IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 4, 2020, accepted May 19, 2020, date of publication May 22, 2020, date of current version June 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2996667

Classification of Neurodegenerative Diseases via Topological Motion Analysis–A Comparison Study for Multiple Gait Fluctuations

YAN YAN[1,2], (Member, IEEE), OLATUNJI MUMINI OMISORE[1,2], (Member, IEEE), YU-CHENG XUE[1], HUI-HUI LI[1], (Member, IEEE), QIU-HUA LIU[1], ZE-DONG NIE[1,2], (Member, IEEE), JIANPING FAN[1,2], AND LEI WANG[1,2], (Member, IEEE)

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2]Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China

Corresponding authors: Jianping Fan (jp.fan@siat.ac.cn) and Lei Wang (wang.lei@siat.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71531004, in part by the Shenzhen Innovation Project under Grant KQJSCX20170731163142116, and in part by the Enhancement Project for Shenzhen Biomedical Electronics Technology Public Service Platform.

ABSTRACT Neurodegenerative diseases are common progressive nervous system disorders that show intricate clinical patterns. The gait fluctuations reflect the physiology and pathologic alterations in the locomotor control system. Using gait fluctuations for disease state evaluation is an essential way for clinical trials and healthcare monitoring. The classification of gait fluctuations helps improve the life quality and enhance clinical diagnosis ability in neuro-degenerative patients. In this work, we firstly embed the time series of multiple gait fluctuations into the phase space. Then we use persistent homology to extract the topological signatures of barcodes. Together with a random forest classifier, we proposed a topological motion analysis (TMA) framework to analyze the gait fluctuations. Further, we proposed a comprehensive comparison study using the TMA framework in the neuro-degenerative classification tasks for stance-, stride-, and swing-based gait fluctuations. In the tasks of comparing amyotrophic lateral sclerosis (ALS), Huntington's disease (HD), and Parkinson's disease (PD) to the healthy control (HC) group, the best-achieved AUC scores were 0.9135, 0.9906 and 0.9667 respectively, which show the effectiveness of TMA framework. In summary, our study proposed a TMA framework towards gait fluctuations classification in the neuro-degenerative analysis tasks. The proposed method shows promising clinical application value in earlier interventions and state monitoring for neurodegenerative patients.

INDEX TERMS Topological motion analysis, gait fluctuations analysis, neurodegenerative disease analysis, feature extraction, persistence landscape, biomedical signal processing.

I. INTRODUCTION

Neurodegenerative disease is a common progressive disorder of the nervous system, which might lead to serious liver problems like the tremor of limbs, jaw or face, and stiffness of slowing of movement [1]. This disease usually emerges gradually and causes movement problems and difficulty walking, including amyotrophic lateral sclerosis, Huntington's disease, and Parkinson's disease. Amyotrophic lateral sclerosis affects upper motor neurons and lower motor neurons, which lead to symptoms of muscle weakness, atrophy, fasciculation, and cramping [2]. Huntington's disease is a hereditary disorder that leads by the harm from the gradual loss of motor neurons in basal ganglia and cerebral cortex, which manifested as unwanted movements, behavioral and psychiatric disturbances, and dementia [3]. Parkinson's disease occurs due to the lack of dopamine production as a result of the deterioration of neurons in basal ganglia, which leads to a progressive movement disorder like hands and legs tremor, movement slowness, rigidity, or postural instability [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

The disorder patterns lead by the three kinds of neurodegenerative diseases are quite different, which brought challenges for early detection and preventing of progression.

While the gait abnormality as a deviation of walking gait fluctuation may reflect different disorder patterns, gait fluctuations analysis is an essential tool for assessing neurodegenerative disease [5]–[7]. The gait information describes the human walking states, which exhibits periodic patterns termed as gait cycle [8]. Each gait cycle has a sequence of ordered gait events, which occur at specific temporal locations and thereby referred to as temporal gait parameters, as illustrated in Figure 1. In this study, the stride, stance, and swing parameters are involved for human walking pattern analysis and distinguishing abnormal neurodegenerative class. Hausdorff *et al.* [9]–[11] studied the stride interval time series of the gait in subjects with Huntington's disease and also the healthy elderly subjects in compared to control subjects, which is supposed to be the earliest work based on the gait fluctuations-based disease analysis. Their results showed that stride interval variations are more random in subjects with Huntington's disease and elderly subjects than in control subjects. Further investigations for the discussion on gait fluctuations analysis toward disease state analysis are [5], [12], [13].

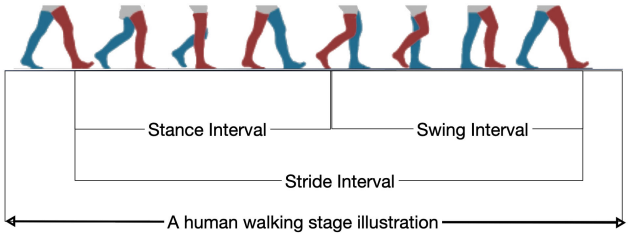


FIGURE 1. The example illustrations for the gait fluctuations of stride interval for each group.

Inspired by their work, Kamruzzaman uses two basic temporal-spatial gait parameters (stride length and cadence) as input features and support vector machine method to analyze the cerebral palsy gait [14]. [15] reported multiple regression normalization strategies that incorporated patient physical properties and self-selected speed for Parkinson's Disease Gait analysis. In [16], Wu used a nonparametric Paren-window method to estimate the probability density functions of stride interval and its sub-phases: swing interval and stance interval, with the statistical analysis of gait rhythm. Frequency range distribution [17], tensor decomposition [18], texture-based images with fuzzy recurrence plots [19] were proposed for the gait dynamics analysis for neurodegenerative disease classification. Studying the dynamics of gait patterns in neurodegenerative disease to diagnose the severity could lead to the fall prediction, treatment, and rehabilitation strategies improvement.

Recently, the techniques based on ample dataset training such as deep neural networks [20], [21], and ensemble learning methods [22] accomplished great success in different applications. However, in the small-sample-size cases, the performances are mainly decided by [23]–[25], how the designed feature reveal the characteristics of the pathological disease path and human body state dynamics. The nonlinear time series analysis provided insights for feature designing based on the dynamical system theory, such as recurrence plot [26], Poincare plot [27], Lyapunov exponents [28], detrended fluctuation analysis (DFA) [29], approximate entropy, and sample entropy-based analysis [30]. In nonlinear time series analysis, the phase spaces represented by data point clouds reveal important information of the underlying dynamical system.

Meanwhile, the topological data analysis (TDA) provided powerful tools based on algebra topology theory proposed for data point cloud analysis [31], [32]. The TDA techniques had contributed in lots of recognition and classification tasks: 3-D object recognition [33]; protein folding analysis [34] and complex data analysis & visualization [35]–[37]. In nonlinear time series analysis problems, TDA also provided alternative viewpoints for signal analysis [38]–[41].

We were motivated by the nonlinear analysis and topological analysis techniques. Thus, we proposed a topological motion analysis (TMA) framework that combines the nonlinear motion analysis and TDA techniques. The proposed work could contribute to gait fluctuations analysis and further applications in neurodegenerative disease classification.

Contributions:

1) We proposed a topological motion analysis framework toward neurodegenerative classification tasks. The TMA of gait fluctuations is supposed to be a novel technique to distinguish disease states.
2) We performed a comprehensive comparison study on multiple gait fluctuations, including stride-interval, stance-interval, and swing-interval-based time series from each foot.
3) Using an open dataset of gait neuro-degenerative fluctuation analysis with the TMA framework, achieved results show promising ability in state recognition and neurodegenerative disease classification.

Outline of the Paper:

1) Section 2 provides the proposed framework and the corresponding background on phase space reconstruction, and TDA technique: the simplicial complex theory, the filtration building process, and how the topological features are extracted.
2) Section 3 introduces the data source and subject information, and the experimental settings for the neuro-degenerative classification tasks.
3) In Section 4, we illustrated the performance of neurodegenerative classification tasks with the proposed framework with different parameters.
4) Then, in Section 5, we discussed the results and reviewed the related work of neurodegenerative disease classification. Finally, Section 6 concludes the paper and illustrates the limitations and future directions.

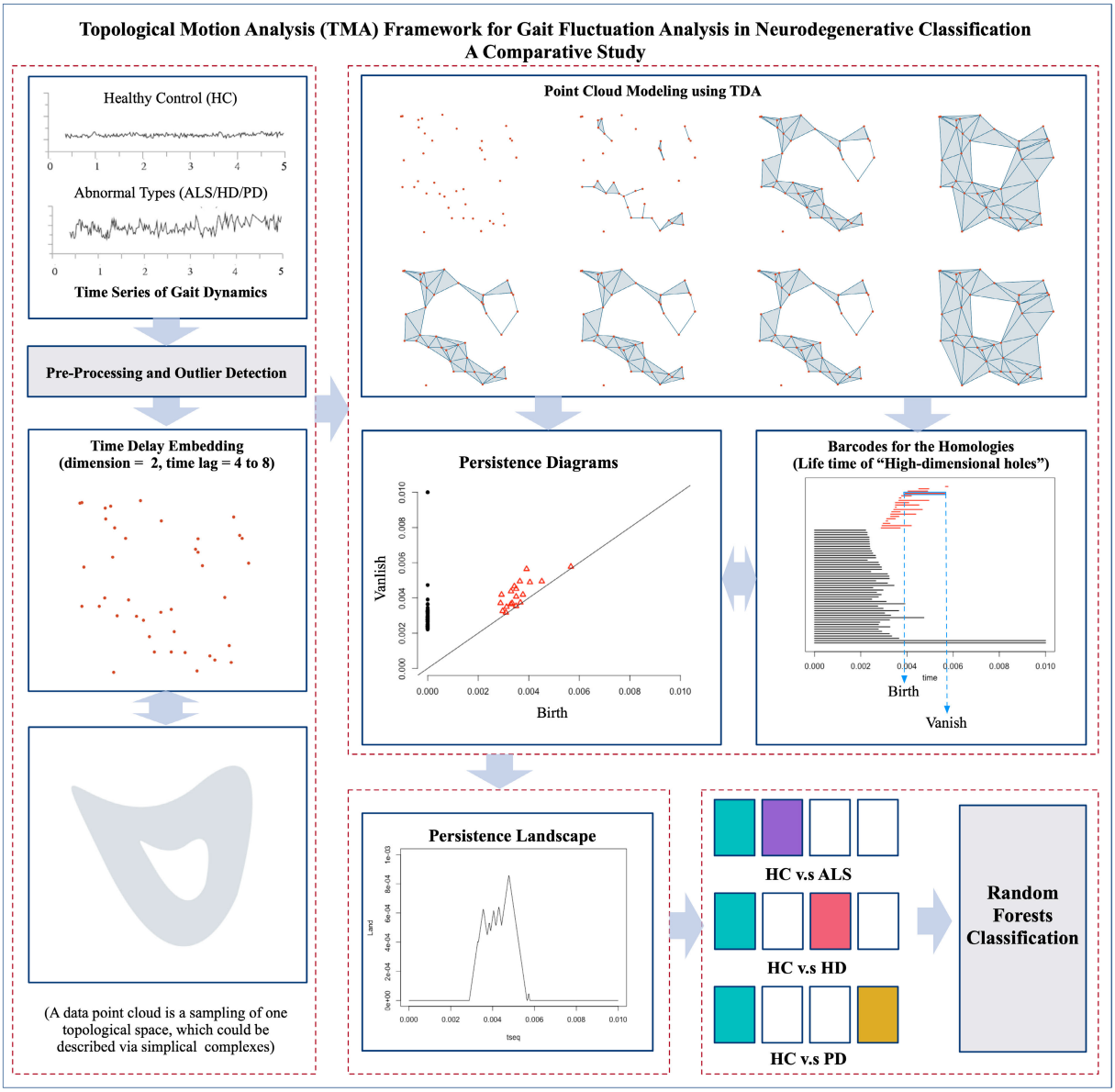


FIGURE 2. The proposed TMA framework and neuro-degenerative classification application.

