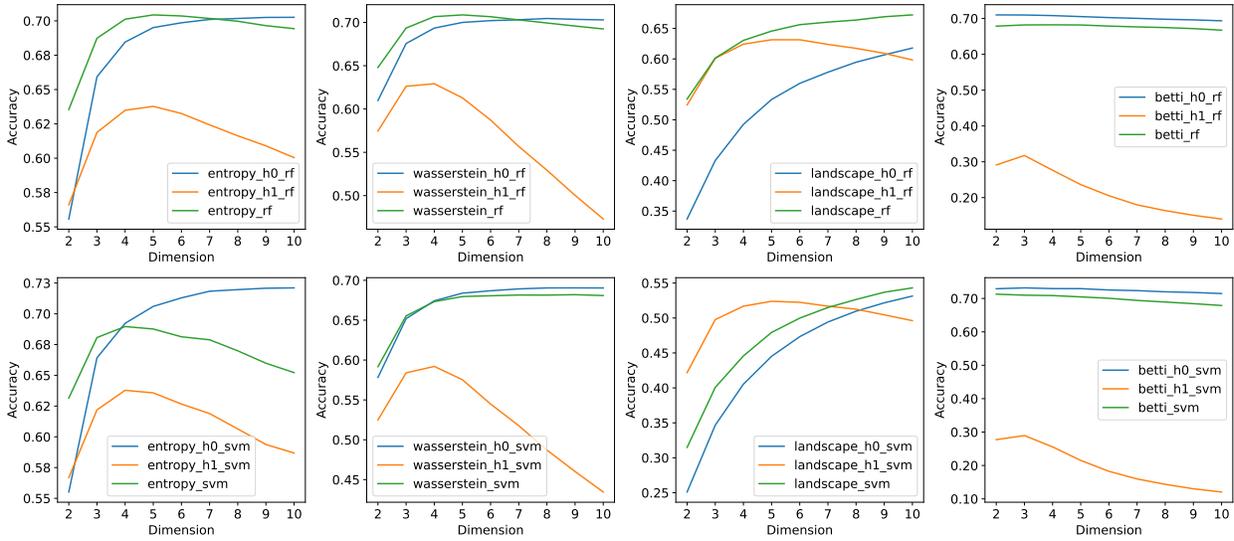


Fig. 3. The mean of the accuracy in each embedding dimension across all delays.

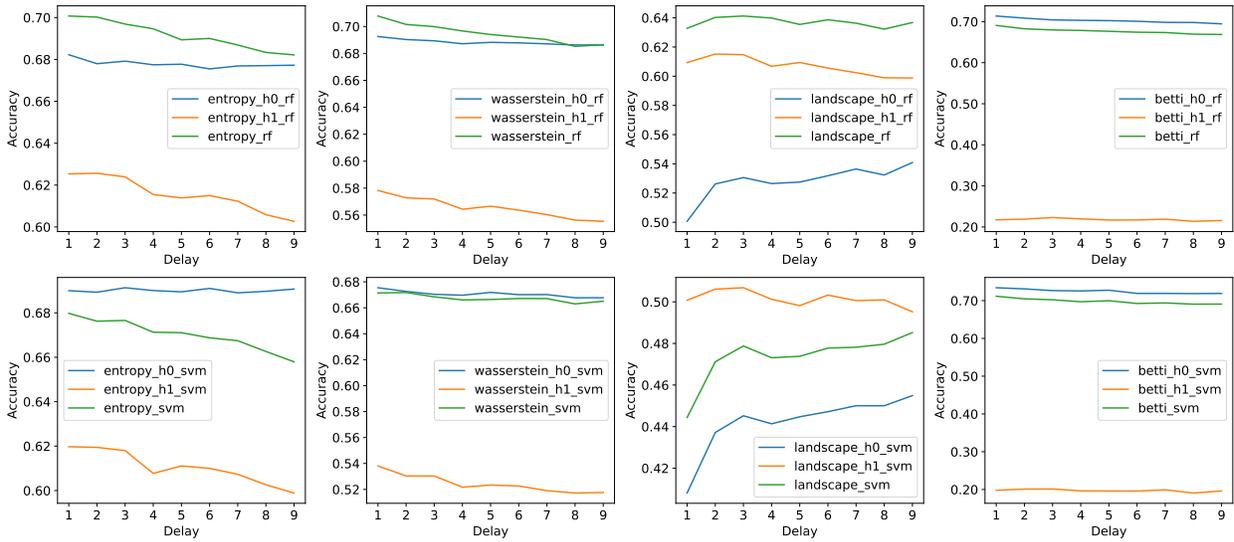


Fig. 4. The mean of the accuracy in each embedding delay across all dimensions.

