

Article

Towards Analysis of Multivariate Time Series Using Topological Data Analysis

Jingyi Zheng ^{1,*}, Ziqin Feng ^{1,†} and Arne D. Ekstrom ²¹ Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA; zzzf0006@auburn.edu² Department of Psychology and Evelyn McKnight Brain Institute, University of Arizona, Tucson, AZ 85721, USA; adekstrom@arizona.edu

* Correspondence: jzz0121@auburn.edu

† These authors contributed equally to this work.

Abstract: Topological data analysis (TDA) has proven to be a potent approach for extracting intricate topological structures from complex and high-dimensional data. In this paper, we propose a TDA-based processing pipeline for analyzing multi-channel scalp EEG data. The pipeline starts with extracting both frequency and temporal information from the signals via the Hilbert–Huang Transform. The sequences of instantaneous frequency and instantaneous amplitude across all electrode channels are treated as approximations of curves in the high-dimensional space. TDA features, which represent the local topological structure of the curves, are further extracted and used in the classification models. Three sets of scalp EEG data, including one collected in a lab and two Brain–computer Interface (BCI) competition data, were used to validate the proposed methods, and compare with other state-of-art TDA methods. The proposed TDA-based approach shows superior performance and outperform the winner of the BCI competition. Besides BCI, the proposed method can also be applied to spatial and temporal data in other domains such as computer vision, remote sensing, and medical imaging.

Keywords: topological data analysis; Hilbert–Huang transform; scalp EEG; persistent homology; brain–computer interface



Citation: Zheng, J.; Feng, Z.; Ekstrom, A.D. Towards Analysis of Multivariate Time Series Using Topological Data Analysis. *Mathematics* **2024**, *12*, 1727. <https://doi.org/10.3390/math12111727>

Academic Editor: Moo K. Chung

Received: 27 April 2024

Revised: 15 May 2024

Accepted: 30 May 2024

Published: 1 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

MSC: 55-11; 91E10

1. Introduction

Topological data analysis (TDA) is an innovative approach increasingly employed in statistical applications and machine learning, offering a fresh perspective for exploring complex data structures, especially the shapes of data. One of the most widely used TDA algorithms is persistent homology (PH), introduced in [1,2]. Homology is a fundamental tool in the area of algebraic topology employed to capture the topological structures of simplicial complexes such as components, loops, and voids, etc. PH computes the homology of a Vietoris–Rips complex filtration built on the data, which is typically a finite metric space in high dimensions. The PH algorithm reveals the changes of these topological structures in the simplicial complex filtration and gains insights from the data.

In recent years, there has been a growing interest in leveraging TDA to extract pertinent topological features from signals in the field of signal processing. For the analysis of EEG signals, recent research [3–10] has demonstrated that the extracted topological features contain valuable information relevant to various neurological disorders. As summarized in the survey [11], the following three approaches are commonly employed in the TDA analysis of EEG signals:

- (i) Applying TDA to the EEG signals channel by channel using sublevel filtration [9];

- (ii) Applying TDA to the point cloud using Vietoris–Rips filtration, with each channel as a point and distance measured by connectivity measures such as Pearson correlation coefficients [6];
- (iii) Applying TDA to the reconstructed state space using time-delay embedding backed by Taken’s embedding theorem [12,13].

Persistent diagrams, also called barcodes, are then generated along with the filtration processes. Various methods exist for extracting topological features from these persistent diagrams. For instance, areas of one-dimensional Betti curves has been used to detect delirium using bispectral EEG (BSEEG) [7].

However, the current TDA analyses are applied either on the EEG signals or on the frequency information extracted from the Fourier transform, which are not comprehensive representation of both the time and frequency information embedded in EEG signals. In methods (i) and (iii), TDA is applied to the signals channel by channel; hence, the information of spatial distribution and correlation among the channels are not captured; and method (ii) misses the local information of the signals, which is hidden in short time intervals.

To address these challenges, in this paper, we propose a new way to extract TDA features from scalp EEG signals. We consider the multivariate brain signals as a curve in a Banach space. The recorded multi-channel scalp EEG signals can be treated as an approximation of the true brain signal curves with sampled spatial and temporal points (i.e., electrode channels and samples). The signal curve can be further segmented based on time intervals (2 ms/4 ms/8 ms, 100 ms, and 200 ms), from which local topological features are extracted. Moreover, the instantaneous frequency and instantaneous power of the scalp EEG signals are extracted via Hilbert–Huang Transform (HHT), and are also treated as curves in the Banach space. The TDA features are further extracted from the segmented frequency and power curves. In this paper, we utilized two ways of extracting topological features from a persistent diagram and compared their discriminative capability in the classifiers. The first one is the Betti numbers on five evenly distributed scales [5] and the second one is the areas of Betti curves, with dimensions being 0 and 1 in both ways. The contributions of this work can be summarized as follows:

- The multivariate time series are treated as approximations of curves in a Banach space and further segmented to extract local topological features.
- The time–frequency representation of EEG signals are extracted via HHT, and the instantaneous frequency and power series are treated as approximations of curves in a Banach space.
- The proposed TDA approach extracts spatial, temporal, and frequency information embedded in the multivariate time series and can be applied in other domains such as computer vision, remote sensing, medical imaging etc.

This is the first study that extracts time–frequency representation from multi-channel scalp EEG signals and considers the multivariate frequency and power series as curves in a Banach space to extract local topological features. To evaluate the proposed method and compare with other state-of-art TDA approaches, three scalp EEG datasets including two from a Brain–computer Interface (BCI) competition and one dataset collected as part of a cognitive neuroscience study are utilized. Besides comparing with other TDA approaches via cross-validation, we also compare the proposed method with the winner of the competition as well as the classification performance reported in the literature on the test data.

Compared to existing TDA approaches, our proposed method demonstrates significantly superior performance. Additionally, when compared to various signal processing methods, our TDA approach exhibits robust and stable performance across subjects, with average results being comparable and even better than some advanced methods. In the BCI application, it is important to have an accurate and robust classifier that translates the brain signals into commands for external devices. Employing TDA techniques enable

the researchers to extract meaningful features from BCI data and reduce noise, which promotes the accuracy of signal processing and classification. Furthermore, TDA allows for the exploration of topological properties across different brain regions, which enhances a deeper comprehension of neural processes.

The remainder of this paper is organized as follows. In Section 2, we introduce the three datasets used for validation and comparison. In Section 3, we discuss the details of the proposed TDA method. In Section 4, we validate the proposed method on three EEG datasets and compare it with other state-of-art TDA approaches as well as the winner of the BCI competition. Finally, we conclude the paper in Section 5.

2. Data Description

To validate the proposed method and further compare it with the state-of-art TDA approaches in the literature, we employed three scalp EEG datasets. To encompass a comprehensive representation of cases encountered in EEG-based BCI, the following three datasets with varying numbers of electrode channels, ranging from three to one hundred and eighteen, are selected.

Data 1 are from Dataset IIb from BCI competition IV [14]. The data contain both the EEG and electrooculogram (EOG) activity of nine subjects, but, in this study, we only leveraged the EEG signals, which were recorded from three electrode channels (C3, Cz, and C4) with a sampling frequency of 250 Hz. Each subject participated in the cue-based screening paradigm that consisted of two classes, motor imagery (MI) of the left hand (class 1) and the right hand (class 2). There were a total of three online feedback sessions, with the first session (03T) serving as the training data and the other two sessions (04E, 05E) as test data. Each session contained four runs with smiley feedback, and each run consisted of twenty trials per class, resulting in a total of eighty trials per class. The data provided by the competition were bandpass-filtered between 0.5 and 100 Hz and filtered by a notch filter at 50 Hz. No extra pre-processing was conducted in this study. The winner of the competition was the algorithm giving the largest kappa value on the test data.

Data 2 are from Dataset IVa from BCI competition III [15]. It contains 118-channel scalp EEGs from five subjects with two motor imagery tasks, concerning the right hand and foot. It is worth mentioning that the challenge of these data are the small training samples. Specifically, subject aa underwent 168 trials for training and 112 trials for testing, subject al underwent 224 trials for training and 56 for testing. However, subject av underwent 84 trials for training and 196 trials for testing. Subject aw only underwent 56 trials for training but 224 trials for testing. Subject ay only underwent 28 trials for training but 252 trials for testing. To overcome the problem, training trials from other subjects could also be considered during the model training, as recommended by the competition [15]. Therefore, we included the training trials from subject av when training the classifier of subject ay. The performance measure of the competition was the overall accuracy.