II. METHOD

A. PROPOSED FRAMEWORK

As Figure 2 illustrated, the proposed work includes the following stages:

1) The time series of multiple gait fluctuations are processed with an outlier detection procedure, the preprocessing of the time series is necessary.
2) The processed time series are embedded into data point clouds, in which the dimension adopted is 2, and the lag parameters are from the set of {4, 5, 6, 7, 8}. The time series samples are transformed into point clouds. Each point cloud represents one topological space.
3) To model the corresponding topological spaces, the data points are adopted to construct simplicial complexes with different radius parameter. As one gradually increases the radius from 0 to ∞, a sequence of simplicial complexes is generated, termed as filtration. The filtration of the point cloud can be used for modeling the topological space.
4) The barcodes are extracted as features of the space, which illustrates the duration for the topological objects.
5) The persistence diagram and persistence landscape are generated based on the barcodes, which is more favorable for statistical analysis and machine learning tasks.
6) With the former transformations, the time series set is transformed into a persistence landscape set, from which the HC v.s ALS, HC v.s HD, and HC v.s PD binary classification tasks are performed with leave-one-out cross-validation. The random forest classifier is used in the classification tasks.

We briefly describe these stages with corresponding technical details in the following sections and involve the more in-depth mathematical descriptions in the appendix sections with provided references.

B. TIME-DELAY EMBEDDING

The time-delay embedding technique is a typical method for phase space reconstruction. In the proposed TMA framework, i.e., the time series are converted into point clouds. Reference [42] proposed a standard description of time-delay embedding. Time-delay embeddings aim to reconstruct the state and dynamics of unknown dynamical systems from the observations (i.e., the measured time series). Phase space reconstruction can embed the time series into points in abstract space. The Takens' theorem [43] tells that if one variable of the system is observable, then a complete system can be reconstructed from the observations. The reconstructed phase spaces reveal the dynamical system's character, which can act as the extracted features in recognition applications [44].

In this study, the stride, stance, and swing time series are observations for each subjects' body systems. Each time series from one subject embedded into the corresponding data point cloud reveals the properties for the body dynamical system. Mathematically, suppose the gait fluctuations interval sequence $f(n), n \in \mathbb{Z}^+$, in which n is the sampling index. Based on the Takens' theorem, let $\tau \in \mathbb{Z}^+$ be the delay step and the dimension as $d \in \mathbb{Z}^+$, then the time delay embedding process can be illustrated as:

$$DE(f, t; \tau, d) = \{f(t), f(t+\tau), \cdots, f(t+(d-1)\tau)\} \quad (1)$$

Then the data point clouds are generated by this phase space reconstruction method. A point cloud is a sampling representation of its underlying topological space of the real dynamical system. A good approximation of the real dynamics needs appropriate embedding parameters of d and τ. Lots of work had been done for searching for optimal embedding dimension d and time-delay lag parameters τ, refer [45] for details. However, for the classification task, we use a practical choice for embedded dimension $d = 2$ to reduce the computation cost for topological feature generation. The time lag parameter set is set as $\tau \in \{4, 5, 6, 7, 8\}$.

C. TOPOLOGICAL GAIT FLUCTUATION ANALYSIS

The core of the TMA framework is the topological feature extraction. We adopt the TDA techniques in the gait fluctuations point cloud analysis to build the feature set for further classification. In this section, we briefly introduce the process of topological feature extraction. We explain the corresponding mathematical explanations in the appendix and references further.

1) TOPOLOGICAL SPACE AND HOMOLOGY

In the phase space, the point clouds are supposed to be lying on some abstract space, i.e., one point cloud is a sampling of a topological space. Once the time series are embedded into point clouds, the time series classification problem is converted into point clouds classification problem. We try to study the characteristics of the corresponding topological spaces in which the point clouds are lying on to distinguish different time series classes.

Two shapes in the space can be distinguished by examining the structures like connected components and holes. Consider a disk and a circle, and the corresponding topological spaces are different since the circle has a hole, while the disk contains a solid connected component. The topological objects of connected components and holes are called 0-dimensional homology and 1-dimensional homology, respectively. The higher dimensional homologies can be understood as 'high-dimensional holes.' Thus one topological space can be described as a set of such 'high-dimensional holes.' Informally, for a topological space $\mathbb{X}$ there is one corresponding set of homologies, i.e., homology groups:

$$H_0(\mathbb{X}), H_1(\mathbb{X}), H_2(\mathbb{X}), \ldots \quad (2)$$

We explore a dynamical process based on homologies with diverse settings of the topological space in the following section, which forms the topological feature set we use in a later stage.

2) PERSISTENT HOMOLOGY AND BARCODES

With a given data point, the underlying topological space shows different characteristics with different resolutions. The persistent features based on the point clouds are detected over a range of spatial scales, which better represent the underlying space's essential features compared to the noise [31]. Simplicial complexes are discrete objects used to study the topological spaces, such as points, line segments, triangles, or even higher n-dimensional objects. As Figure 3 illustrated, we consider a resolution parameter of ϵ (the radius of a unit ball). Consider the process of increasing the resolution parameter to the values of $\{\epsilon_0, \epsilon_1, \epsilon_2, \ldots\}$, each value corresponds to one simplicial complex. For example, the original embedded point cloud ($\epsilon = 0$), the two-hole simplicial complex ($\epsilon = \epsilon_3$), and the one connected component when $\epsilon = \epsilon_8$. We can see that in this ϵ growing process, different topological objects appear and disappear, as the 0-dimensional homology (connected objects), and 1-dimensional homology (holes) in Figure 3. The set contains the simplicial complexes generated with different ϵ values is a nested sequence (termed as filtration).

The topological space characteristics are illustrated by the durations of these topological objects, which termed as barcodes. In Figure 3, one small hole appears when $\epsilon = \epsilon_3$ and lasts to $\epsilon = \epsilon_5$ and finally disappear at $\epsilon = \epsilon_6$, while another bigger hole lasts from $\epsilon = \epsilon_4$ to $\epsilon = \epsilon_8$. The two hole-based objects with the birth-vanish (or birth-death) period are the two red bars in the barcodes plot (bottom left of Figure 3), i.e., the two barcodes represent the 1-dimensional persistent homologies. Meanwhile, the dark bars represent the lifetime periods of the 0-dimensional homologies. Thus, we use the birth time of b_α and vanish (death) time of d_α to

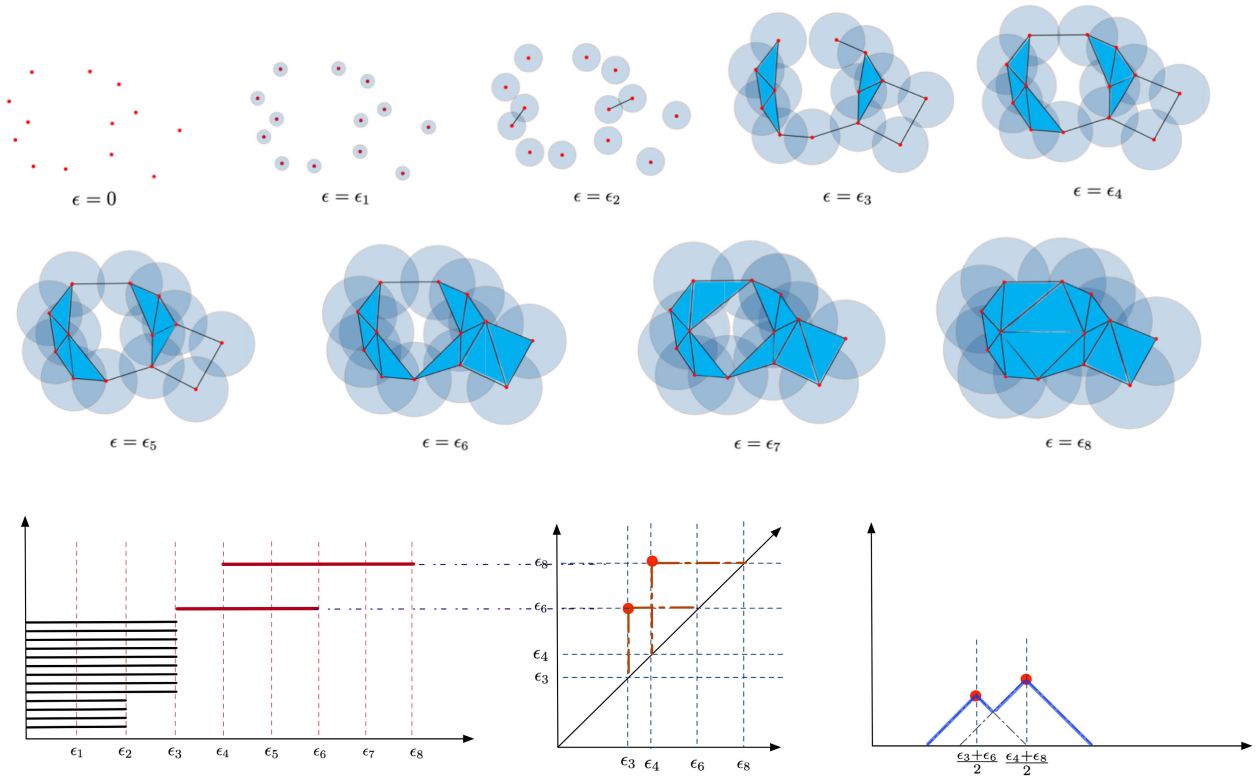


FIGURE 3. The point cloud modeling using simplicial complexes for topological feature extraction: (top) the topological space illustrations using different radius parameters, with the corresponding simplicial complexes; (bottom left) the barcodes for the topological objects, red bars for holes while darks bars for connected components; (bottom middle) the persistence diagram corresponds to the barcodes of topological holes; (bottom right) the persistence landscapes illustration, which is favorable for statistical analysis.

track the homology of α for topological space representation. Mathematically, the barcodes of the corresponding topological is a set of $\{(b_1, d_1), (b_2, d_2), \ldots, (b_m, d_m)\}$.