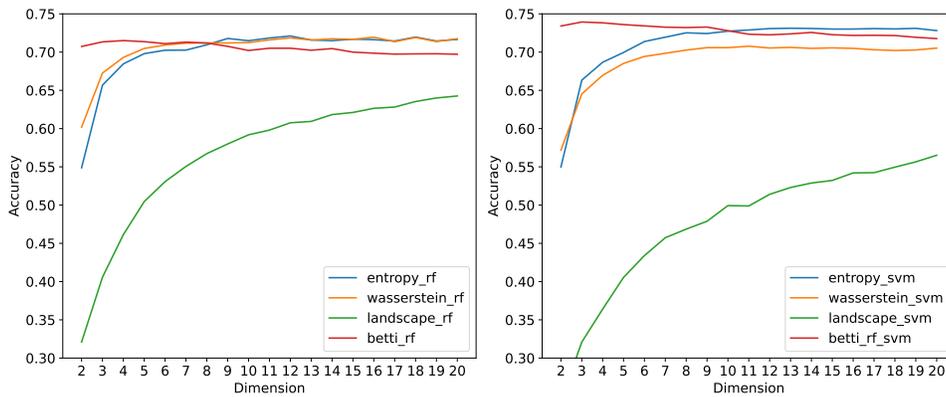


Fig. 5. The classification accuracy with embedding dimensions ranged from 2 to 20.

TABLE I. CLASSIFICATION ACCURACY USING TOPOLOGICAL FEATURES AND OTHER TIME DOMAIN FEATURES

Features		Classifier	Mean accuracy	Peak accuracy	
Topological features	Persistent entropy	SVM	73.02	73.12	
	Wasserstein amplitude		70.49	70.76	
	Landscape amplitude		56.32	56.51	
	Betti amplitude		73.50	73.93	
Time domain features	RMS (Root mean square)		70.36	70.63	
	TD (Time Domain Statistics)		68.19	68.48	
Topological features	Persistent entropy		Random forest	71.51	71.69
	Wasserstein amplitude			71.69	72.07
	Landscape amplitude	64.35		64.57	
	Betti amplitude	71.37		71.77	
Time domain features	RMS (Root mean square)	71.26		71.60	
	TD (Time Domain Statistics)	69.85		70.19	

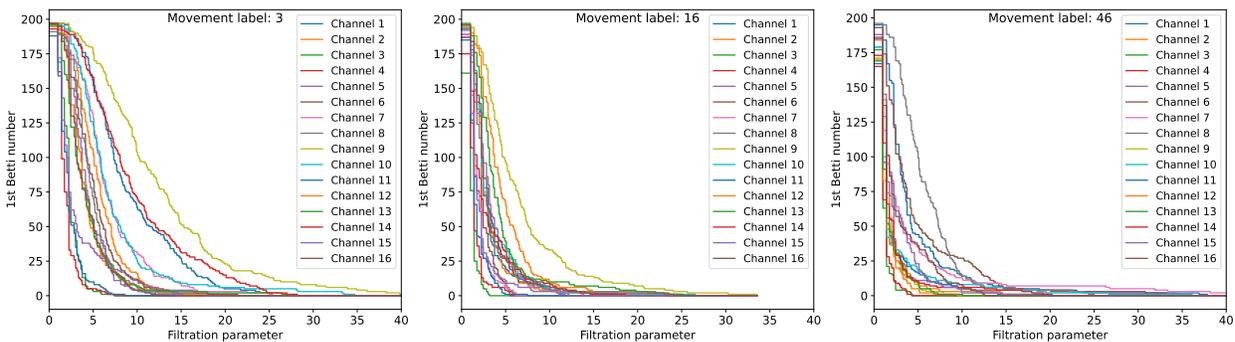


Fig. 6. Betti curves of one sample randomly selected from each of three types of hand movements, respectively.

### C. Compare to Other Time-domain Features

To evaluate the performance of topological features relative to other feature types, we conducted an additional experiment using two time-domain features: Root Mean Square (RMS) and Time Domain Statistics (TS). RMS computes the root-mean-square value for each channel and constructs a 16-dimensional feature vector. In contrast, TS incorporates the mean absolute value, number of zeros, number of slope changes, and waveform length calculated from each channel [36]. We separately fed these two feature types, as well as the topological features with embedding delay and dimension determined based on the findings from previous experiments, into Support Vector Machines (SVM) and Random Forest for movement classification. We repeated the training process 10 times and recorded the classification accuracies. The mean and peak accuracy are displayed in Table I. Our results indicate that topological features, with the exception of landscape amplitude, outperform RMS and TS. Notably, the SVM model employing persistent entropy or Betti amplitude achieved a high accuracy of over 73%.

