International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(4s) 2026 pp. 208-215
https://doi.org/10.70917/ijcisim-2026-2504
Article

Topological Feature Extraction For Human Activity Recognition Using Persistent Homology

D. Sasikala[1], M. Renukadevi[2]

[1]Department of Mathematics, PSGR Krishnammal College for Women, Coimbatore-641004, Tamil Nadu, India. E-mail: dsasikala@psgrkcw.ac.in
[2]Department of Mathematics, KPR College of Arts Science and Research, Coimbatore-641407, Tamil Nadu, India. E-mail: renukadevi.velumani@gmail.com

Abstract: The focus of this article is to investigate the effectiveness of Topological Data Analysis (TDA) for Human Activity Recognition (HAR) using smartphone sensor data. Using persistent homology, topological characteristics were extracted from the Human Activity Recognition Using Smartphones dataset. Representative persistence intervals were chosen using the AvgInt sampling technique. Three machine learning methods were then used to classify the retrieved topological descriptors: Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Confusion matrix-based classification accuracy and scatter plot visualizations were used to assess the suggested framework. According to experimental findings, TDA-derived features are highly discriminative for classification and efficiently capture activity-related patterns. Among the classifiers that were assessed, Random Forest demonstrated relatively lower accuracy, whereas SVM and KNN reached the same classification performance. These findings demonstrate the potential of integrating TDA with machine learning techniques for smartphone-based activity recognition and highlight the importance of classifier selection when working with topological feature representations.
Keywords: Topological Data Analysis (TDA), Human Activity Recognition (HAR), Persistent Homology, Multi-class Classification and Topology-based Machine Learning.

1. Introduction

Human Activity Recognition (HAR) has become a significant research domain in machine learning, pattern recognition, and ubiquitous computing due to its broad applicability in healthcare monitoring, elderly assistance systems, fitness tracking, smart homes, and context-aware applications. The widespread use of smartphones with built-in sensors like gyroscopes and accelerometers has made it possible to continuously track people's motions, producing vast amounts of sensor data that may be used to categorize and identify everyday activities. As a result, both academic researchers and business professionals have given the creation of precise and effective HAR systems a great deal of attention.

Statistical, temporal, and frequency-domain properties that are derived from sensor inputs are the mainstay of traditional HAR techniques. Machine learning techniques are then used to process these features in order to categorize human behaviors. Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests (RF), which have shown excellent performance in a variety of activity identification tasks, are among the most widely used classification approaches. The literature on pattern recognition and contemporary machine learning frameworks go into great detail about the theoretical underpinnings and real-world applications of these machine learning techniques [4], [8].

Even though traditional feature extraction techniques have shown promise, they frequently fall short of capturing the underlying geometric and structural features present in complex datasets. As a result, different data representation methods that can uncover hidden correlations in high-dimensional data have been investigated by researchers more and more. Topological Data Analysis (TDA), a mathematical framework that uses ideas from algebraic topology to examine the form and connectedness of data, is one such method. TDA has emerged as a

powerful tool for extracting robust and interpretable features from complex datasets by identifying topological structures that remain stable across multiple scales [5].

The idea of topological persistence, which was developed to separate significant topological patterns from noise in datasets, serves as the theoretical basis for TDA [1]. One of the most popular methods in TDA is persistent homology, which offers a multiscale description of loops, voids, connected components, and other topological properties seen in data [3]. Several mathematical structures, such as witness complexes and Vietoris–Rips complexes, have been devised to make persistent homology computations easier [2], [6]. Edelsbrunner and Harer [7] provided detailed descriptions of computational topology and persistent homology, while later research [13] thoroughly examined the stability aspects of persistence modules.

Various representations that allow topological information to be included into machine learning workflows have been developed as a result of recent developments in TDA. While persistence pictures and kernel-based methods convert persistence diagrams into vectorized representations appropriate for classification problems [15], [19], persistence landscapes offer a statistical framework for examining topological properties [11]. Additionally, TDA-based feature extraction has been shown to be successful in complicated pattern recognition tasks using deep learning models that incorporate topological signatures [16]. The actual implementation of TDA in data analysis and machine learning research has also been made easier by the growing availability of computational tools like the GUDHI library [10]. The work of Otter et al. [17] provides a thorough overview of computational techniques and applications of persistent homology.

Smartphone-based datasets are already common benchmarks for assessing classification algorithms and feature extraction methods in the context of Human Activity Recognition. One of the most important HAR datasets was presented by Anguita et al. [9] and has been widely used in activity recognition studies. Reyes-Ortiz et al. [14] then showed how machine learning techniques can effectively identify human activity and transitions from smartphone sensor data. Even while HAR research has made great strides, the majority of current studies concentrate mostly on statistical and signal-processing characteristics, with the use of topological descriptors being relatively understudied.