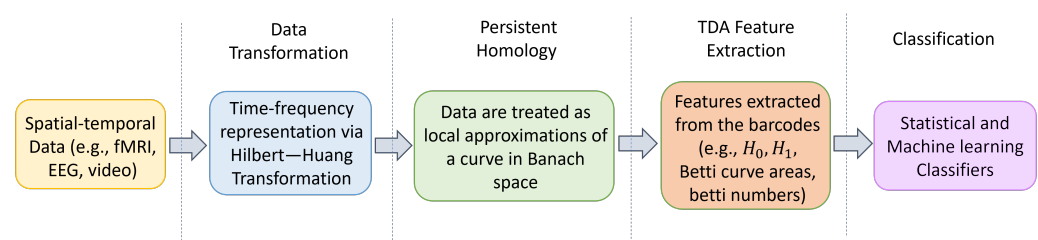
Data 3 are from the Human Spatial Cognition Laboratory at the University of Arizona [16,17]. It consists of 64-channel scalp EEGs from 19 subjects (written informed consent was obtained in accordance with the Institutional Review Board at the University of Arizona) who performed a spatial distance monitoring task to identify binary outcomes, short vs. long distance in an immersive virtual environment. There were 48 trials per subject with 24 trials in each class. The signals were processed by 1–50 Hz bandpass filtering, artifact amelioration, and eye/muscle artifacts removal. The details of the experimental design and data pre-processing can be found in [16]. The information about the three datasets are summarized in Table 1. All datasets are publicly available.

Table 1. The configurations of datasets.

	Data 1 BCI IV Iib	Data 2 BCI III IVa	Data 3 Lab Data
Subject Number	9	5	19
Channel Number	3	118	64
Trials per Class	80	14–112	24
Trial Length	1000	2500	2828
Sampling Rate	250 Hz	1000 Hz	500 Hz

3. Method

The proposed TDA approach is composed of four parts, including data transformation, persistent homology, TDA feature extraction, and classification, as outlined in Figure 1. Details of each component are discussed in the following section.

**Figure 1.** The proposed TDA-based processing pipeline.

3.1. Data Transformation

Existing TDA-based approaches for analyzing EEG signals extract the topological features from signals either in the original time domain or frequency domain via Fourier transform. Scalp EEG signals are assumed to be stationary in many methods, but temporal drift and experimental manipulations may introduce non-stationarities. In other words, the frequency and power of the signals change over time. Fourier transform has limitations for revealing the frequency information of EEG signals because it estimates a constant power for each frequency during the entire time span. To better capture the frequency and power shift over time, we leveraged the Hilbert–Huang Transformation (HHT), proposed by [18], to reveal the dynamic time–frequency representation of the EEG signals.

Different from the Fourier or wavelet transform, HHT is a data-driven approach. It first decomposes the signals into sub-signals, named Intrinsic Mode Functions (IMFs), via Empirical Mode Decomposition (EMD), and then reveals the time–frequency information of each IMF via Hilbert Transform (HT). Specifically, we denoted the signal from one channel during one trial as $x(t)$. The EMD decomposes the signal into a collection of IMFs, denoted as $\{c_i(t), i = 1, \dots, n\}$, via a sifting process [18]. The instantaneous frequency $\omega_i(t)$ and instantaneous amplitude $a_i(t)$ of each IMF is then revealed by HT.

$$x(t) = \sum_{i=1}^n c_i(t) + r(t) = \sum_{i=1}^n \text{Re}\{a_i(t) \exp(i \int \omega_i(t) dt)\} + r(t).$$

For scalp EEG signals, the dynamic frequency and power revealed by HHT are essential information because they are used to reveal the subtle changes in the underlying dynamic process in the brain. Moreover, due to the sifting process in EMD, the IMFs are in descending frequency bands, with the first IMF carrying the highest frequency sub-signal. Considering the frequency bands of brain oscillations, in this study, we chose the first four IMFs that cover delta, theta, alpha, and beta waves.

We denoted the EEG signals during each trial as $X(t)_{N \times T}$, with each row $x_n(t)$ ($n = 1, 2, \dots, N$) being length T signals from each of the N channels. HHT was applied on each $x_n(t)$, and the resulting instantaneous frequency and amplitude of the first four IMFs are denoted as $\omega_{ni}(t)$ and $a_{ni}(t)$, respectively, where $i = 1, 2, 3, 4$. Figure 2 shows an example of the time-varying frequency and power (i.e., square of the amplitude) of the first four IMFs under two different tasks. The color represents the instantaneous amplitude of the corresponding instantaneous frequency. By organizing the dynamic frequency and power information from each channel in a matrix form, we obtained one frequency matrix, denoted as F_i , and one power matrix, denoted as P_i , for each IMF, resulting in a total of eight matrices. All matrices are $N \times T$, with each row being $\omega_{ni}(t)$ or $a_{ni}(t)$. In the following, TDA features are extracted from the eight matrices that represent the dynamic time–frequency information of the signals.

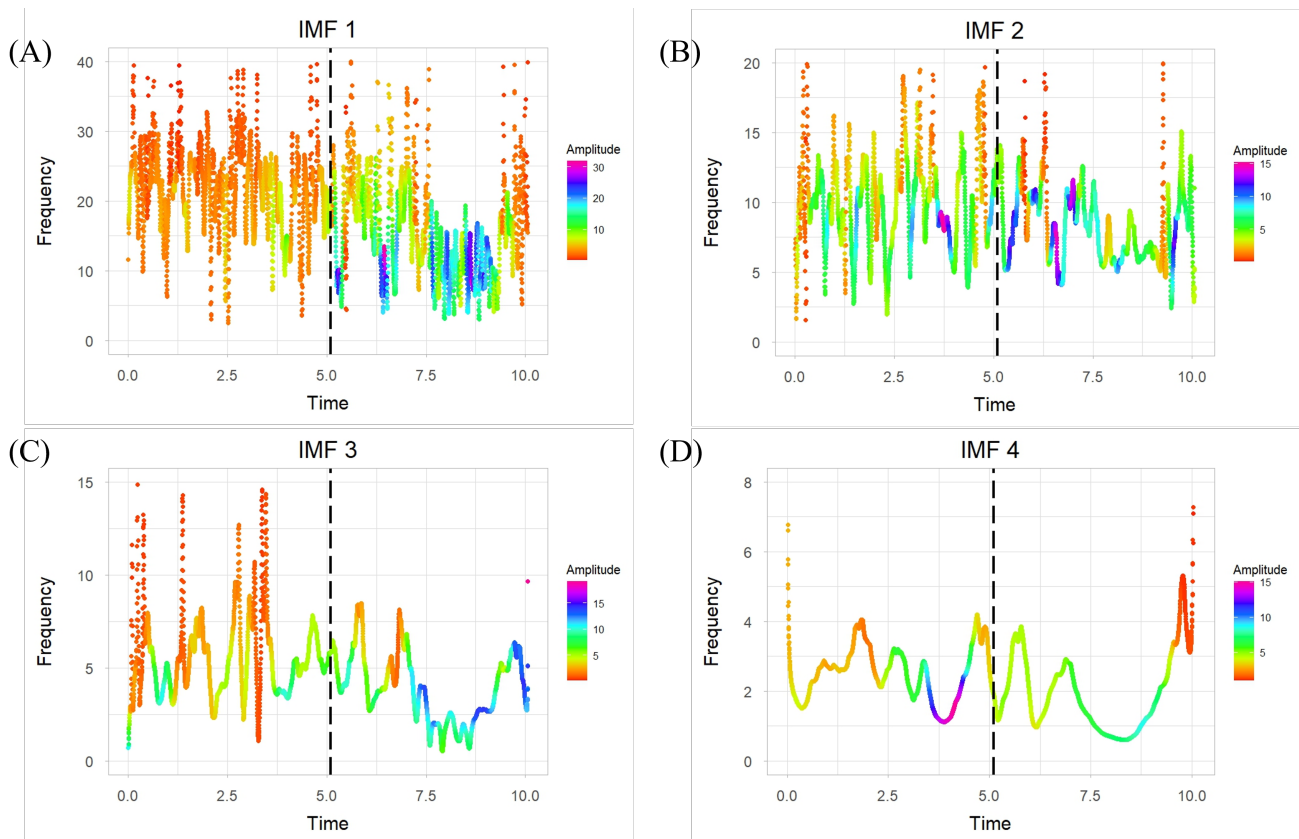


Figure 2. An example of the HHT-transformed data during two different tasks: the instantaneous frequency and instantaneous amplitude (represented by color) of IMF 1 (A), IMF 2 (B), IMF 3 (C), and IMF 4 (D).