In summary, the topological space characteristics are described by points sampled from some $\mathbb{R}_n$ with neighbors. With which a graph built for neighbor relations study. Using the graph vertices and edges, we use the geometrical objects of simplexes and simplicial complexes as the basic building blocks to describe the way how to glue a space (a triangulation of a topological space).

There are different ways to build the complex to approximate the topological space, such as Vietoris-Rips complex [46], [47], Graph-induced complex [48], and Sparsified Čech complex [49]. In this work, we use the Vietoris-Rips complex in which sometimes was called Rips complex for its simplicity, a fast construction algorithm of the Rips complex using the point cloud can be referred from [50]. Since the embedding process is two-dimensional, so only the points, edges, triangles are applied. The formal definition of a simplicial complex, filtration, and mathematical description of the process can be referred from the appendix contents and formerly mentioned works of literature.

D. PERSISTENCE LANDSCAPE FEATURE GENERATION

1) PERSISTENCE DIAGRAM

The persistence diagrams provide a concise description of the topological changes over the data point cloud [38]. From the above process, we get the barcodes from the point cloud with a set of $\{(b_1, d_1), (b_2, d_2), \ldots, (b_m, d_m)\}$. Now we make a simple transformation with the barcodes by setting the birth time as the horizontal axis while the vanishing time as the vertical axis (bottom middle of Figure 3). Then the barcodes of the topological space are converted into the points on the persistence diagram. We mainly focus on the information of 1-dimensional homology, which corresponds to the red bars in the barcodes plot and the red points in the persistence diagram, as Figure 3 illustrates.

Persistence diagrams represent different topological spaces from the point cloud. By distinguishing the diagrams' dissimilarities, we can tell the differences to classify the point clouds from different neuro-degenerative gait fluctuations time series. Typical methods of using persistence diagrams for classification are Wasserstein distances-based, Bottleneck distances-based, or topological kernels [34]. In this study, we use the persistence landscape as the topological feature, based on the Wasserstein distance.

2) PERSISTENCE LANDSCAPE

An intuition of persistence landscapes is a rotation of the persistence diagram plot (Figure 3 bottom right). Mathematically, the persistence diagram P_k encoded from the k-dimensional homology α information in all scales. As last section described, the homology α was born at b_α

and vanish at d_α which make a pair (b_α, d_α) and pair set $\{(b_1, d_1), (b_2, d_2), \ldots (b_m, d_m)\}$. We use $z_\alpha \in \mathbb{R}$ to denote each pair and consider such pairs as a point.

The Wasserstein distance of the persistence diagram space as:

$$W_p(P_k^1, P_k^2) = inf_\phi [\sum_{q \in P_k^1} ||x - \phi(x)||_\infty^p]^{\frac{1}{p}} \tag{3}$$

The equation is termed as the p-th Wasserstein distance, when $p = \infty$ the metric is known as Bottleneck distance. Persistence landscapes are based on the definition of the Wasserstein metric, which is suitable for statistical analysis [51].

For each birth-death point $(b_\alpha, d_\alpha) \in P_k$, a piecewise linear function:

$$f(b_\alpha, d_\alpha) = \begin{cases} x - b_\alpha, & if\ x \in \left(b_\alpha, \dfrac{b_\alpha + d_\alpha}{2}\right] \\ -x + d_\alpha, & if\ x \in \left(\dfrac{b_\alpha + d_\alpha}{2}, d_\alpha\right) \\ 0, & if\ x \notin (b_\alpha, d_\alpha) \end{cases}. \tag{4}$$

with which a sequence of functions λ can be given by:

$$\lambda_k(x) = k - max\{f_{(b_\alpha, d_\alpha)}(x) | (b_\alpha, d_\alpha) \in P_k\} \tag{5}$$

where the k-max denotes the k^{th} largest value of a function.

The persistence landscape lies in a vector space. It is easy to combine with tools from statistics and machine learning [51]–[53]. The persistence landscapes are designed features based on the barcodes, which represent the gait fluctuations for our neurodegenerative analysis task. We use the persistence landscape as the features for the classification task.

E. PATTERN RECOGNITION

In the previous sections, we extract the barcode information for topological feature generation. We use the persistence landscapes as the features for distinguishing different diseases. In this work, we adopt the random forests as the classifier for the neurodegenerative diseases classification task.

A random forest classifier is often used as one nonparametric method to build an ensemble model of decision trees from random subsets of features from the training set. Random forest models have shown excellent performance in classification problems, mainly when the feature set contains irrelevant features or noise. In the applications, when the number of features is much more than the number of instances, random forests-based models show good classification ability. In this work, the topological features based on barcodes of abstract objects are not significantly related. Meanwhile, the size of the feature set is much larger than the number of instances in each task. So, we choose the random forest classifier as the classifier for the pattern recognition tasks.

Random forests classifier is an ensemble learning method where a large number of decision trees are exploited [54]. The decision trees separate the data samples to reduce the entropy of the dataset. In this work, the analysis tasks are binary classification problems, i.e., the output of the classifier $\mathbf{Y} = \{0, 1\}$ based on the probabilities. Given the vector-based topological features $\mathbf{X}$, the classifier with rule m_n gives the corresponding probability of error [55]:

$$L(m_n) = P[m_n(\mathbf{X} \neq)] \underset{n \to \infty}{\longrightarrow} L^* \tag{6}$$

where L^* is the error of Bayes classifier:

$$m^*(\mathbf{x}) = \begin{cases} 1 & \text{if } P[\mathbf{Y} = 1 | \mathbf{X} = \mathbf{x}] > P[\mathbf{Y} = 0 | \mathbf{X} = \mathbf{x}] \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The RF classifier output is obtained via a majority voting using classification trees using the involved decision trees. The Gini impurity parameter is used for the optimization of random forest classifier parameter training and testing [55]. The optimal rules of the forests built with the decision tree's set are achieved with the optimization process.

III. MATERIAL AND EXPERIMENTS

A. DATASET DESCRIPTION AND PRE-PROCESSING

In this study, we adopt the Gait Dynamics in Neuro-Degenerative Disease Dataset [12] from the Physionet database [56] as the validation dataset. The dataset was proposed for understanding the pathophysiology of neurodegenerative diseases. The dataset includes 16 healthy control (HC) subjects, 15 patients with Parkinson's disease (PD), 20 Huntington's disease HD, and 13 subjects with amyotrophic lateral sclerosis (ALS).

The HC group consists of subjects with age range 20–74, the ALS group consists of subjects with age range 36–70, the HD group consists of subjects with age range 29–71, and the PD group consists of subjects with age range 44–80. All the subjects were instructed to walk at their average pace along a 77-meter long hallway for 5 min. The raw data of the database were pressure signals obtained using force-sensitive resistors. The stride-to-stride measurements of footfall contact times have been derived from the signals from the force-sensitive resistors, which includes time series of left-foot-stride intervals, right-foot-stride intervals, left-foot-swing intervals, right-foot-swing intervals, left-foot-stance intervals, right-foot-stance intervals, and double support intervals. Also, the left/right swing interval percentages of stride, left/right stance interval percentages of stride, and double support interval percentages of stride are included.

In this study, we consider the six gait interval time series: left-stride-interval time series, right-stride-interval time series, left-swing-interval time series, right-swing-interval time series, left-stance-interval time series and right-stance-interval time series (Figure 1 illustrates a single walking cycle with stride, stance, and swing). In order to minimize startup effects, the first 20s were excluded. With a median filter, outliers brought by turnarounds at the end of the hallway were replaced with the median value described in [16], [57]. Data normalization may improve pattern recognition and reduce computational time [58], for each time series a Z-score normalization was implemented before further

processing [59]. Illustrations for each group in the dataset are shown in Figure 4.

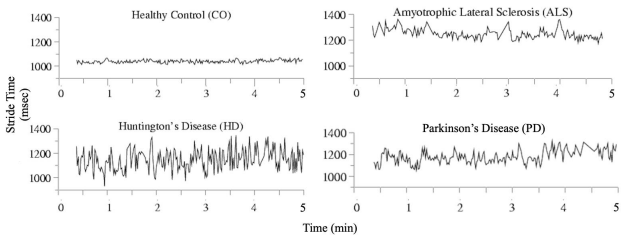


FIGURE 4. The example illustrations for the gait fluctuations of stride interval for each group.

B. EXPERIMENTS SETUP

We perform three binary classification tasks with the dataset, i.e., HC v.s ALS, HC v.s HD, and HC v.s PD. For each classification task, we use the time series of L-Stride, R-Stride, L-stance, R-Stance, L-Swing, and R-Swing separately. We call the binary classification with one of the six tracks of dynamics time series a trail, in which only two groups each with one gait fluctuations time series are considered. For example, in the trail of HC v.s ALS using L-Stride data, only the left stride time series from the HC group and ALS group is considered, which includes 29 L-Stride time series (16 from the HC group, while the rest are from the ALS group). Since the trails are small sample classification tasks, a better choice for the binary classification is leave-one-out cross-validation, i.e., each time using one sample as the testing sample while the rest samples are used as training samples. We train the classifiers with the topological features set generated from each time series via the TMA framework.

C. PERFORMANCE ASSESSMENT

For binary classification, the confusion matrices are calculated as illustrated in Table 1, where TP stands for true positive; TN stands for true negative; FN stands for false negative; FP stands for false positive. A confusion matrix contains information about actual labels and predicted labels by the classification system.

TABLE 1. The confusion matrix illustration.

	Predicted labels	
Actual	Positive	Negative
Positive	TP	FN
Negative	FP	TN

From the confusion matrix, we can get the accuracy, sensitivity and specificity parameters as follows respectively:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{FP + TN} \tag{10}$$

Moreover, for the comparison with other related works, we also consider the area under the receiver operating characteristic curve (AUC), which is from the confusion matrix information, where an AUC score of 1 means a perfect test, and an AUC score of 0.5 represents a random guess. A higher AUC score means the model is better for the classification task.

IV. EXPERIMENTAL RESULTS

In this section, we describe the classification results of the proposed framework, and the intermediate results, including the illustrations of point clouds, barcodes plots with comparison, persistence diagrams. Finally, we illustrate the results of the classification tasks and corresponding explanation.

A. PRE-PROCESSING AND TIME DELAY EMBEDDING

The gait fluctuations are converted into point clouds for topological feature extraction. Firstly, we perform the outlier detection to remove the extreme values lead by the turn-back during the walking experiment. After the outlier detection, then the time series were normalized with the settings in Section III.A. Then the same dimension parameter $d = 2$ and time delay lag set $\{4, 5, 6, 7, 8\}$ were used for time-delay embedding.