### D. Visualize EMG Signals using Betti Curves

As previously mentioned, Betti amplitude is recognized for its robustness to changes in dimension and delay, and it has demonstrated the highest classification accuracy among other topological and time-domain features. This suggests that Betti amplitude may be better equipped to reveal the topology of EMG signals. However, Betti curve encodes even more topological information than Betti amplitude, as Betti ampli-

tude calculates only the  $L_2$  distance between Betti curves. As a result, Betti curve can effectively differentiate topological patterns of distinct hand movements. This has inspired us to employ Betti curve for visualizing the topological patterns of various hand movements. For instance, we have plotted Betti curves of randomly selected samples from three types of hand movements in Fig. 6. It is important to note that only the 0-th Betti number, which counts the connected components in the topological space, is displayed in the figure, as we have demonstrated that only the information of  $H_0$  is effective in classifying EMG signals. The figure clearly illustrates the distinct patterns of Betti curves in each channel for different hand movements.

To gain a broader understanding of the topological patterns for a specific type of hand movement, we calculated the average Betti curve by taking the arithmetic mean of Betti curves from all samples within that particular type of hand movement. This offers a more accurate representation of the general topological patterns for a given movement. By visualizing the average Betti curves of various hand movements in a single figure, we can easily observe the clear differences in topological patterns among different movements. This is demonstrated in Fig. 7, where the curve labeled “0” represents the “rest” movement, which is the fastest curve that drops to 0. This suggests that the topological space of “rest” has minimal topological complexity. Defining a distance metric on the average Betti curve could potentially aid in measuring the difficulty of discriminating a movement from EMG signals.

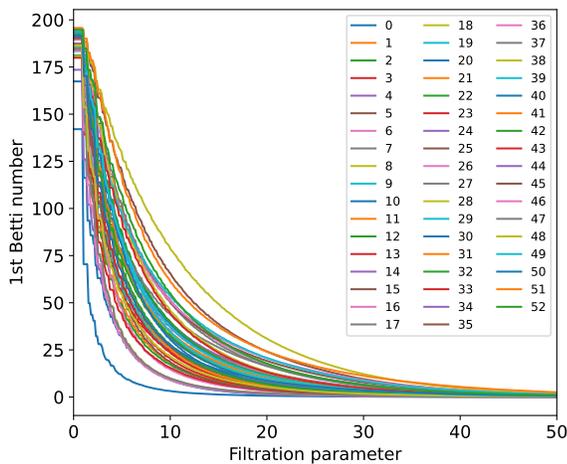


Fig. 7. The average Betti curves calculated for each of the 53 hand movements based on all training samples.

#### IV. CONCLUSIONS

The goal of this study was to determine whether topological features could enhance the accuracy of EMG signal classification for hand movements and contribute a methodology to the analysis of bio-electrical signals. The main findings of this study are as follows: (1) Topological features can effectively classify EMG signals, achieving the highest classification accuracy of 73.93%, outperforming the other tested time-domain features by nearly 2% in the experiment. (2) Topological features of  $H_0$  prove more effective than those of  $H_1$ . In general, higher embedding dimensions lead to increased accuracy, while embedding delay has a smaller impact on classification accuracy. (3) Among the four tested topological features, Betti amplitude is the most stable, and we have introduced Betti curves for visualizing the shape of hand movement EMG signals. Future research will explore whether these findings can be applied to other types of bio-electrical signals.

#### ACKNOWLEDGMENT

This research was funded by Scientific Research Foundation of Chongqing University of Technology and supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (No. KJQN202101108).

#### REFERENCES

- [1] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, pp. 11–35, 2006.
- [2] C. Sapsanis, G. Georgoulas, and A. Tzes, "Emg based classification of basic hand movements based on time-frequency features," in *21st Mediterranean conference on control and automation*. IEEE, 2013, pp. 716–722.
- [3] M. A. Aceves-Fernandez, J. Ramos-Arreguin, E. Gorrostieta-Hurtado, J. Pedraza-Ortega *et al.*, "Methodology proposal of emg hand movement classification based on cross recurrence plots," *Computational and mathematical methods in medicine*, vol. 2019, 2019.
- [4] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by emg signals and arm motions," *IEEE transactions on robotics and automation*, vol. 19, no. 2, pp. 210–222, 2003.