3.2. Persistent Homology and Vietoris–Rips Filtration

An abstract simplicial complex K over a finite set S is a collection of nonempty subsets of the set S , such that $\bigcup K = S$ and, for any two subsets τ, σ of S , $\tau \subseteq \sigma \in K$ implies that $\tau \in K$. Every abstract simplicial complex has a geometric realization, which is a set of points, line segments, triangles, and general n -simplices. An n -simplex (n -dimensional simplex) is the convex hull of $n + 1$ many affine independent points, making it naturally a geometric object. The dimension of a simplicial complex is the maximum of the dimensions of its simplices. For instance, any graph (G, E) is a one-dimensional simplicial complex, with G being the set of points and E being the set of 1-simplices. Consequently, graphs capture the pairwise relations between the points, and simplicial complexes yield the higher-order relationships among the points.

For any metric space X and scale $r \geq 0$, the Vietoris–Rips complex $VR(X; r)$ is the complex with simplices to be the nonempty finite subsets of X of diameter $\leq r$. In TDA, the metric space X is the data sampled from some unknown topological space Y and one would like to use the state space X to uncover the topological features of the unknown space Y . The idea behind TDA/PH is to investigate the ‘shape’ of the complex $VR(X; r)$ using homology as the scale r varies from small to large and trust the persisting topological features to be the representative of the topological properties of Y . This idea is supported by a fundamental result, the nerve theorem [19], which states that the nerve complex of a nice open cover of a topological space Y is homotopic to Y . This theorem shows that the topological features of a space Y is encoded in finite abstract combinatorial structures built on the space Y .

In algebraic topology, (co)homologies are used to capture the topological properties of simplicial complexes such as the numbers of components, loops, and voids, etc. In the following, we provide a brief overview of the computation process for homologies with coefficients in \mathbb{Z}_2 , which is the field with two elements, 0 and 1, where $1 + 1 = 0$.

Let K be a finite simplicial complex. We fix an integer $p \geq 0$. We define $C_p(K)$ to be the vector space on the field \mathbb{Z}_2 with a base to be the collection of p -simplices. Furthermore, we define a boundary map ∂_p from $C_p(X)$ to $C_{p-1}(X)$ to be the linear extension of the intuitive boundary map on any p -simplices. The kernel of ∂_p , denoted by $Z_p(K)$, is the collection of p -cycles and the image of ∂_{p+1} , denoted by $B_p(K)$, is the collection of p -boundaries, i.e., the ‘filled’ p -cycles. Then, the p^{th} simplicial homology group with coefficients in \mathbb{Z}_2 is defined to be

$$H_p(K) = Z_p(K) / B_p(K).$$

Therefore, the dimension of the vector space $H_p(K)$ gives the count of the ‘unfilled’ p -dimensional ‘hole’ in K , which is also called the p th Betti number. Topologically, a one-dimensional ‘hole’ is a circle and a two-dimensional ‘hole’ is a void. The dimension of the 0th homology group $H_0(K)$ gives the number of connected components in the simplicial complex K .

To extract a filtration of simplicial complexes from a state space X , Vietoris–Rips filtration was employed. It starts with expanding each point in the state space to a disk with a radius of zero. The radii of the disks grow uniformly, and then the procedure ends when they reach a predetermined value. The predetermined value in our calculation is the one such that the resulting simplicial complex loses all the topological structures, i.e., homotopy equivalent to a singleton. For each radius $r_i \geq 0$ with $i = 0, 1, 2, \dots, n$, we obtain a Vietoris–Rips complex $VR(X; r_i)$, denoted by K_{r_i} . This yields a filtration of nested simplicial complexes, $K_{r_0}, K_{r_1}, K_{r_2}, \dots, K_{r_n}$ with

$$K_{r_0} \subseteq K_{r_1} \subseteq K_{r_2} \subseteq \dots \subseteq K_{r_n}.$$

The persistent homology of this filtration of simplicial complexes $\{K_{r_i} : i = 0, 1, \dots, n\}$ is the homology groups $\{H_p(K_{r_i}) : p \geq 0 \text{ and } i = 1, \dots, n\}$ connected by the mapping $\{\phi_p^{i,j} : p \geq 0 \text{ and } 0 \leq i < j \leq n\}$, where each $\phi_p^{i,j}$ is the linear transformation from $H_p(K_{r_i})$ to $H_p(K_{r_j})$ induced by the inclusion map. For each homology dimension $p \geq 0$, we obtain a finitely generated persistence module \mathbb{V}_p over the field \mathbb{Z}_2 :

$$H_p(K_{r_0}) \xrightarrow{\phi_p^{0,1}} H_p(K_{r_1}) \xrightarrow{\phi_p^{1,2}} \dots \xrightarrow{\phi_p^{n-1,n}} H_p(K_{r_n}).$$

By the structure theorem [20], such a persistence module can be decomposed to a direct sum of interval modules $\mathbb{I}[r_i, r_j)$ with $i < j$ in the following form:

$$0 \Rightarrow \dots \Rightarrow 0 \Rightarrow \mathbb{Z}_2 \xrightarrow{id} \dots \xrightarrow{id} \mathbb{Z}_2 \Rightarrow 0 \Rightarrow \dots \Rightarrow 0$$

where 0 is the trivial groups and id represents the identity map. Each of the interval modules represents a topology structure that persists in the interval $[r_i, r_j)$. We denote the collection of such intervals as \mathcal{J}_p . So, we obtain the the following decomposition of the persistence module:

$$\mathbb{V}_p \cong \bigoplus_{I \in \mathcal{J}_p} \mathbb{I}(I)$$

In this decomposition, topological invariants (components, circles, voids, etc) persist in these Vietoris–Rips complexes on the corresponding intervals. Hence, each p -dimensional interval module corresponds to one of these topological invariants. These intervals are called p -dimensional Betti intervals, in the form $[r_{\text{birth}}, r_{\text{death}})$, which defines the scales at which a p -dimensional hole appears in the simplicial complex $K_{r_{\text{birth}}}$ and dies in the simplicial complex $K_{r_{\text{death}}}$. These topological features are not observable through the analytic approach with a fixed scale. We then denote the persistent barcode to be the collection of all these intervals. Then, the Betti curve of dimension p is defined to be

$$\beta_p(r) = \sum_{I \in \mathcal{J}_p} w(I) \mathbf{1}_{r \in I}$$

where w is the pre-chosen weight function and $\mathbf{1}_{r \in I}$ is the characteristic function, i.e., its value is 1 if $r \in I$, otherwise 0. If the weight function $w(I)$ is the constant function 1, then the value of the Betti curve at r is the number of Betti intervals containing the scale r . It is not hard to see that if the data are sampled from a line segment and the weight function is chosen to be the constant 1 function, then the area of the Betti curve at dimension 0 is exactly half of the length of the line segment. Hence, the areas of the Betti curves are closely related to the topological/geometric properties of the underlying space. In our proposed approach, the constant function 1 is chosen to be the weight function for the Betti curves.

To illustrate the process, we discuss two simple examples (A) and (B) in Figure 3. Example (A) contains three data points with the only zero-dimensional persistent homology being nontrivial (as shown in the figure, only red barcodes, H_0 , appear in the persistent barcode of (A)); and Example (B) contains five data points whose filtration has nontrivial zero- and one-dimensional persistent homology (both H_0 and H_1 appear in the persistent barcode as shown in the figure). A graphical representation of the persistent barcode of the datasets is also included and it is associated with the filtration of Vietoris–Rips complexes.

Persistent homology possesses a crucial property [21], wherein the persistence barcodes from Vietoris–Rips filtrations demonstrate remarkable stability when the data are contaminated. Specifically, the distance between the persistence barcodes (bottleneck distance) obtained via applying persistent homology on two datasets in a given dimension p is controlled by the distance of these two datasets (Gromov–Hausdorff metric). The computation of persistent homology is carried out using Python 3.12.1 [22] named Giotto-tda, which allows the users to perform Vietoris–Rips filtration analysis along with a time-delayed embedding of time series.