Since the length of each data sample varies due to the different walking speed of the subjects, the achieved point clouds have different scales. Meanwhile, we use a subsampling strategy to reduce the point cloud scale to lower the computational cost without changing the topological properties of the point cloud [60]. We use the scale parameter of 50 to represent the point cloud, i.e., all the embedded point clouds are sub-sampled into a 50-point scale (see Figure 5).

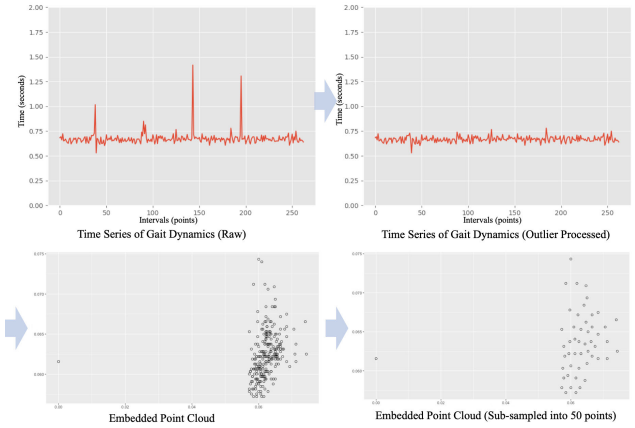


FIGURE 5. The illustration of pre-processing: (top left) raw gait fluctuations time series with outliers; (top right) outlier-processed time series; (bottom left) the point cloud from the time series after normalization; (bottom right) subsampled point cloud.

B. TOPOLOGICAL SIGNATURES

1) BARCODES

The information of the topological objects represents the topological spaces for the point clouds. Each barcode

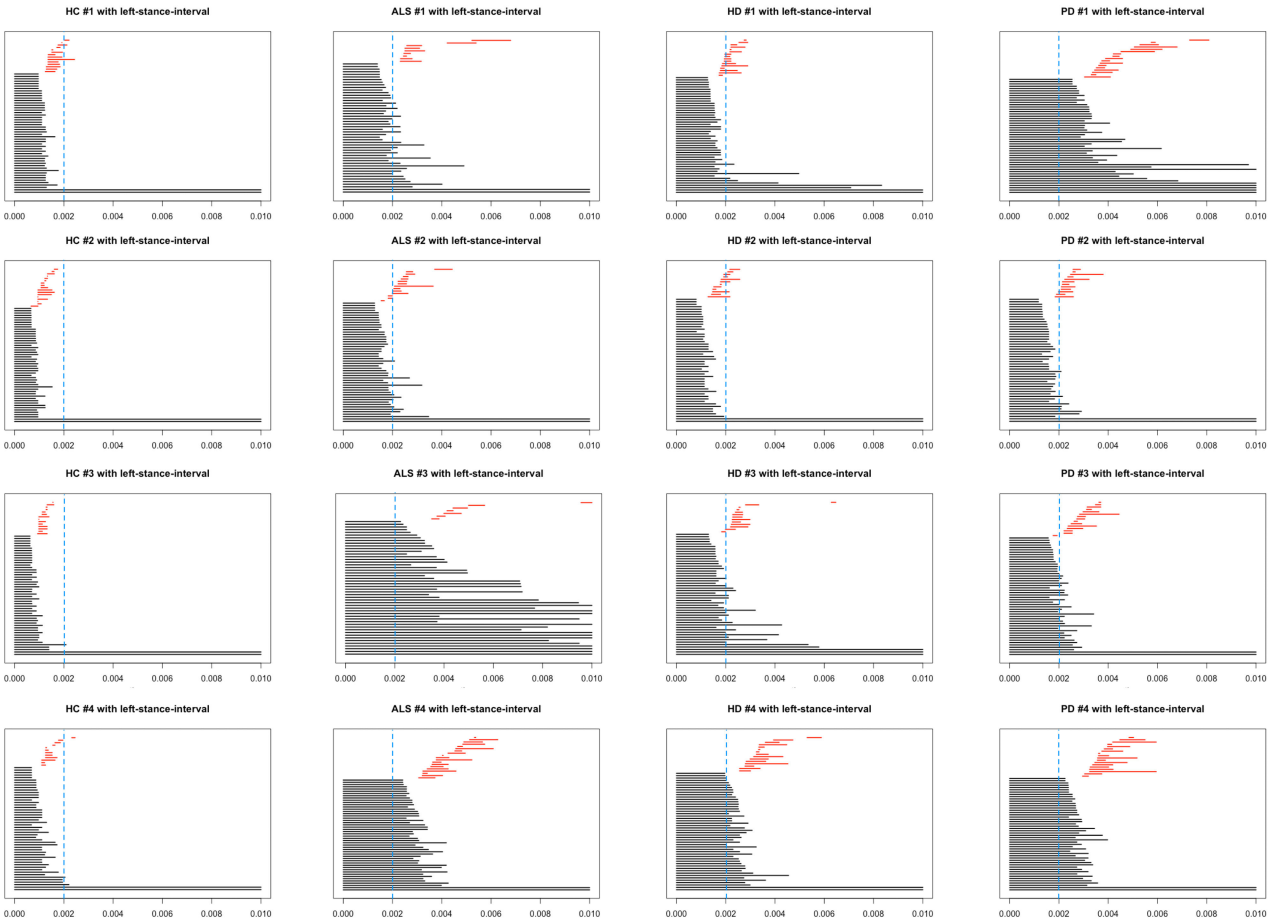


FIGURE 6. Barcode for point clouds from left-stance-interval gait fluctuations: 1^{st} column for HC; 2^{nd} column for ALS, 3^{rd} column for HD, 4^{th} column for PD.

describes the process of increasing the topological space's radius for a specific homology. In Figure 6, we illustrate the barcodes comparison from each group based on the gait fluctuations time series (we illustrate more results from the stance-, stride- and swing-based gait fluctuation time series of each foot in the supplement materials). Here we consider four subjects' barcodes for left-stance-interval gait fluctuations ($\tau = 7$ for time-delay embedding before barcodes extraction) from the HC, ALS, HD, and PD groups, respectively.

For each barcode plot in Figure 6, the dark bars are illustrations for the 0-dimensional homologies, i.e., the connected object; the red bars are illustrations for the 1-dimensional homologies namely the holes in the topological spaces. The horizon axis is the time parameter that reveals the duration of each topological object. The left end of the bar is the birth time, while the right end is the vanish (death) time. From the illustrations, we can tell that in the HC group (first row of Figure 6) the birth time and the vanish time are both earlier compare to other groups (2^{nd} row for ALS group, 3^{rd} row for HD group, and 4^{th} row for PD group). The significant differences are the foundations for the topological feature extraction process.

2) TOPOLOGICAL FEATURES

The persistent diagrams are another topological signature based on the barcodes of the point clouds. The red triangles and dark points illustrate the barcodes of 1-dimensional homologies and 0-dimensional homologies in Figure 7, respectively. The horizontal axis of the persistent diagram graph is the birth time while the vertical for the vanish (death) time. From the persistence diagram, we can see that in different groups (rows), the distributions for the birth-vanish coordinates are similar in intra-group and significantly different in inter-group comparisons. Also, we provide more results for the comparisons using different gait fluctuations time series, which share the same conclusion visually. Persistence landscapes are built with the persistence diagrams based on b_α and d_α. Examples of persistence landscapes are illustrated in Figure 8. Visually we cannot directly tell the differences since the values are accumulation values. Details of persistence landscape computation can be accessed from [52].

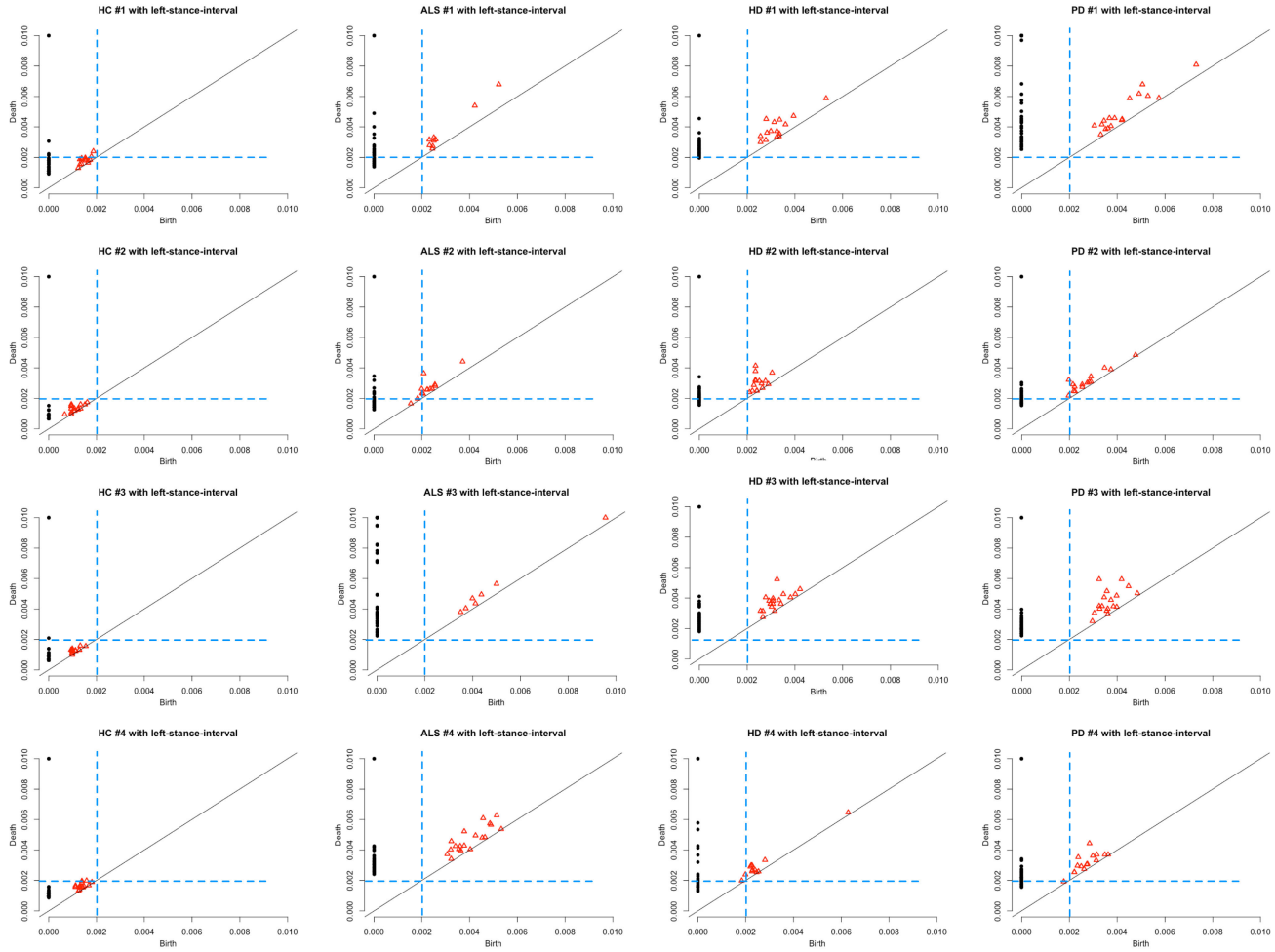


FIGURE 7. The corresponding persistence diagrams of the barcodes plot for Figure 6: 1^{st} column for HC; 2^{nd} column for ALS, 3^{rd} column for HD, 4^{th} column for PD.