- [5] N. Rabin, M. Kahlon, S. Malayev, and A. Ratnovsky, "Classification of human hand movements based on emg signals using nonlinear dimensionality reduction and data fusion techniques," *Expert Systems with Applications*, vol. 149, p. 113281, 2020.
- [6] M. Köppen, "The curse of dimensionality," in *5th online world conference on soft computing in industrial applications (WSC5)*, vol. 1, 2000, pp. 4–8.
- [7] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, pp. 1–13, 2010.
- [8] A. Phinyomark, S. Thongpanja, H. Hu, P. Phukpattaranont, and C. Limsakul, "The usefulness of mean and median frequencies in electromyography analysis," *Computational intelligence in electromyography analysis-A perspective on current applications and future challenges*, vol. 81, p. 67, 2012.
- [9] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 27, no. 4, pp. 760–771, 2019.
- [10] Y. He, O. Fukuda, N. Bu, H. Okumura, and N. Yamaguchi, "Surface emg pattern recognition using long short-term memory combined with multilayer perceptron," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 5636–5639.
- [11] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, "Persistence barcodes for shapes," *International Journal of Shape Modeling*, vol. 11, no. 02, pp. 149–187, 2005.
- [12] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.
- [13] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.
- [14] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson, "Extracting insights from the shape of complex data using topology," *Scientific reports*, vol. 3, no. 1, pp. 1–8, 2013.
- [15] E. Carlsson, G. Carlsson, and V. De Silva, "An algebraic topological method for feature identification," *International Journal of Computational Geometry & Applications*, vol. 16, no. 04, pp. 291–314, 2006.
- [16] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.
- [17] Y. Lee, S. D. Barthel, P. Dłotko, S. M. Moosavi, K. Hess, and B. Smit, "Quantifying similarity of pore-geometry in nanoporous materials," *Nature communications*, vol. 8, no. 1, pp. 1–8, 2017.
- [18] D. Chen, J. Liu, J. Wu, G.-W. Wei, F. Pan, and S.-T. Yau, "Path topology in molecular and materials sciences," *Journal of Physical Chemistry Letters*, vol. 14, no. 4, p. 954–964, 2023.
- [19] Z. Cang, L. Mu, and G.-W. Wei, "Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening," *PLoS computational biology*, vol. 14, no. 1, p. e1005929, 2018.
- [20] M. Wang, Z. Cang, and G.-W. Wei, "A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation," *Nature Machine Intelligence*, vol. 2, no. 2, pp. 116–123, 2020.
- [21] A. Bukkuri, N. Andor, and I. K. Darcy, "Applications of topological data analysis in oncology," *Frontiers in artificial intelligence*, vol. 4, p. 659037, 2021.
- [22] V. De Silva, R. Ghrist *et al.*, "Homological sensor networks," *Notices of the American mathematical society*, vol. 54, no. 1, 2007.
- [23] V. Snášel, J. Nowaková, F. Xhafa, and L. Barolli, "Geometrical and topological approaches to big data," *Future Generation Computer Systems*, vol. 67, pp. 286–296, 2017.
- [24] N. Ravishanker and R. Chen, "Topological data analysis (tda) for time series," *arXiv preprint arXiv:1909.10604*, 2019.
- [25] S. Gholizadeh and W. Zadrozny, "A short survey of topological data analysis in time series and systems analysis," *arXiv preprint arXiv:1809.10745*, 2018.

- [26] Y. Umeda, J. Kaneko, and H. Kikuchi, "Topological data analysis and its application to time-series data analysis," *Fujitsu Scientific & Technical Journal*, vol. 55, no. 2, pp. 65–71, 2019.
- [27] C. M. Pereira and R. F. de Mello, "Persistent homology for time series and spatial data clustering," *Expert Systems with Applications*, vol. 42, no. 15-16, pp. 6026–6038, 2015.
- [28] F. A. Khasawneh and E. Munch, "Chatter detection in turning using persistent homology," *Mechanical Systems and Signal Processing*, vol. 70, pp. 527–541, 2016.
- [29] M. Gidea and Y. Katz, "Topological data analysis of financial time series: Landscapes of crashes," *Physica A: Statistical Mechanics and its Applications*, vol. 491, pp. 820–834, 2018.
- [30] 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.
- [31] P. S. Ignacio, C. Dunstan, E. Escobar, L. Trujillo, and D. Uminsky, "Classification of single-lead electrocardiograms: Tda informed machine learning," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019, pp. 1241–1246.
- [32] H. Dlugas, "Electrocardiogram feature extraction and interval measurements using optimal representative cycles from persistent homology," *bioRxiv*, pp. 2022–02, 2022.
- [33] M. Atzori, A. Gijsberts, S. Heynen, A.-G. M. Hager, O. Deriaz, P. Van Der Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the ninapro database: A resource for the biorobotics community," in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2012, pp. 1258–1265.
- [34] S. Pizzolato, L. Tagliapietra, M. Cognolato, M. Reggiani, H. Müller, and M. Atzori, "Comparison of six electromyography acquisition setups on hand movement classification tasks," *PloS one*, vol. 12, no. 10, p. e0186132, 2017.
- [35] U. Bauer, "Ripser: efficient computation of vietoris–rips persistence barcodes," *Journal of Applied and Computational Topology*, vol. 5, no. 3, pp. 391–423, 2021.
- [36] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE transactions on biomedical engineering*, vol. 40, no. 1, pp. 82–94, 1993.