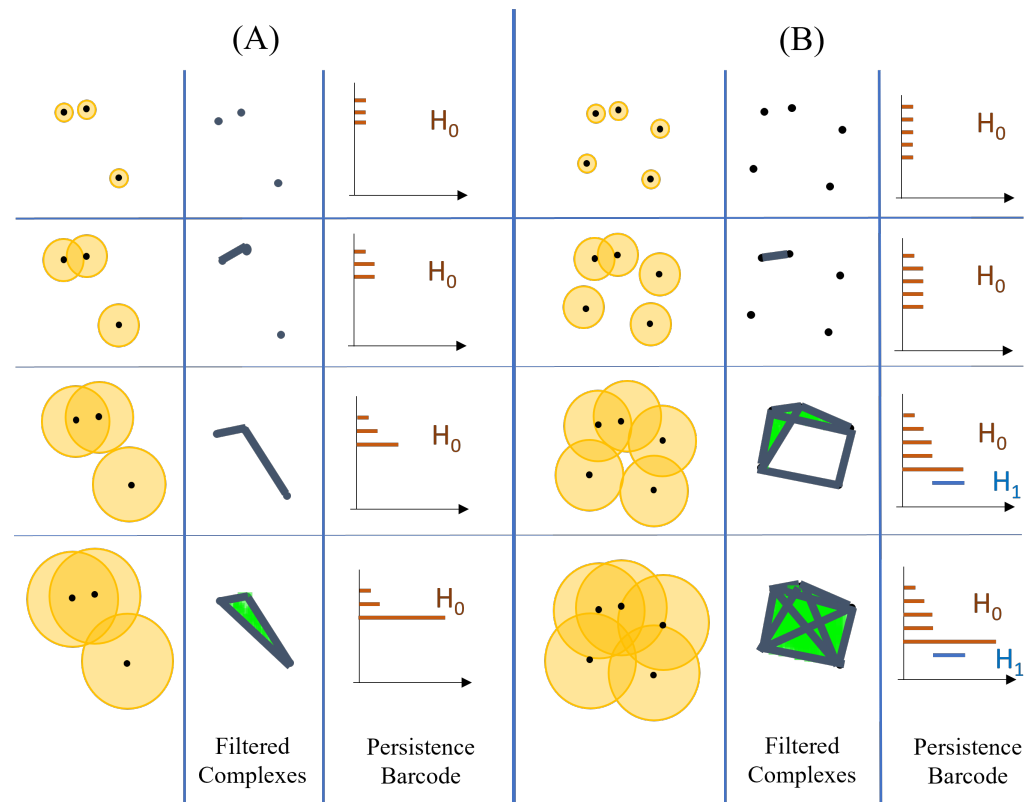


Figure 3. Illustrations of filtration and the associated barcode using three data points (A) and five data points (B), with black points representing the data. Both (A) and (B) show four stages of filtration along with the associated simplicial complexes and persistent barcodes. In stage 1, all points are isolated, giving the same number of bars and data points. As balls grow, some balls merge together resulting in the death of certain bars, and the length of each line segment (red/blue) represents the life span of certain topological features during the filtration. The red line segments denote the life span of connected components. Therefore, (A) has three red lines and (B) has five red lines in their barcodes. During the filtration of (B), a loop emerges in stage 3 of the filtration but disappears later. The life span of this feature is represented by the blue line segment in the barcode of (B). In (A), only the collection of zero-dimensional Betti intervals is nontrivial while both collections of zero- and one-dimensional Betti intervals are nontrivial in (B).

3.3. TDA Feature Extraction

The existing TDA-based frameworks extract topological features from the signal or transformed signal recorded in each channel, respectively, without considering the spatial information across channels. To overcome this limitation, we consider the transformed data across different locations as an approximation of some curves in a Banach space and further extract their local topological structures as follows.

Let M denote a compact manifold and $C(M)$ denote the collection of all real-valued continuous functions, which is a Banach space when it is equipped with the supreme norm. We consider the true brain signals during a time period as a function from a time interval to the Banach space $C(M)$ with M representing the scalp, a compact two-dimensional manifold. The inherent brain signals exhibit continuity both temporally and spatially, meaning that signals are present throughout the scalp at any given moment. Hence, the true brain signals can be considered as a curve in the Banach space $C(M)$. The scalp EEG signals are recorded via electrode channels placed at certain locations on the scalp with a specific sampling frequency. Therefore, the recorded brain signals (after pre-processing) can be considered as an approximation of the true brain signals, effectively capturing both temporal and spatial characteristics. The electrode channels act as an approximation of the entire scalp, and the signals are recorded discretely at specific time points determined by the

sampling rate. Thus, the recorded scalp EEG signals can be considered as an approximation of the curve. The same idea can be applied to the instantaneous frequency $\{F_i : i = 1, 2, 3, 4\}$ and instantaneous amplitude $\{P_i : i = 1, 2, 3, 4\}$, which are the transformed signals obtained in Section 3.1. For illustration purposes, we use F_1 as an example to discuss the following TDA feature extraction.

We recall that F_1 is a multivariate time series:

$$F_1(t)_{N \times T} = \{\mathbf{x}(t) \in \mathbb{R}^N : t = 1, 2, \dots, T\}$$

where T is the length of the time series and N is the number of channels. To obtain local topological properties, we first divide $F_1(t)_{N \times T}$ into shorter time segments with the length being ΔT . We denote the number of time intervals as L , i.e., $L \times \Delta T$ is a number equal to or slightly less than T . The data during each time segment,

$$F_{1\ell} = \{\mathbf{x}(t) \in \mathbb{R}^N : t = (\ell - 1)\Delta T + 1, (\ell - 1)\Delta T + 2, \dots, \ell\Delta T\}$$

with $1 \leq \ell \leq L$, are considered as a finite metric space with ΔT points in the Euclidean space \mathbb{R}^N .

Now, we consider $F_{1\ell}$ with a fixed ℓ such that $1 \leq \ell \leq L$. We build the Vietoris–Rips filtration $\text{VR}(F_{1\ell}; r)$ at scales $r \geq 0$, and then compute the homology groups of the filtration objects $\text{VR}(F_{1\ell}; r)$ at scales $r \geq 0$ with the homology dimension p being 0 and 1 using the package Giotto-tda [22]. These groups generate the persistent barcodes of the filtration objects, consisting of intervals in the form $[r_{\text{birth}}, r_{\text{death}})$, which represent the birth and death of some topological structure, either a connected component (zero-dimensional homology) or a loop (one-dimensional homology). To extract the TDA features, we calculate the areas $C_{p,\ell}$ of p -dimensional Betti curves of the persistent homology of $F_{1\ell}$ with $p = 0, 1$. The weight function used in the Betti curves is the constant function 1. Hence, the areas $C_{p,\ell}$ can also be calculated as the sum of all the life spans of each topological structure in the filtration objects, i.e.,

$$\sum \{|r_{\text{death}} - r_{\text{birth}}| : [r_{\text{birth}}, r_{\text{death}}) \in \mathcal{J}_{p,\ell}\},$$

where $\mathcal{J}_{p,\ell}$ is the collection of intervals in a persistent barcode of the segment $F_{1\ell}$, which correspond to the existences of p -dimensional topological structures during the filtration. A summary of this TDA feature extraction process is given below Algorithm 1.

Algorithm 1 TDA feature extraction.

INPUT $\mathbf{X}_{N \times T}$ and time interval length ΔT .