C. PATTERN CLASSIFICATION

We evaluate the TMA framework with the topological features applying leave-one-out cross-validation. For comparison using different gait fluctuation time series, we keep the classifier parameters fixed, the tree estimator number is set as 1200, and the maximum depth value is 5. The results for the binary classification of HC vs. ALS, HC vs. HD, and HC vs. PD using random forests classifier with different time-delay parameters are illustrated in Table 2. We overstrike the best score for each classification task. The best AUC scores are 0.9135, 0.9906, and 0.9667 for HC vs. ALS, HC vs. HD, and HC vs. PD classification, respectively.

1) HC V.S ALS

From the 3^{rd} column to 6^{th} column in Table 2, the results for the HC v.s ALS classification task are illustrated. The best achieved model represented by the AUC scores for each time series are 0.9135, 0.7885, 0.8293, 0.8317, 0.8558, and 0.8654 for left-foot-swing ($\tau = 4$), right-foot-swing ($\tau = 7$), left-foot-stride($\tau = 6$), right-foot-stride($\tau = 6$), left-foot-stance($\tau = 5$), and right-foot-stance ($\tau = 6$) respectively. The corresponding average accuracies are 79.31%, 82.76%, 75.86%, 79.31%, 79.31%, and 75.86% for the totally 29 subjects (16 HC subjects, and 13 ALS subjects). The sensitivities are 76.92%, 69.23%, 76.92%, 61.54%, 76.92%, and 61.54%, which means there are 10, 9, 10, 8, 10, and 8 subjects recognized from the total 13 ALS subjects with the left-foot-swing ($\tau = 4$), right-foot-swing ($\tau = 7$), left-foot-stride($\tau = 6$), right-foot-stride($\tau = 6$), left-foot-stance($\tau = 5$), and right-foot-stance ($\tau = 6$) information.

The overall best AUC score is 0.9135 when using the left-foot-swing time series, and embedded with a time lag $\tau = 4$. The corresponding confusion matrix of the best case is illustrated in Table 3.

2) HC V.S HD

From the 7^{th} column to 10^{th} column in Table 2, the results for the HC v.s HD analysis are illustrated. The best model represented by the AUC scores for each time series are 0.9125, 0.9188, 0.9906, 0.9625, 0.9281, and 0.9875 for

TABLE 2. Achieved Results of TMA framework: each row corresponds the results for different gait flutuation time series with different embedding time lag.

Gait Fluctuations	Time Delay	HC vs ALS				HC vs HD				HC vs PD			
		Spe	Sen	Acc	AUC	Spe	Sen	Acc	AUC	Spe	Sen	Acc	AUC
Left-foot-stride	$\tau = 4$	75.00%	69.23%	72.41%	0.7644	87.50%	90.00%	88.89%	0.9656	68.75%	80.00%	74.19%	0.7625
	$\tau = 5$	87.50%	76.92%	82.76%	0.8269	75.00%	80.00%	77.78%	0.8969	68.75%	93.33%	80.65%	0.9083
	$\tau = 6$	75.00%	76.92%	75.86%	0.8293	81.25%	95.00%	88.89%	0.9596	81.25%	80.00%	80.65%	0.8667
	$\tau = 7$	93.75%	69.23%	82.76%	0.7788	81.25%	85.00%	83.33%	0.9156	81.25%	73.33%	77.42%	0.8125
	$\tau = 8$	93.75%	76.92%	86.21%	0.8221	**87.50%**	**95.00%**	**91.67%**	**0.9906**	87.50%	80.00%	83.87%	0.9542
Right-foot-stride	$\tau = 4$	81.25%	61.54%	72.41%	0.7692	75.00%	85.00%	80.56%	0.9250	75.00%	73.33%	74.19%	0.7750
	$\tau = 5$	87.50%	69.23%	79.31%	0.7548	68.75%	90.00%	80.56$	0.9219	81.25%	80.00%	80.65%	0.8792
	$\tau = 6$	93.75%	61.54%	79.31%	0.8317	87.50%	85.00%	86.11%	0.9594	75.00%	73.33%	74.19%	0.8542
	$\tau = 7$	81.25%	69.23%	75.86%	0.8173	68.75%	80.00%	75.00%	0.8953	68.75%	80.00%	74.19%	0.8000
	$\tau = 8$	75.00%	61.54%	68.97%	0.6803	81.25%	95.00%	88.89%	0.9625	75.00%	86.67%	80.65%	0.8083
Left-foot-stance	$\tau = 4$	81.25%	69.23%	75.86%	0.7740	87.50%	85.00%	86.11%	0.8969	75.00%	80.00%	77.42%	0.9125
	$\tau = 5$	81.25%	76.92%	79.31%	0.8558	81.25%	90.00%	86.11%	0.9281	68.75%	86.67%	77.42%	0.8250
	$\tau = 6$	75.00%	69.23%	72.41%	0.7260	75.00%	75.00%	75.00%	0.8062	56.25%	60.00%	58.06%	0.7021
	$\tau = 7$	87.50%	61.54%	75.86%	0.7212	81.25%	85.00%	83.33%	0.8875	75.00%	66.67%	70.97%	0.8000
	$\tau = 8$	75.00%	69.23%	72.41%	0.7260	68.75%	85.00%	77.78%	0.8938	68.75%	60.00%	64.52%	0.6792
Right-foot-stance	$\tau = 4$	81.25%	61.54%	72.41%	0.8462	68.75%	85.00%	77.78%	0.8750	81.25%	80.00%	80.67%	0.8833
	$\tau = 5$	81.25%	69.23%	75.86%	0.8125	93.75%	90.00%	91.67%	0.9875	68.75%	73.33%	70.97%	0.8125
	$\tau = 6$	87.50%	61.54%	75.86%	0.8654	75.00%	75.00%	75.00%	0.9219	75.00%	73.33%	74.19%	0.8708
	$\tau = 7$	87.50%	61.54%	75.86%	0.8101	81.52%	90.00%	86.11%	0.9688	87.50%	93.33%	90.32%	0.9208
	$\tau = 8$	93.75%	69.23%	82.76%	0.8077	87.50%	85.00%	86.11%	0.9125	**87.50%**	**86.67%**	**87.10%**	**0.9667**
Left-foot-swing	$\tau = 4$	**81.25%**	**76.92%**	**79.31%**	**0.9135**	100.00%	75.00%	86.11%	0.9125	93.75%	73.33%	83.87%	0.8938
	$\tau = 5$	75.00%	76.92%	75.86%	0.8510	93.75%	85.00%	88.89%	0.8750	100.00%	80.00%	90.32%	0.8708
	$\tau = 6$	87.50%	76.92%	82.76%	0.8702	93.75%	80.00%	86.11%	0.8984	93.75%	80.00%	87.10%	0.8750
	$\tau = 7$	81.25%	61.54%	72.41%	0.8510	81.25%	80.00%	80.56%	0.8875	87.50%	80.00%	83.87%	0.8667
	$\tau = 8$	87.50%	69.23%	79.31%	0.8053	75.00%	80.00%	77.78%	0.8344	87.50%	66.67%	77.42%	0.8875
Right-foot-swing	$\tau = 4$	62.50%	53.85%	58.61%	0.5962	81.25%	75.00%	77.78%	0.8188	87.50%	80.00%	83.87%	0.9167
	$\tau = 5$	87.50%	61.54%	75.86%	0.7596	87.50%	70.00%	77.78%	0.8500	100.00%	93.33%	96.77%	0.9333
	$\tau = 6$	75.00%	46.15%	62.07%	0.6298	87.50%	75.00%	80.56%	0.8906	87.50%	86.67%	87.10%	0.8583
	$\tau = 7$	93.75%	69.23%	82.76%	0.7885	87.50%	80.00%	83.33%	0.8719	93.75%	93.33%	93.55%	0.9625
	$\tau = 8$	81.25%	53.85%	68.97%	0.7188	93.75%	75.00%	83.33%	0.9188	100.00%	0.8667%	93.55%	0.9125

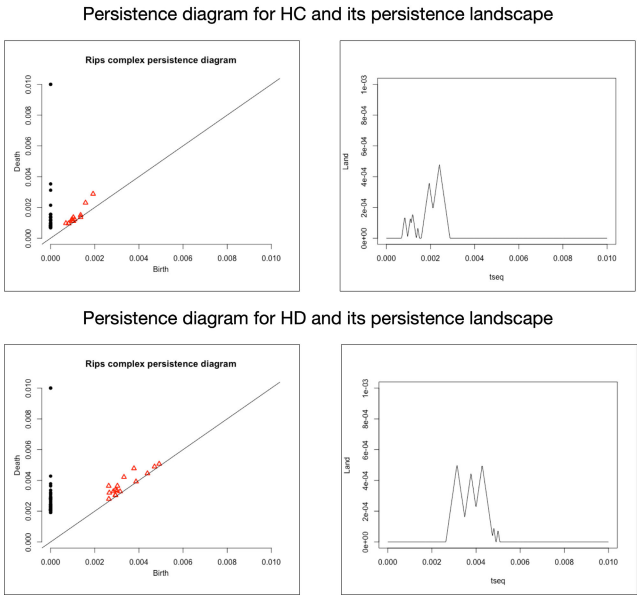


FIGURE 8. The illustrations for persistence landscapes of the barcodes and persistence diagrams (one HC example and a single HD example illustrated).

left-foot-swing ($\tau = 4$), right-foot-swing ($\tau = 8$), left-foot-stride($\tau = 8$), right-foot-stride($\tau = 8$), left-foot-stance ($\tau = 5$), and right-foot-stance ($\tau = 5$) respectively.

TABLE 3. Confusion matrix for the best case of HC v.s ALS.

	Predicted labels	
Actual	ALS	HC
ALS	10	3
HC	3	13

The corresponding average accuracies are 86.11%, 83.33%, 91.67%, 88.89%, 86.11%, and 91.67% for the totally 29 subjects (16 HC subjects, and 20 HD subjects). The sensitivities are 75%, 75%, 95%, 95%, 90%, and 90%, which means there are 15, 15, 19, 19, 18, and 18 subjects recognized from the total 20 HD subjects with the left-foot-swing ($\tau = 4$), right-foot-swing ($\tau = 8$), left-foot-stride ($\tau = 8$), right-foot-stride ($\tau = 8$), left-foot-stance($\tau = 5$), and right-foot-stance ($\tau = 5$) information.