1. **Time Segment** Divide $\mathbf{X}_{N \times T}$ into a collection of $\{X_\ell : 1 \leq \ell \leq L\}$ with length of the time intervals to be ΔT satisfying that $L \times \Delta T \leq T$.
2. **Vietoris–Rips Filtration** Build Vietoris–Rips complexes $\text{VR}(X_\ell; r)$ by the collection of subsets of X_ℓ whose diameter is less or equal to some given $r > 0$.
3. **Persistent Homology** Compute homology groups of the filtration objects $\text{VR}(X_\ell; r)$ with $r \geq 0$ and homology dimensions being 0 and 1; then, obtain the persistent barcodes consisting of Betti intervals in the form $[r_{\text{birth}}, r_{\text{death}})$.
4. **TDA Features** Calculate the areas $C_{p,\ell}$ of p -dimensional Betti curves of each persistent homology on the dataset X_ℓ with $p = 0, 1$ as

$$\sum \{|r_{\text{death}} - r_{\text{birth}}| : [r_{\text{birth}}, r_{\text{death}}) \in \mathcal{J}_{p,\ell}\}$$

where $\mathcal{J}_{p,\ell}$ is the collection of p th dimensional Betti intervals in the persistent barcodes of X_ℓ for $1 \leq \ell \leq L$.

RETURN $C_{p,\ell}$, $p = 0, 1$, and $\ell = 1, 2, \dots, L$.

With the time interval length being ΔT , the total number of TDA features $\{C_{p,m} : p = 0, 1 \text{ and } \ell = 1, 2, \dots, L\}$ extracted from F_1 is $2L$. For each of the eight matrices, F_i and P_i with $i = 1, 2, 3, 4$ obtained in Section 3.1, we extract the TDA features with the same process, resulting in a total of $16L$ TDA features for each trial. To better capture the intrinsic local topological properties of the transformed signals, we use three different ΔT , 8 ms, 100 ms, and 200 ms, and denote the corresponding number of shorter time segments as L_1 , L_2 , and L_3 . Therefore, the total TDA features extracted for each trial is $16 \times (L_1 + L_2 + L_3)$.

Besides the proposed method, there are other widely used TDA approaches in the literature that we will compute and compare with. Therefore, we briefly introduce them in the following.

In the existing literature, there are two different ways to create the state space for EEG data, spatial embedding and time-delay embedding. For spatial embedding, a signal from each channel is considered as a point in the state space. Since there are only three channels in Data 1, TDA features from the state space obtained by spatial embedding do not provide any valid information about Data 1. Based on Taken's embedding theorem (see [12,13]), time-delay embedding is a very useful way to reconstruct state space for the single signal from one channel. One can reconstruct the state space of each signal independently with time-delay embedding in the following way. For a time series $x(t)$, an embedding with time delay Δt and embedding dimension d is a mapping ϕ from $x(t)$ to \mathbb{R}^d such that, for each t ,

$$\phi(x(t)) = (x(t), x(t + \Delta t), \dots, t + (d - 1)\Delta t).$$

The key to successfully reconstructing the state space is to choose the parameters, time-delay, embedding dimensions, and the time interval length. Selecting parameters arbitrarily can distort the reconstruction of the state space, leading to the obscuring of underlying assets and the highlighting of noise. The false nearest neighbor (FNN) test firstly proposed in [23] has been widely applied in determining the proper embedding dimension of a nonlinear system [24]. The authors in [5] use FNN to determine the best parameters for the TDA approach to reconstruct the state space using channel-wise time-delay embedding for original EEG data and find out that time delay $\Delta t = 1$ and embedding dimensions $d = 3$ and 5 with time interval lengths of 100 or 250 have the best performance. Hence, there are four different sets of parameters that perform best. We adopt all of them to extract the corresponding TDA features.

To illustrate the channel-wise time-delay embedding TDA approach, we use the time delay $\Delta t = 1$, embedding dimension $d = 3$, and time interval length of 100 as an example. Let $\{x(t) : t = 1, 2, \dots, T\}$ denote the original EEG signal from one channel during one trial. Then, we apply time-delay embedding ϕ with $d = 3$ and $\Delta t = 1$ to reconstruct the state space $X(t)$ in \mathbb{R}^3 . We notice that, in this case, the reconstructed state space $X(t)$ is a $3 \times (T - 2)$ matrix, i.e.,

$$X(t) = \{\phi(x(t))^T : t = 1, 2, \dots, T - 2\}.$$

Then, we divide the reconstructed state space $X(t)$ into segments

$$X_\ell(t) = \{\phi(x(t))^T : t = 100(\ell - 1) + 1, 100(\ell - 1) + 2, \dots, 100\ell\}$$

of equal time window 100 with $\ell = 1, 2, \dots, L$ and $100L \leq T - 2$. For each ℓ with $1 \leq \ell \leq L$, we build the Vietoris–Rips filtration, compute their homologies, and obtain the persistent barcodes of dimensions 0 and 1. We group the Betti intervals into $\mathcal{J}_{p,\ell}$ for $p = 0, 1$ and $\ell = 1, 2, \dots, L$ which contains all the p th dimensional Betti intervals in the persistent barcodes of the state space $X_\ell(t)$.

We adopted two different methods (A) and (B) in the literature to extract TDA features from each Betti interval collection $\mathcal{J}_{p,\ell}$ of the persistent barcodes, or, equivalently, the Betti curve $\beta_{p,\ell}$ with the pre-chosen weight function being the constant function 1, with $p = 0, 1$ and $\ell = 1, 2, \dots, L$:

- Method (A) extracts TDA features in the same way as Algorithm 1. We calculate the areas of the Betti curves $\beta_{p,\ell}$ mentioned above, and equivalently, the sum of the life spans of each p -dimensional topological structure in the Vietoris–Rips filtration of $X_\ell(t)$ for $p = 0, 1$ and $\ell = 1, 2, \dots, L$. The areas of the 1-dimensional Betti curve for each state space are used in [7] to detect delirium through BSEEG with different embedding dimensions, time delay, and time windows. Here, we adopted the best parameters according to [5] in the time-delay embedding process.
- Method (B) extracts the p -dimensional Betti number of the Vietoris–Rips complex $VR(X_\ell(t), r)$ at a pre-chosen set of scales for homology dimensions $p = 0, 1$. The scales are chosen as a fraction of a scale R such that the complex $VR(X_\ell(t), R)$ is contractible. Following the procedure in [25], the scale R is determined as follows: we choose an arbitrary point \mathbf{x}_0 in $X_\ell(t)$ and take R to be the maximum of $\{\|\mathbf{x}_0 - \mathbf{x}\| : \mathbf{x} \in X_\ell(t)\}$, which is also called the radius of the spherical volume of the corresponding state space. Then, we extract the TDA features of X_ℓ as the values of the Betti curve $\beta_{p,\ell}$ on the complex $VR(X_\ell(t), R)$ for $p = 0, 1$ at the scales $R/100, R/50, R/25, R/10, R/4$, i.e., the one-dimensional Betti number of each Vietoris–Rips complex $VR(X_\ell, r)$ at the corresponding scales [5].

We suppose that there are $2 \times L \times N$ many collections of Betti intervals with dimensions 0 and 1. Then, Method (A) extracts $2 \times L \times N$ number of TDA features, while Method (B) offers $10 \times L \times N$ as many TDA features. These two methods were applied on the EEG data as well as the transformed data, and compared with our proposed approach.

3.4. Classification

The classifier was constructed for each subject, respectively, predicting the label (i.e., MI in the BCI or tasks in the lab experiment) of the test trial. For each subject during each trial, $16 \times (L_1 + L_2 + L_3)$ TDA features were extracted using the proposed TDA approach. The number of features varied depending on the methods for TDA feature extraction. The classifiers we considered include both linear and nonlinear models: logistic regression with LASSO (Logistic), Linear Discriminate Analysis (LDA), Support Vector Machine (SVM) with linear and RBF kernel, K-Nearest Neighbors (KNN), and Random Forest (RF). For the machine learning models (SVM, KNN, and RF), recursive feature elimination (RFE) was employed to select the most relevant features for the classifiers. For the logistic regression, no extra feature selection was performed because LASSO preforms both variable selection and regularization.

The comparison was carried out in two phases: different classifiers were firstly compared using the same TDA features, then the best classifier was used to compare various TDA approaches including the proposed approach and other state-of-art TDA approaches in the literature. The performance of classifiers was assessed by accuracy and Cohen's kappa values. Specifically, accuracy measures the proportion of correct predictions among the total number of predictions. Kappa is a statistic that measures the the agreement between the prediction and the observed classes while considering the possibility of the agreement occurring by chance. It is computed as $\frac{P_o - P_e}{1 - P_e}$, where P_o is the proportion of agreement between the predicted and observed classifications and P_e is the hypothetical probability of agreement by chance. The kappa ranges from -1 to 1 with 1 indicating perfect agreement between predictions and observations. In short, for both accuracy and kappa values, a higher value indicates better classification performance.

When assessing the performance of classifiers, repeated five-fold cross-validation (CV) was used for all datasets. Regarding the BCI data (Data 1 and 2), an additional assessment was conducted on the test data, which were provided by the competition organizers. The classifiers submitted by all teams were evaluated using this test data with the performance ranked on a leader board. Hence, for BCI data, repeated CV was applied on the training data and the classifiers that were re-trained using all training trials were also tested on the test trials. The performance of the testing data was compared with the top-performing teams on the leader board as well as the results reported in the literature.

4. Results

For the comparison of classifiers and TDA features, we used Data 1 as an illustrating example, showing the detailed performance of each subject on the training trials. For Data 2 and Data 3, only the average training performance is reported.

With the proposed TDA features, we first compared the statistical and machine learning models based on their training performance estimated by repeated CV, including accuracy (Acc) and kappa (κ), which are summarized in Table 2 with the highest values highlighted in bold. Overall, Random Forest (RF) shows the superior performance with the highest accuracy and kappa, on average, as well as the highest accuracy and kappa for all subjects except one. Therefore, we chose RF as our classifier in the further analysis.

With RF as the classifier, Table 3 summarizes the classification performance, obtained via CV on the training data, of the proposed TDA approach and other state-of-art TDA methods in the literature, as discussed in Section 3.3. The highest values are highlighted in bold. The first two approaches in Table 3 compare method (A) and (B) applied on EEG signals, and method (A) shows better performance on average. Then, we applied a channel-wise time-delay embedding TDA approach to the instantaneous frequency F_i and amplitude P_i with $i = 1, 2, 3, 4$, and obtained collections of persistent barcodes of each segment. The same time delay, embedding dimensions, and time interval lengths were used in the process. Then, we applied method (A) on the persistent barcodes to obtain the third group of comparison features with a total number of TDA features being $16 \times L \times N$ for each set of time-delay parameters. Comparing with the first two TDA approaches, extracting TDA features from the transformed data achieved significantly higher classification performance. However, the three existing TDA approaches were all focused on individual channels, which means that the TDA features were extracted separately from each channel, ignoring the spatial correlation in the spatial-temporal data. Compared with the existing TDA approaches, our proposed TDA method shows superior performance with the highest accuracy and kappa value on all subjects.

Table 2. Comparison of different classifiers: classification accuracy on the training data with cross-validation (Data 1).

Subject ID	LASSO Logistic		LDA		SVM (Linear)	
	Acc	κ	Acc	κ	Acc	κ
1	0.745	0.49	0.845	0.69	0.865	0.73
2	0.756	0.513	0.799	0.599	0.904	0.808
3	0.724	0.448	0.729	0.459	0.809	0.619
4	0.752	0.504	0.742	0.484	0.869	0.738
5	0.730	0.46	0.823	0.645	0.891	0.781
6	0.750	0.5	0.799	0.598	0.857	0.714
7	0.786	0.573	0.768	0.535	0.908	0.815
8	0.771	0.543	0.803	0.606	0.849	0.698
9	0.748	0.495	0.80	0.6	0.865	0.73
Average	0.751	0.503	0.790	0.580	0.868	0.737
Subject ID	SVM (RBF)		KNN		RF	
	Acc	κ	Acc	κ	Acc	κ
1	0.878	0.755	0.838	0.675	0.958	0.915
2	0.872	0.744	0.811	0.623	0.953	0.906
3	0.838	0.675	0.810	0.62	0.912	0.825
4	0.876	0.753	0.872	0.744	0.945	0.89
5	0.896	0.791	0.771	0.543	0.945	0.89
6	0.869	0.739	0.795	0.59	0.945	0.89
7	0.899	0.798	0.789	0.578	0.924	0.848
8	0.840	0.68	0.716	0.431	0.834	0.668
9	0.863	0.726	0.834	0.669	0.934	0.868
Average	0.870	0.74	0.804	0.608	0.927	0.854

Table 3. Comparison of TDA features from literature: classification performance on the training data with cross-validation (Data 1).

Subject ID	EEG (A)		EEG (B)		HHT (A)		Proposed Features	
	Acc	κ	Acc	κ	Acc	κ	Acc	κ
1	0.546	0.093	0.534	0.068	0.788	0.575	0.958	0.915
2	0.528	0.055	0.585	0.17	0.631	0.263	0.953	0.906
3	0.530	0.06	0.543	0.085	0.598	0.195	0.912	0.825
4	0.715	0.43	0.536	0.073	0.926	0.853	0.945	0.89
5	0.721	0.443	0.551	0.103	0.759	0.518	0.945	0.89
6	0.523	0.045	0.510	0.02	0.616	0.233	0.945	0.89
7	0.588	0.175	0.555	0.11	0.7	0.4	0.924	0.848
8	0.724	0.448	0.649	0.298	0.801	0.603	0.834	0.668
9	0.618	0.235	0.506	0.013	0.76	0.52	0.934	0.868
Average	0.610	0.220	0.552	0.104	0.731	0.462	0.927	0.854

Table 4 summarizes the training performance of the proposed TDA approaches on all three datasets including the averaged performance and the standard deviation across all subjects. The training performance reported in the literature is also included in the table. For all three datasets, our proposed TDA approaches exhibit superior performance.

Table 4. Summary of classification performance on the training data with cross-validation.

	Data 1	Data 2	Data 3
Accuracy	0.927 (0.038)	0.955 (0.038)	0.987 (0.020)
Kappa	0.854 (0.076)	0.910 (0.076)	0.973 (0.041)
Literature			
Accuracy	0.843 (0.15) [26]	0.923 (0.04) [27]	0.961 (0.027) [17]

In addition to validating our approach on the training data and comparing it with existing TDA methods, we also assessed the proposed method on the testing data provided by the BCI competition and compared our results with those of top-performing teams on the leader board as well as those reported in the literature. Specifically, the classifier was re-trained on the entire training trials provided by the competition organizers. Subsequently, it was utilized to predict the label of testing trials for each subject. Table 5 summarizes the top three teams in the public leader board as well as the testing performance reported in the literature. Our testing performance not only exceeds that of the competition winner but also outperforms that reported in the literature. Moreover, our proposed method consistently demonstrates strong performance across all subjects with minimal variation, a notable contrast to other results that exhibit low kappa values, particularly for subjects 2 and 3.

Table 5. Test Performance of the BCI competition IV dataset IIb (Data 1) from leader board and literature.

Kappa	Mean	1	2	3	4	5	6	7	8	9
Leader Board										
Our Method	0.70	0.70	0.71	0.61	0.74	0.64	0.78	0.65	0.75	0.73
1st [28]	0.60	0.40	0.21	0.22	0.95	0.86	0.61	0.56	0.85	0.74
2nd	0.58	0.42	0.21	0.14	0.94	0.71	0.62	0.61	0.84	0.78
3rd	0.46	0.19	0.12	0.12	0.77	0.57	0.49	0.38	0.85	0.61

Table 5. Cont.

Kappa	Mean	1	2	3	4	5	6	7	8	9
Literature										
Ang et al. [29]	0.6	0.43	0.21	0.24	0.94	0.84	0.59	0.58	0.86	0.66
Yang et al. [30]	0.62	0.44	0.24	0.25	0.93	0.86	0.70	0.55	0.85	0.75
Kumar et al. [31]	0.56	0.55	0.21	0.01	0.99	0.66	0.53	0.72	0.77	0.61
Ang et al. [32]	0.61	0.36	0.17	0.26	0.96	0.87	0.67	0.56	0.86	0.75
Shahid et al. [33]	0.61	0.43	0.36	0.19	0.95	0.63	0.66	0.59	0.90	0.76

For Data 2, Table 6 summarizes the testing performance of the proposed TDA approach alongside the top-performing teams on the leader board and those reported in the literature. Our approach is ranked between second and third place. It is important to note that the competition does not require the same approach for all subjects. For example, the winning team [34] employed a common spatial pattern (CSP) for subjects al, aw, and ay, while a combination of CSP, AR, and LDA were utilized for subjects aa and av. Additionally, they utilized the testing trials during the training phase for subjects with a small number of training trials such as subject ay.

Overall, the proposed TDA based approach outperforms existing TDA approaches and demonstrates competitive performance when compared to other state-of-art methods.