The overall best AUC score is 0.9906 when using the left-foot-stride time series, and embedded with a time lag $\tau = 8$. The corresponding confusion matrix of the best case is illustrated in Table 4.

TABLE 4. Confusion matrix for the best case of HC v.s HD.

	Predicted labels	
Actual	HD	HC
HD	19	1
HC	2	14

3) HC V.S PD

From the 11^{th} column to 14^{th} column in Table 2, the results for the binary classification for HC and PD are illustrated. The best model represented by the AUC scores for each time series are 0.8938, 0.9625, 0.9542, 0.8792, 0.9125, and 0.9667 for left-foot-swing ($\tau = 4$), right-foot-swing ($\tau = 7$), left-foot-stride($\tau = 8$), right-foot-stride($\tau = 5$), left-foot-stance($\tau = 4$), and right-foot-stance ($\tau = 8$) respectively. The corresponding average accuracies are 83.87%, 93.55%, 83.87%, 80.65%, 77.42%, and 87.10% for the totally 29 subjects (16 HC subjects, and 15 HD subjects). The sensitivities are 73.33%, 93.33%, 80.00%, 80.00%, 80.00%, and 86.67%, which means there are 11, 14, 12, 12, 12, and 13 subjects recognized from the total 15 HD subjects with the left-foot-swing ($\tau = 4$), right-foot-swing ($\tau = 7$), left-foot-stride ($\tau = 8$), right-foot-stride ($\tau = 5$), left-foot-stance ($\tau = 4$), and right-foot-stance ($\tau = 8$) information. The best AUC score for the classification of HC and PD is 0.9667, with corresponding specificity = 87.50%,sensitivity = 86.67% and accuracy = 87.10%.

The overall best AUC score for HC v.s PD is 0.9667 when using the right-foot-stride time series, and embedded with a time lag $\tau = 8$. The corresponding confusion matrix of the best case is illustrated in Table 5.

TABLE 5. Confusion matrix for the best case of HC v.s PD.

	Predicted labels	
Actual	PD	HC
PD	13	2
HC	2	14

D. EXPERIMENTAL ENVIRONMENTS

A computer with 2.6 GHz 6-Core Intel Core i7 CPU using 32GB memory is used for the experiments. The whole process of the topological feature generation for the whole dataset costs less than 3 minutes. The cost for one round training-and-testing is about 30 seconds for this 2-dimensional point cloud-based analytical framework.

The adopted software packages are listed below:

1) Time delay embedding: the R package ''nonlinearTseries'';
2) The point cloud sampling: the Matlab package ''JavaPlex'';
3) Barcodes, persistence diagram, and persistence landscape features: the R package ''TDA'';
4) Random forests classifier: the Python package ''scikit-learn''.

V. DISCUSSION

A. TIME DELAY EMBEDDING

The topological features are based on the data point cloud generated by the reconstruction of the state space. The state spaces illustrate one approximation of the real system, from which the data are sampled. State-space reconstruction has been used as one foundation of nonlinear time-series analysis. In this study, we try to explore the features from the reconstructed state-space using the TDA technique, and then later use it in the classification task. However, the determination of the parameters can still be challenging for the task. Here we set two approximations and ignore the theoretical analysis for the embedding dimension d and the time delay lag parameter τ. We simply use $d = 2$ as the embedding dimension for the ease of TDA technique computation. As in [40] states that a larger dimension may not bring better performance for the classification task. We separately considered the time series to observe the subjects' motion system with different time-delay lags τ. As a result, the experimental outcomes are different. A detailed discussion of the determination of d and τ was proposed in [45].

B. RELATED WORK AND COMPARISONS

In this work, we adopt a TMA framework for gait fluctuations analysis for each foot's swing, stride, and stance information. As the experimental results above show, the neurodegenerative disease classification tasks are well handled. The performance evaluation is comparable to the previous related works using the same dataset. In this section, we include the pieces of literature which adopt different techniques in feature extraction rather than feature selection or optimization [61]. The TDA techniques provide alternative insights compared to the previous techniques; simultaneously, the features show an excellent ability to distinguish different neurodegenerative diseases from the healthy control subjects.

Previous studies have indicated that the gait rhythm fluctuations for the subjects are representative for characterizing the neurodegenerative disease. Plenty of gait fluctuation-based features was proposed for the disease classification task. In [16], [62], based on a nonparametric Parzen-Window method, estimate the probability density functions of stride interval, swing interval, and stance interval had been established. From which the gait rhythm standard deviation, and a signal turns count (STC) parameter was derived as a dominant feature. In [63], the radial basis function (RBF) neural networks were adopted to extract the prior knowledge of gait fluctuations, then used as the estimators for different kinds of neurodegenerative disease, i.e., the deterministic learning (DL). Reference [15] adopted the differences in stride length, cadence, stance time, and double support time from the gait signals, namely the spatial-temporal gait features, with which the classifiers were adopted for classification. Reference [2] considered the relations between both feet in the dataset, the features of phase synchronization index and conditional entropy (PSCE) were applied to the five types of time series pairs of gait rhythms (stride time, swing time, stance time, percentage of swing time and percentage of stance time). Ren *et al.* [17] has investigated the properties of frequency range distributions of gait rhythms, after which an empirical mode decomposition (EMD) was implemented for decomposing the time series of gait rhythms into intrinsic mode functions, then Kendall's coefficient of concordance and the ratio for energy change for different IMFs were calculated as

the feature for classification, which shows excellent performance. An adaptive neuro-fuzzy inference system (ANFIS) was presented for identification of the gait of patients with the neurodegenerative disease, in which the neural network adaptive capabilities and the fuzzy logic qualitative approach were combined as a classification model using the swing, stride, and stance information [64]. Reference [65] adopted the recordings of compound force signal and used a 6-level Discrete Wavelet Transform (DWT) with different wavelets as features to deal with ALS detection tasks, [66] also adopt Wavelet transformation as the feature extractor for classification tasks. The multi-fractal detrended cross-correlation Analysis (MFDXA) methodology was adopted in the total force under the left foot and right foot [67]. Reference [68] introduced a hidden Markov model (HMM) with the Gaussian mixtures method to model the gait dynamics to classify the healthy control group and Parkinson's disease group [69].

Based on the right-foot-stride gait fluctuations, by transforming time series into images, the texture features were extracted by texture analysis, novel insights were brought into the disease patterns distinguishing task [19].

In this work, we try to analyze the gait fluctuations patterns from geometrical & topological viewpoint by adopting the proposed TMA framework. Similar to [19] the image-based texture analysis, we can summarize the proposed method to a point cloud-based analysis. This novel insight brought powerful representation tools for neurodegenerative disease analysis.

TABLE 6. Comparisons of AUC and LOO-CV Results for classification of HC and ALS groups.

Methods	Specificity	Sensitivity	Accuracy	AUC
EMD + RF [17]	x	x	x	0.900
EMD + SLR [17]	x	x	x	0.859
EMD + MLP [17]	x	x	x	0.934
EMD + NB [17]	x	x	x	0.891
EMD + SVM [17]	x	x	x	0.906
DL [63]	92.31%	87.50%	89.66%	x
PSCE + MP [2]	81.3%	68.8%	x	0.824
PSCE + RF [2]	93.75%	75.0%	x	0.789
PSCE + NB [2]	87.5%	62.5%	x	0.750
TMA: Left-Swing	**81.25%**	**76.92%**	**79.31%**	**0.914**

TABLE 7. Comparisons of AUC and LOO-CV Results for classification of HC and HD groups.

Methods	Specificity	Sensitivity	Accuracy	AUC
EMD + RF [17]	x	x	x	0.885
EMD + SLR [17]	x	x	x	0.843
EMD + MLP [17]	x	x	x	0.878
EMD + NB [17]	x	x	x	0.898
EMD + SVM [17]	x	x	x	0.900
DL [63]	85.00%	81.25%	83.33%	x
PSCE + MP [2]	100 %	85.00%	x	0.910
PSCE + RF [2]	95.0 %	90.0 %	x	0.959
PSCE + NB [2]	95.0%	80.0%	x	0.920
TMA: Left-Stride	**87.50%**	**95.00%**	**91.67%**	**0.991**

TABLE 8. Comparisons of AUC and LOO-CV Results for classification of HC and PD groups.

Methods	Specificity	Sensitivity	Accuracy	AUC
σ & STC + LS-SVM [16]	x	x	90.32%	0.952
HMM [68]	93.33%	87.50%	90.32	x
EMD + RF [17]	x	x	x	0.865
EMD + SLR [17]	x	x	x	0.949
EMD + MLP [17]	x	x	x	0.910
EMD + NB [17]	x	x	x	0.875
EMD + SVM [17]	x	x	x	0.906
DL [63]	86.67%	86.50%	87.10%	x
PSCE + MP [2]	100%	81.3%	x	0.928
PSCE + RF [2]	93.8%	87.5%	x	0.910
PSCE + NB [2]	87.5%	81.3%	x	0.898
TMA: Right-Stance	**87.50%**	**86.67%**	**87.10%**	**0.967**

Table 6, 7, and 8 illustrate the HC v.s ALS, HC v.s HD, and HC v.s PD classification tasks separately with comparison to some of the mentioned tasks. We only list the best records we achieved in this work using a random forest classifier. The results show the excellent distinguishing ability of the TMA framework. The abbreviations in Table 6, 7, and 8 are: random forests (RF) classifier, simple logistic regression (SLR) classifier, naive Bayesian (NB) classifier, support vector machine (SVM) classifier, and multiple layer perceptron (MLP) classifier. We denote the proposed method results as TMA together with the adopted gait fluctuation time series.

VI. CONCLUSION

A. CONCLUSION

In this study, a TMA framework transforming gait fluctuations time series into data point clouds using time-delay embedding, with a followed TDA technique, was used to extract persistence landscape features for gait fluctuations analysis toward neurodegenerative disease classification. A comprehensive study using the proposed framework with multiple human gait fluctuations is proposed for understanding the pathological characteristics of neurodegenerative diseases. The built classification systems have shown excellent performance and great distinguishing ability in the neurodegenerative analysis. Therefore, the insights brought with the TDA technique greatly enrich the gait analysis tools. The proposed TMA framework shows promising ability in the modeling for disease state analysis or even clinical prediction studies.

B. LIMITATIONS AND FUTURE WORK

We validated the TMA framework in multiple gait fluctuations that shows good distinguishability. However, the dataset adopted is based on a limited number of experiments. Further validations should be performed on a larger scale. Meanwhile, the optimal time delay lag in the embedding stage is hard to determine. We can see that the overall performance could be affected by several stages. The optimal combination of parameters from each stage is hard to search. Thus in the proposed experiments, we have to adopt several arbitrary parameter settings. From the achieved results, we can tell

that the topological features are powerful in revealing class discrepancies. An investigation of shorter length time series distinguishing ability is meaningful in the short-length and sample limited-time series analysis tasks.