Table 6. Test accuracy of the BCI competition III dataset IVa (Data 2) from leader board and literature.

Accuracy	Mean	aa	al	av	aw	ay
Leader Board						
1st [34]	0.947	0.955	1.00	0.806	1.00	0.976
2nd	0.874	0.893	0.982	0.765	0.924	0.806
Our Method	0.852	0.875	0.964	0.816	0.804	0.802
3rd	0.845	0.821	0.946	0.704	0.875	0.881
Literature						
Selim et al. [35]	0.85	0.866	1	0.668	0.906	0.81
Park and Chung [36]	0.845	0.741	1	0.678	0.901	0.893
Dai et al. [37]	0.792	0.681	0.939	0.685	0.884	0.749
Selim et al. [38]	0.788	0.696	0.893	0.592	0.888	0.869
Lotte and Guan [39]	0.786	0.723	0.964	0.602	0.777	0.865
Arvaneh et al. [40]	0.735	0.723	0.964	0.541	0.705	0.734
Belwafi et al. [41]	0.673	0.668	0.961	0.521	0.714	0.50
Herman et al. [42]	0.839	0.768	0.982	0.745	0.929	0.77
Cecchin et al. [43]	0.897	0.824	0.986	0.768	0.94	0.966
Padfield et al. [44]	0.864	0.813	1	0.653	0.933	0.921
Höller et al. [45]	0.861	0.795	1	0.735	0.893	0.885
Zhang et al. [46]	0.76	0.592	0.91	0.585	0.835	0.878
Attallah et al. [47]	0.935	0.922	0.994	0.799	0.989	0.97
Arvaneh et al. [48]	0.882	0.777	1	0.77	0.943	0.921
Lemm et al. [49]	0.736	0.795	0.929	0.526	0.915	0.516
Kee et al. [50]	0.835	0.744	0.985	0.708	0.905	0.834
Meng et al. [51]	0.86	0.83	1	0.735	0.821	0.915
Zhang et al. [52]	0.67	0.75	0.839	0.53	0.741	0.488
Miao et al. [53]	0.895	0.857	0.982	0.766	0.951	0.917

5. Conclusions

In this paper, we proposed a novel TDA-based approach to analyze scalp EEG data. One valuable feature of TDA is its capacity to interpret high-dimensional data with complex topological structures and extract meaningful information from the complex data. Compared to existing TDA approaches in the literature that extract topological features from each channel individually, we treat the multivariate time series in a Banach space and extract local and global topological features from all channels simultaneously. Compared to channel-wise TDA approaches in the literature, the proposed method requires significantly fewer computational resources and less time, making it more practical for interactive applications such as BCI. Our approach is also robust against inter-subject variability and the number of channels, maintaining stable performance across subjects in the three datasets with varying channel counts. One limitation of the proposed TDA approach is the selection of appropriate lengths of the time interval used in segmenting multivariate time series. Since there may not exist a universally optimal choice of interval lengths applicable to all datasets, we recommend tuning this parameter for each new dataset, considering the balance between capturing the local and global topological features of the data. Although the method was introduced in the context of scalp EEG, it can be easily generalized to any spatial and temporal data, making it applicable to various fields such as computer vision, remote sensing, climate change, and more.

Author Contributions: Conceptualization, J.Z. and Z.F.; methodology, J.Z. and Z.F.; software, J.Z. and Z.F.; validation, J.Z. and Z.F.; formal analysis, J.Z. and Z.F.; investigation, J.Z. and Z.F.; resources, J.Z.; data curation, J.Z. and A.D.E.; writing—original draft preparation, J.Z. and Z.F.; writing—review and editing, J.Z., Z.F. and A.D.E.; visualization, J.Z. and Z.F.; supervision, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is based upon work supported by the National Science Foundation under Grant No. 2153492 and the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR003096. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability Statement: All data used in this study are publicly available. Data 1 can be downloaded at <https://www.bbc.de/competition/iv/> (accessed on 1 April 2024). Data 2 can be downloaded at <https://www.bbc.de/competition/iii/index.html> (accessed on 1 April 2024). Data 3 can be downloaded at <https://osf.io/3vxkn/> (accessed on 1 April 2024).

Acknowledgments: The authors would like to thank anonymous referees, an associate editor, and the editor for their constructive comments that improved the quality of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Carlsson, G. Topology and data. *Bull. Am. Math. Soc.* **2009**, *46*, 255–308. [CrossRef]
2. Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. *Discret. Comput. Geom.* **2002**, *28*, 511–533. [CrossRef]
3. Wang, Y.; Behroozmand, R.; Johnson, L.P.; Bonilha, L.; Fridriksson, J. Topological signal processing in neuroimaging studies. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops), Iowa City, IA, USA, 3–7 April 2020; pp. 1–4.
4. Altındış, F.; Yılmaz, B.; Borisenok, S.; İçöz, K. Use of topological data analysis in motor intention based brain–computer interfaces. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1695–1699.
5. Altındış, F.; Yılmaz, B.; Borisenok, S.; İçöz, K. Parameter investigation of topological data analysis for EEG signals. *Biomed. Signal Process. Control* **2021**, *63*, 102196. [CrossRef]
6. Bourakna, A.E.Y.; Chung, M.K.; Ombao, H. Topological Data Analysis for Multivariate Time Series Data. *arXiv* **2022**, arXiv:2204.13799.
7. Yamanashi, T.; Kajitani, M.; Iwata, M.; Crutchley, K.J.; Marra, P.; Malicoat, J.R.; Williams, J.C.; Leyden, L.R.; Long, H.; Lo, D.; et al. Topological data analysis (TDA) enhances bispectral EEG (BSEEG) algorithm for detection of delirium. *Sci. Rep.* **2021**, *11*, 304. [CrossRef] [PubMed]