APPENDIX
NOTATIONS OF TOPOLOGICAL DATA ANALYSIS
A. SIMPLICIAL COMPLEX

Simplexes and simplicial complex are the basic elements to model the abstract topological space. A topological space can be understood as a set of points sampled from some $\mathbb{R}^n$ with neighbors. The neighbor relations could be used to study the connectivities in a graph. In algebraic topology, the basic building blocks are simplicial complex, which is a data structure that explains how to 'glue' a topological space with simplices, i.e., points, edges, triangles, tetrahedra, and even their higher-dimensional generalizations.

Mathematically, a k-dimensional simplex is defined by $k+1$ vertices. Simplex: Given a family of sets, any subset of cardinality $k+1$ is called a k-simplex. The vertices can be considered as 0-simplices, edges for 1-simplices, and triangular faces as 2-simplices. A simplicial complex is a collection of simplices together with their faces. The vertex subset defines the face of a simplex. For instance, the faces of an edge of two vertices are the two endpoint vertices, and the edge itself. The faces of a 2-simplex (triangle with three vertices) include the three vertices, the three edges, and the triangle itself. The collection with a simplex and its all faces is called a simplicial complex as: $\mathcal{K}$ is a finite collection of simplices such that

1) any face of $\sigma \in \mathcal{K}$ is also in $\mathcal{K}$
2) for $\sigma_1, \sigma_2 \in \mathcal{K}$, $\sigma_1 \cap \sigma_2$ is a face of both σ_1 and σ_2

B. THE RIPS COMPLEX AND GRAPH FILTRATION

For the building process of Rips complex, here we use a similar illustration as in [70]. We consider the distance $d(x, y)$ for vertices x and y. For the complex theory, we consider a ϵ-ball with the radius $\epsilon = 0$ for the vertices set, namely the original data point cloud. Intuitively, if we increase the radius ϵ gradually when the distance between any two vertices $d(x, y)$ is less than ϵ, the edge appears. Then it is easy to think that for three vertices, the ϵ-balls intersect mutually (not merge), a triangle appears, while for three vertices, a tetrahedron emerges, and higher dimensional simplices whenever possible. Similarly, there are much more higher-dimensional holes when we consider a more complicated point cloud in practical applications, check [71] for more results. For the 2-dimensional situation in this work, we already describe the process of Figure 3.

Mathematically, consider data point cloud $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^n$, the associate topological space using the Rips simplicial complex construction, which could be denoted by $\mathcal{R}(X, \epsilon)$ in which $\mathcal{R}$ stands for Rips. The process in Figure 3 can be considered as the sequence:

$$\mathcal{R}(X, \epsilon_0), \mathcal{R}(X, \epsilon_1), \cdots \mathcal{R}(X, \epsilon_n) \tag{11}$$

and when the ϵ increase, the previous Rips complex is included in the subsequent one, i.e.

$$\mathcal{R}(X, \epsilon_0) \subseteq \mathcal{R}(X, \epsilon_1) \subseteq \cdots \subseteq \mathcal{R}(X, \epsilon_n) \tag{12}$$

where $\epsilon_0 \leq \epsilon_1 \leq \cdots \leq \epsilon_n$. The increasing sequence of ϵ value produces a *filtration*: given a set X, the K-simplex $\{\sigma_1, \sigma_2, \ldots, \sigma_{k+1}\}$, then we have the definition of filtration:

A filtration of a (finite) simplicial complex K is a sequence of sub-complexes such that

1) $\emptyset = K^0 \subset K^1 \subset K^2 \cdots \subset K^m = K$
2) $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K

ACKNOWLEDGMENT

Y. Yan would like to show appreciation for the directing from Prof. Yannis Goulermas, when he was with the Computer Science Department, University of Liverpool, U.K., in 2017. Dr. Goulermas inspired the idea of applying tools from the Theoretical Computer Science (TCS) area, such as geometrical and topological methods in this work. He would also like to thank the efforts taken by the anonymous reviewers.

REFERENCES

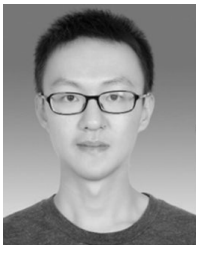
[1] J. Jankovic, "Parkinson's disease: Clinical features and diagnosis," *J. Neurol., Neurosurg. Psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.
[2] P. Ren, W. Zhao, Z. Zhao, M. L. Bringas-Vega, P. A. Valdes-Sosa, and K. M. Kendrick, "Analysis of gait rhythm fluctuations for neurodegenerative diseases by phase synchronization and conditional entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 2, pp. 291–299, Feb. 2016.
[3] C. A. Ross and S. J. Tabrizi, "Huntington's disease: From molecular pathogenesis to clinical treatment," *Lancet Neurol.*, vol. 10, no. 1, pp. 83–98, Jan. 2011.
[4] R. Pahwa and K. E. Lyons, *Handbook of Parkinson's Disease*. Boca Raton, FL, USA: CRC Press, 2013.
[5] J. M. Hausdorff, Y. Ashkenazy, C.-K. Peng, P. C. Ivanov, H. E. Stanley, and A. L. Goldberger, "When human walking becomes random walking: Fractal analysis and modeling of gait rhythm fluctuations," *Phys. A, Stat. Mech. Appl.*, vol. 302, nos. 1–4, pp. 138–147, Dec. 2001.
[6] M. Yang, H. Zheng, H. Wang, and S. McClean, "Feature selection and construction for the discrimination of neurodegenerative diseases based on gait analysis," in *Proc. 3rd Int. ICST Conf. Pervas. Comput. Technol. Healthcare*, 2009, pp. 1–7.
[7] J. D. Schaafsma, N. Giladi, Y. Balash, A. L. Bartels, T. Gurevich, and J. M. Hausdorff, "Gait dynamics in Parkinson's disease: Relationship to parkinsonian features, falls and response to levodopa," *J. Neurological Sci.*, vol. 212, nos. 1–2, pp. 47–53, Aug. 2003.
[8] H. Zhao, Z. Wang, S. Qiu, J. Wang, F. Xu, Z. Wang, and Y. Shen, "Adaptive gait detection based on foot-mounted inertial sensors and multi-sensor fusion," *Inf. Fusion*, vol. 52, pp. 157–166, Dec. 2019.
[9] J. M. Hausdorff, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Is walking a random walk? Evidence for long-range correlations in stride interval of human gait," *J. Appl. Physiol.*, vol. 78, no. 1, pp. 349–358, Jan. 1995.
[10] J. M. Hausdorff, P. L. Purdon, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Fractal dynamics of human gait: Stability of long-range correlations in stride interval fluctuations," *J. Appl. Physiol.*, vol. 80, no. 5, pp. 1448–1457, May 1996.
[11] J. M. Hausdorff, S. L. Mitchell, R. Firtion, C. K. Peng, M. E. Cudkowicz, J. Y. Wei, and A. L. Goldberger, "Altered fractal dynamics of gait: Reduced stride-interval correlations with aging and Huntington's disease," *J. Appl. Physiol.*, vol. 82, no. 1, pp. 262–269, Jan. 1997.
[12] J. M. Hausdorff, A. Lertratanakul, M. E. Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger, "Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis," *J. Appl. Physiol.*, vol. 88, no. 6, pp. 2045–2053, Jun. 2000.
[13] J. M. Hausdorff, "Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking," *Hum. Movement Sci.*, vol. 26, no. 4, pp. 555–589, Aug. 2007.

[14] J. Kamruzzaman and R. K. Begg, "Support vector machines and other pattern recognition approaches to the diagnosis of cerebral palsy gait," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2479–2490, Dec. 2006.
[15] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, "Classification of Parkinson's disease gait using spatial-temporal gait features," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 6, pp. 1794–1802, Nov. 2015.
[16] Y. Wu and S. Krishnan, "Statistical analysis of gait rhythm in patients with Parkinson's disease," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 150–158, Apr. 2010.
[17] P. Ren, S. Tang, F. Fang, L. Luo, L. Xu, M. L. Bringas-Vega, D. Yao, K. M. Kendrick, and P. A. Valdes-Sosa, "Gait rhythm fluctuation analysis for neurodegenerative diseases by empirical mode decomposition," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 52–60, Jan. 2017.
[18] T. D. Pham and H. Yan, "Tensor decomposition of gait dynamics in Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1820–1827, Aug. 2018.
[19] T. D. Pham, "Texture classification and visualization of time series of gait dynamics in patients with neuro-degenerative diseases," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 188–196, Jan. 2018.
[20] P. Pławiak, M. Abdar, J. Pławiak, V. Makarenkov, and U. R. Acharya, "DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring," *Inf. Sci.*, vol. 516, pp. 401–418, Apr. 2020.
[21] Y. Yan, X. Qin, Y. Wu, N. Zhang, J. Fan, and L. Wang, "A restricted Boltzmann machine based two-lead electrocardiography classification," in *Proc. IEEE 12th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Jun. 2015, pp. 1–9.
[22] P. Pławiak, M. Abdar, and U. Rajendra Acharya, "Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105740.
[23] T. Tuncer, S. Dogan, P. Pławiak, and U. Rajendra Acharya, "Automated arrhythmia detection using novel hexadecimal local pattern and multilevel wavelet transform with ECG signals," *Knowl.-Based Syst.*, vol. 186, Dec. 2019, Art. no. 104923.
[24] M. Abdar, W. Książek, U. R. Acharya, R.-S. Tan, V. Makarenkov, and P. Pławiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease," *Comput. Methods Programs Biomed.*, vol. 179, Oct. 2019, Art. no. 104992.
[25] M. Abdar, V. N. Wijayaningrum, S. Hussain, R. Alizadehsani, P. Plawiak, U. R. Acharya, and V. Makarenkov, "IAPSO-AIRS: A novel improved machine learning-based system for wart disease treatment," *J. Med. Syst.*, vol. 43, no. 7, p. 220, Jul. 2019.
[26] Z. Ju, G. Ouyang, M. Wilamowska-Korsak, and H. Liu, "Surface EMG based hand manipulation identification via nonlinear feature extraction and classification," *IEEE Sensors J.*, vol. 13, no. 9, pp. 3302–3311, Sep. 2013.
[27] A. S. Khaled, M. I. Owis, and A. S. Mohamed, "Employing time-domain methods and poincaré plot of heart rate variability signals to detect congestive heart failure," *BIME J.*, vol. 6, no. 1, pp. 35–41, 2006.
[28] E. D. Übeyli, "Lyapunov exponents/probabilistic neural networks for analysis of EEG signals," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 985–992, Mar. 2010.
[29] T. Penzel, J. W. Kantelhardt, L. Grote, J. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1143–1151, Oct. 2003.
[30] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000.
[31] G. Carlsson, "Topology and data," *Bull. Amer. Math. Soc.*, vol. 46, no. 2, pp. 255–308, 2009.
[32] F. Chazal and B. Michel, "An introduction to topological data analysis: Fundamental and practical aspects for data scientists," 2017, *arXiv:1710.04019*. [Online]. Available: http://arxiv.org/abs/1710.04019
[33] G. Singh, F. Mémoli, and G. E. Carlsson, "Topological methods for the analysis of high dimensional data sets and 3d object recognition," in *Proc. SPBG*, 2007, pp. 91–100.
[34] V. Kovacev-Nikolic, P. Bubenik, D. Nikolić, and G. Heo, "Using persistent homology and dynamical distances to analyze protein binding," *Stat. Appl. Genet. Mol. Biol.*, vol. 15, no. 1, pp. 19–38, Jan. 2016.
[35] 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," *Sci. Rep.*, vol. 3, no. 1, p. 1236, Dec. 2013.
[36] B. Rieck and H. Leitte, "Persistent homology for the evaluation of dimensionality reduction schemes," *Comput. Graph. Forum*, vol. 34, no. 3, pp. 431–440, Jun. 2015.
[37] L. Duponchel, "Exploring hyperspectral imaging data sets with topological data analysis," *Analytica Chim. Acta*, vol. 1000, pp. 123–131, Feb. 2018.
[38] L. M. Seversky, S. Davis, and M. Berger, "On time-series topological data analysis: New data and opportunities," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 59–67.
[39] S. Emrani, T. Gentimis, and H. Krim, "Persistent homology of delay embeddings and its application to wheeze detection," *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 459–463, Apr. 2014.
[40] Z. Zhang, Y. Song, H. Cui, J. Wu, F. Schwartz, and H. Qi, "Topological analysis and Gaussian decision tree: Effective representation and classification of biosignals of small sample size," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2288–2299, Sep. 2017.
[41] B. Safarbali and S. M. R. Hashemi Golpayegani, "Nonlinear dynamic approaches to identify atrial fibrillation progression based on topological methods," *Biomed. Signal Process. Control*, vol. 53, Aug. 2019, Art. no. 101563.
[42] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, vol. 7. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[43] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*. Berlin, Germany: Springer, 1981, pp. 366–381.
[44] J. Frank, S. Mannor, and D. Precup, "Activity and gait recognition with time-delay embeddings," in *Proc. 24th AAAI Conf. Artif. Intell.*, 2010, pp. 1581–1586.
[45] E. Bradley and H. Kantz, "Nonlinear time-series analysis revisited," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 9, Sep. 2015, Art. no. 097610.
[46] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*. Providence, RI, USA: American Mathematical Society, 2010.
[47] L. Vietoris, "Über den höheren zusammenhang kompakter räume und eine klasse von zusammenhangstreuen abbildungen," *Mathematische Annalen*, vol. 97, no. 1, pp. 454–472, 1927.
[48] T. K. Dey, F. Fan, and Y. Wang, "Graph induced complex on point data," in *Proc. 29th Annu. Symp. Symp. Comput. Geometry (SoCG)*. New York, NY, USA: ACM, 2013, pp. 107–116.
[49] B. Osting, S. Palande, and B. Wang, "Spectral sparsification of simplicial complexes for clustering and label propagation," 2017, *arXiv:1708.08436*. [Online]. Available: http://arxiv.org/abs/1708.08436
[50] A. Zomorodian, "Fast construction of the vietoris-rips complex," *Comput. Graph.*, vol. 34, no. 3, pp. 263–271, Jun. 2010.
[51] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 77–102, 2015.
[52] P. Bubenik and P. Dłotko, "A persistence landscapes toolbox for topological statistics," *J. Symbolic Comput.*, vol. 78, pp. 91–114, Jan. 2017.
[53] P. Bubenik, "The persistence landscape and some of its properties," 2018, *arXiv:1810.04963*. [Online]. Available: http://arxiv.org/abs/1810.04963
[54] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[55] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016.
[56] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
[57] Y. Yan, K. Ivanov, O. Mumini Omisore, T. Igbe, Q. Liu, Z. Nie, and L. Wang, "Gait rhythm dynamics for neuro-degenerative disease classification via persistence landscape-based topological representation," *Sensors*, vol. 20, no. 7, p. 2006, Apr. 2020.
[58] N. Md. Tahir and H. H. Manap, "Parkinson disease gait classification based on machine learning approach," *J. Appl. Sci.*, vol. 12, no. 2, pp. 180–185, Feb. 2012.
[59] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.
[60] V. De Silva and G. Carlsson, "Topological estimation using witness complexes," in *Proc. 1st Eurographics Conf. Point-Based Graph. (SPBG)*. Goslar, Germany: Eurographics Association, 2004, p. 157–166.
[61] A. Athisakthi and M. P. Rani, "Recognize vital features for classification of neurodegenerative diseases," in *Advances in Computational Intelligence*. Singapore: Springer, 2020, pp. 287–301.

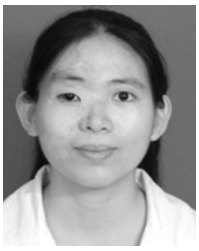
[62] Y. Wu and L. Shi, "Analysis of altered gait cycle duration in amyotrophic lateral sclerosis based on nonparametric probability density function estimation," *Med. Eng. Phys.*, vol. 33, no. 3, pp. 347–355, Apr. 2011.
[63] W. Zeng and C. Wang, "Classification of neurodegenerative diseases using gait dynamics via deterministic learning," *Inf. Sci.*, vol. 317, pp. 246–258, Oct. 2015.
[64] Q. Ye, Y. Xia, and Z. Yao, "Classification of gait patterns in patients with neurodegenerative disease using adaptive neuro-fuzzy inference system," *Comput. Math. Methods Med.*, vol. 2018, pp. 1–8, Sep. 2018.
[65] S. Bilgin, "The impact of feature extraction for the classification of amyotrophic lateral sclerosis among neurodegenerative diseases and healthy subjects," *Biomed. Signal Process. Control*, vol. 31, pp. 288–294, Jan. 2017.
[66] S. Bilgin, "The comparison of neurodegenerative diseases and healthy subjects using discrete wavelet transform in gait dynamics," *J. Med. Bioeng.*, vol. 6, no. 1, pp. 35–38, Jun. 2017.
[67] S. Dutta, D. Ghosh, and S. Samanta, "Non linear approach to study the dynamics of neurodegenerative diseases by multifractal detrended cross-correlation analysis—A quantitative assessment on gait disease," *Phys. A, Stat. Mech. Appl.*, vol. 448, pp. 181–195, Apr. 2016.
[68] A. Khorasani and M. R. Daliri, "HMM for classification of Parkinson's disease based on the raw gait data," *J. Med. Syst.*, vol. 38, no. 12, p. 147, Dec. 2014.
[69] S. Marziyeh Ghoreshi Beyrami and P. Ghaderyan, "A robust, cost-effective and non-invasive computer-aided method for diagnosis three types of neurodegenerative diseases with gait signal analysis," *Measurement*, vol. 156, May 2020, Art. no. 107579.
[70] E. Munch, "A user's guide to topological data analysis," *J. Learn. Anal.*, vol. 4, no. 2, pp. 47–61, 2017.
[71] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proc. 41st Annu. Symp. Found. Comput. Sci.*, 2000, pp. 454–463.

YAN YAN (Member, IEEE) received the B.Eng. and M.Sc. degrees in instrument engineering from the Harbin Institute of Technology, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree in computer science. He worked as a Research Assistance with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, from 2012 to 2014. He has been an Honorary Research Assistant with the Department of Computer Science, University of Liverpool. His research interests are biomedical signal processing, machine learning, dynamical systems, and topological data analysis.

OLATUNJI MUMINI OMISORE (Member, IEEE) received the B.Tech. and M.Tech. degrees in computer science and the Ph.D. degree in pattern recognition and intelligent systems from the University of Chinese Academy of Sciences. He worked as a Software Engineer, a System Analyst, and a University Lecturer. He has nine years of research experience. He currently holds a postdoctoral position with the Centre for Medical Robotics and MIS Devices, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China. His research interests include constraints control of flexible and MIS robotic systems, computational intelligence, and digital libraries with specialty in data and knowledge mining.



YU-CHENG XUE received the bachelor's degree in computer science from Donghua University, Shanghai, China, in 2014, and the master's degree in computer science from The Chinese University of Hong Kong. His research interests include electromyogram signals processing and machine learning.



HUI-HUI LI (Member, IEEE) received the B.S. and M.S. degrees from Shenzhen University, Shenzhen, China, in 2003 and 2006, respectively, and the Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2011. She is currently an Assistant Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen. Her research interests include biomedical signal processing, medical ultrasound, and miniature antenna design.

QIU-HUA LIU received the master's degree in business management from Peking University, in 2017. She has been engaged in scientific research management for eight years. During these years, she has acquired rich knowledge on biomedical engineering and accumulated rich experience in project management. She also got Project Management Professional Qualification (PMP), in 2019, and won the Shenzhen Technology Invention Award, in 2017.

ZE-DONG NIE (Member, IEEE) has long been engaged in human body communication and wearable health technology. His research interests include noninvasive blood glucose monitoring, wearable sensors, communication system and protocol, biometric verification, and machine learning. He has applied for more than 60 invention patents and holds seven PCT patents. In addition, he has published more than 30 SCI/EI articles and holds 33 authorized invention patents. His research achievement won the second prize of Guangdong Technical Invention Award, in 2018, and the second prize of Shenzhen Technical Invention Award, in 2017. He won the honor of innovation and entrepreneurial talent of Shenzhen and excellent employee of SIAT four times. He is an Evaluation Expert of the National Natural Science Foundation of China and Guangdong Provincial Science and Technology Department. He is also a reviewer of several IEEE journals.

JIANPING FAN received the Ph.D. degree from the Chinese Academy of Sciences, in 1990. He is currently the Dean of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, and the Director of the Research Center for High Performance Computing. Prior to SIAT, he had been the Project Leader with the Institute of Computing Technology, the Deputy Chief Engineer with the National Research Center for Intelligent Computing Systems, the Director of the National Engineering Center on High Performance Computing, and the Vice Director of the Institute of Computing Technology. He had been in charge of and involved in developing the Dawning 1, Dawning 1000, Dawning 3000, Dawning 4000, and other series of Dawning supercomputers.

LEI WANG (Member, IEEE) received the B.Eng. degree in information and control engineering and the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 1995 and 2000, respectively. He was with the University of Glasgow, Glasgow, U.K., and Imperial College London, London, U.K., from 2000 to 2008. He is currently a Full Professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. He has published over 200 scientific articles, authored four book chapters, and holds 60 patents. His current research interests include body sensor networks, digital signal processing, and biomedical engineering.