8. Yan, Y.; Wu, X.; Li, C.; He, Y.; Zhang, Z.; Li, H.; Li, A.; Wang, L. Topological EEG nonlinear dynamics analysis for emotion recognition. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *15*, 625–638. [[CrossRef](#)]
9. Wang, Y.; Ombao, H.; Chung, M.K. Topological data analysis of single-trial electroencephalographic signals. *Ann. Appl. Stat.* **2018**, *12*, 1506. [[CrossRef](#)] [[PubMed](#)]
10. Zheng, J.; Feng, Z.; Li, Y.; Liang, F.; Cao, X.; Ge, L. Topological Data Analysis for Scalp EEG Signal Processing. In Proceedings of the 2023 8th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 8–10 July 2023; pp. 549–553. [[CrossRef](#)]
11. Xu, X.; Drougard, N.; Roy, R.N. Topological Data Analysis as a New Tool for EEG Processing. *Front. Neurosci.* **2021**, *15*, 761703. [[CrossRef](#)] [[PubMed](#)]
12. Takens, F.; Rand, D.; Young, L.S. *Detecting Strange Attractors in Turbulence*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 898.
13. Sauer, T.; Yorke, J.A.; Casdagli, M. Embedology. *J. Stat. Phys.* **1991**, *65*, 579–616. [[CrossRef](#)]
14. Tangermann, M.; Müller, K.R.; Aertsen, A.; Birbaumer, N.; Braun, C.; Brunner, C.; Leeb, R.; Mehring, C.; Miller, K.J.; Mueller-Putz, G.; et al. Review of the BCI competition IV. *Front. Neurosci.* **2012**, *6*, 55. [[CrossRef](#)]
15. Blankertz, B.; Müller, K.R.; Krusienski, D.J.; Schalk, G.; Wolpaw, J.R.; Schlogl, A.; Pfurtscheller, G.; Millan, J.R.; Schroder, M.; Birbaumer, N. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2006**, *14*, 153–159. [[CrossRef](#)] [[PubMed](#)]
16. Liang, M.; Zheng, J.; Isham, E.; Ekstrom, A. Common and distinct roles of frontal midline theta and occipital alpha oscillations in coding temporal intervals and spatial distances. *J. Cogn. Neurosci.* **2021**, *33*, 2311–2327. [[CrossRef](#)] [[PubMed](#)]
17. Zheng, J.; Liang, M.; Sinha, S.; Ge, L.; Yu, W.; Ekstrom, A.; Hsieh, F. Time-frequency analysis of scalp EEG with Brain-computer Interface transform and deep learning. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 1549–1559. [[CrossRef](#)] [[PubMed](#)]
18. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London. Ser. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
19. Chazal, F.; Michel, B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front. Artif. Intell.* **2021**, *4*, 108. [[CrossRef](#)] [[PubMed](#)]
20. Bakke Bjerkevik, H. On the Stability of Interval Decomposable Persistence Modules. *Discrete Comput. Geom.* **2021**, *66*, 92–121. [[CrossRef](#)]
21. Chazal, F.; de Silva, V.; Glisse, M.; Oudot, S. *The Structure and Stability of Persistence Modules*; Springer: Berlin/Heidelberg, Germany, 2016.
22. Tauzin, G.; Lupo, U.; Tunstall, L.; Pérez, J.B.; Caorsi, M.; Medina-Mardones, A.; Dassatti, A.; Hess, K. giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration. *arXiv* **2020**, arXiv:2004.02551.
23. Liebert, W.; Pawelzik, K.; Schuster, H. Optimal embeddings of chaotic attractors. *Europhys. Lett.* **1991**, *14*, 521–526. [[CrossRef](#)]
24. Kennel, M.; Brown, R.; Abarbanel, H. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **1992**, *45*, 3403–3411. [[CrossRef](#)]
25. Adams, H.; Tausz, A.; Vejdemo-Johansson, M. javaPlex: A research software package for persistent (co)homology. *Lect. Notes Comput. Sci.* **2014**, *8592*, 129–136.
26. Zhang, Y.; Nam, C.S.; Zhou, G.; Jin, J.; Wang, X.; Cichocki, A. Temporally Constrained Sparse Group Spatial Patterns for Motor Imagery BCI. *IEEE Trans. Cybern.* **2019**, *49*, 3322–3332. [[CrossRef](#)] [[PubMed](#)]
27. Park, Y.; Chung, W. Frequency-Optimized Local Region Common Spatial Pattern Approach for Motor Imagery Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1378–1388. [[CrossRef](#)] [[PubMed](#)]
28. Ang, K.K.; Chin, Z.Y.; Zhang, H.; Guan, C. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 1–6 June 2008; pp. 2390–2397.
29. Ang, K.K.; Chin, Z.Y.; Zhang, H.; Guan, C. Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs. *Pattern Recognit.* **2012**, *45*, 2137–2144. [[CrossRef](#)]
30. Yang, Y.; Chevallier, S.; Wiart, J.; Bloch, I. Time-frequency selection in two bipolar channels for improving the classification of motor imagery EEG. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August–1 September 2012; pp. 2744–2747. [[CrossRef](#)]
31. Kumar, S.; Mamun, K.; Sharma, A. CSP-TSM: Optimizing the performance of Riemannian tangent space mapping using common spatial pattern for MI-BCI. *Comput. Biol. Med.* **2017**, *91*, 231–242. [[CrossRef](#)] [[PubMed](#)]
32. Ang, K.K.; Chin, Z.Y.; Zhang, H.; Guan, C. Robust filter bank common spatial pattern (RFBCSP) in motor-imagery-based brain-computer interface. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MI, USA, 3–6 September 2009; pp. 578–581. [[CrossRef](#)]
33. Shahid, S.; Sinha, R.K.; Prasad, G. A bispectrum approach to feature extraction for a motor imagery based brain-computer interfacing system. In Proceedings of the 2010 18th European Signal Processing Conference, Aalborg, Denmark, 23–27 August 2010; pp. 1831–1835.

34. Rashid, M.; Sulaiman, N.; PP Abdul Majeed, A.; Musa, R.M.; Ab Nasir, A.F.; Bari, B.S.; Khatun, S. Current status, challenges, and possible solutions of EEG-based brain–computer interface: A comprehensive review. *Front. Neurobot.* **2020**, *14*, 515104. [[CrossRef](#)] [[PubMed](#)]
35. Selim, S.; Tantawi, M.M.; Shedeed, H.A.; Badr, A. A csp\am-ba-svm approach for motor imagery bci system. *IEEE Access* **2018**, *6*, 49192–49208. [[CrossRef](#)]
36. Park, Y.; Chung, W. BCI classification using locally generated CSP features. In Proceedings of the 2018 6th International Conference on Vietoris–Rips (BCI), Resort, Republic of Korea, 15–17 January 2018; pp. 1–4.
37. Dai, M.; Zheng, D.; Liu, S.; Zhang, P. Transfer kernel common spatial patterns for motor imagery brain–computer interface classification. *Comput. Math. Methods Med.* **2018**, *2018*, 9871603. [[CrossRef](#)] [[PubMed](#)]
38. Selim, S.; Tantawi, M.; Shedeed, H.; Badr, A. Reducing execution time for real-time motor imagery based BCI systems. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 555–565.
39. Lotte, F.; Guan, C. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* **2010**, *58*, 355–362. [[CrossRef](#)] [[PubMed](#)]
40. Arvaneh, M.; Guan, C.; Ang, K.K.; Quek, H.C. Spatially sparsed common spatial pattern to improve BCI performance. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 2412–2415.
41. Belwafi, K.; Romain, O.; Gannouni, S.; Ghaffari, F.; Djemal, R.; Ouni, B. An embedded implementation based on adaptive filter bank for brain–computer interface systems. *J. Neurosci. Methods* **2018**, *305*, 1–16. [[CrossRef](#)]
42. Herman, P.; Prasad, G.; McGinnity, T.M.; Coyle, D. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2008**, *16*, 317–326. [[CrossRef](#)]
43. Cecchin, T.; Ranta, R.; Koessler, L.; Caspary, O.; Vespignani, H.; Maillard, L. Seizure lateralization in scalp EEG using Hjorth parameters. *Clin. Neurophysiol.* **2010**, *121*, 290–300. [[CrossRef](#)] [[PubMed](#)]
44. Padfield, N.; Zabalza, J.; Zhao, H.; Masero, V.; Ren, J. EEG-based brain–computer interfaces using motor-imagery: Techniques and challenges. *Sensors* **2019**, *19*, 1423. [[CrossRef](#)] [[PubMed](#)]
45. Höller, Y.; Thomschewski, A.; Uhl, A.; Bathke, A.C.; Nardone, R.; Leis, S.; Trinka, E.; Höller, P. HD-EEG based classification of motor-imagery related activity in patients with spinal cord injury. *Front. Neurol.* **2018**, *9*, 955. [[CrossRef](#)]
46. Zhang, Y.; Liu, B.; Ji, X.; Huang, D. Classification of EEG signals based on autoregressive model and wavelet packet decomposition. *Neural Process. Lett.* **2017**, *45*, 365–378. [[CrossRef](#)]
47. Attallah, O.; Abougharbia, J.; Tamazin, M.; Nasser, A.A. A BCI system based on motor imagery for assisting people with motor deficiencies in the limbs. *Brain Sci.* **2020**, *10*, 864. [[CrossRef](#)] [[PubMed](#)]
48. Arvaneh, M.; Guan, C.; Ang, K.K.; Quek, C. Mutual information-based optimization of sparse spatio-spectral filters in brain–computer interface. *Neural Comput. Appl.* **2014**, *25*, 625–634. [[CrossRef](#)]
49. Lemm, S.; Blankertz, B.; Curio, G.; Muller, K.R. Spatio-spectral filters for improving the classification of single trial EEG. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 1541–1548. [[CrossRef](#)]
50. Kee, C.Y.; Ponnambalam, S.G.; Loo, C.K. Multi-objective genetic algorithm as channel selection method for P300 and motor imagery data set. *Neurocomputing* **2015**, *161*, 120–131. [[CrossRef](#)]
51. Meng, J.; Huang, G.; Zhang, D.; Zhu, X. Optimizing spatial spectral patterns jointly with channel configuration for brain–computer interface. *Neurocomputing* **2013**, *104*, 115–126. [[CrossRef](#)]
52. Zhang, H.; Chin, Z.Y.; Ang, K.K.; Guan, C.; Wang, C. Optimum spatio-spectral filtering network for brain–computer interface. *IEEE Trans. Neural Netw.* **2010**, *22*, 52–63. [[CrossRef](#)]
53. Miao, M.; Wang, A.; Liu, F. Application of artificial bee colony algorithm in feature optimization for motor imagery EEG classification. *Neural Comput. Appl.* **2018**, *30*, 3677–3691. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.