For HAR applications, where sensor measurements frequently display intricate geometric structures, TDA's capacity to describe the fundamental geometry of data makes it more appealing. Numerous studies have shown that topological representations can enhance classification performance by collecting structural information that traditional feature extraction techniques might miss [18], [20]. However, further research is needed to determine how well various persistence interval selection techniques work and how they affect machine learning classifiers in HAR issues.

Inspired by these findings, this study uses sensor data from smartphones to explore a TDA-based framework for Human Activity Recognition. Researchers at the University of Genoa in Italy created the dataset used in this study, which includes 5,000 samples that reflect ten daily activities, namely Walking, Standing, Sitting, Running, Lying Down, Jumping, Driving, Descending Stairs, Cycling, and Climbing Stairs. Each sample consists of 40 numerical features derived from accelerometer and gyroscope measurements obtained from smartphones positioned at the waist of participants.

Three sample strategies—RandInt, MaxInt, and AvgInt—are used in the suggested framework to choose representative persistence intervals and extract topological features via persistent homology. SVM, KNN, and Random Forest classifiers built in Python are then used to classify the generated feature representations. According to experimental results, RandInt performs noticeably worse than MaxInt and AvgInt sampling strategies, which both reach 100% classification accuracy across all assessed classifiers. Additionally, when paired with topological characteristics, SVM and KNN consistently show higher classification capabilities, demonstrating the discriminative power of TDA-based representations for activity recognition.

2. Materials and Methods

The dataset used in this study, the machine learning classifiers used for activity recognition, the persistence interval sampling methods used to obtain representative topological descriptors, and the Topological Data Analysis framework used for feature extraction are all described in this section.

2.1.Dataset Description

A smartphone-based Human Activity Recognition (HAR) dataset created by researchers at the University of Genoa, Italy, was used for the experiments in this paper. Because of its well-organized and labeled activity records,

this dataset has been used extensively as a standard in machine learning and pattern recognition research [9]. Thirty volunteers, ages 19 to 48, participated in data collection while carrying out ten distinct daily tasks.

Each participant carried a smartphone positioned at the waist, allowing the embedded accelerometer and gyroscope sensors to continuously record body movements. Raw sensor measurements were processed to derive numerical descriptors representing the characteristics of body movement.

This dataset is commonly used to train and evaluate procedures for plan identification & analysis due to its well-structured and labeled data.

Total Samples in Dataset: 5000 **Number of Features in Each Sample:** 40 (derived from accelerometer and gyroscope readings) **Number of Classes:** 10 (Walking, Standing, Sitting, Running, Lying Down, Jumping, Driving, Descending Stairs, Cycling, Climbing Stairs)

2.2.Topological Data Analysis

A mathematical approach called Topological Data Analysis (TDA) uses ideas from algebraic topology to describe the inherent structure of data [5]. TDA examines the connectedness and geometric structure of data points, in contrast to traditional statistical techniques that mainly concentrate on numerical connections. One of the most used tools in TDA is persistent homology, which finds topological characteristics that hold true at many sizes [3]. These features, which together offer a condensed depiction of the underlying geometry of the data, include loops, related components, and higher-dimensional structures.

The following phases make up the overall TDA workflow used in this study:

1. Construction of a simplicial complex from the feature space.
2. Filtration of the simplicial complex.
3. Computation of persistent homology.
4. Extraction of persistence intervals.
5. Selection of representative intervals.
6. Classification using machine learning algorithms.

This framework enables the transformation of sensor-derived numerical features into topological descriptors suitable for classification.

2.3.Persistent Homology

The development and death of topological structures during the filtration process are quantified by persistent homology [1], [3]. Different topological properties arise when simplices are progressively added to the simplicial complex as the filtering parameter increases.

While short-lived features are frequently linked to noise, features that endure throughout a wide range of filtering values are typically considered important structures [7]. Topological feature extraction is based on the resulting persistence intervals, which offer a multiscale overview of the data.

Persistent homology is especially useful for evaluating complicated sensor datasets since it has proven to be resistant against noise and sampling variability [13], [17].

2.4.Persistence Interval Sampling Techniques

The choice of persistence intervals has a significant impact on topological representations' classification ability. In this work, three persistence interval sampling approaches were examined in accordance with the approach used in topological classification research [20].

2.4.1. **MaxInt Sampling**

From the calculated persistence diagram, RandInt chooses persistence intervals at random. This method may not always capture the most representative topological structures, which could have an impact on classification results even if it offers a straightforward mechanism for interval selection.

2.4.2. **MaxInt Sampling**

MaxInt chooses intervals with the highest persistence values. This approach highlights the most important structures found in the data since large persistence intervals are associated with stable topological features.

2.4.3. **AvgInt Sampling**

AvgInt chooses intervals based on the persistence diagram's average persistence properties. This approach seeks a balance between retaining stable topological information and preserving representative structural variability.

The selected intervals are subsequently transformed into feature vectors suitable for machine learning classification.

2.5. Machine Learning Classifiers

Three supervised machine learning techniques were used to assess the retrieved topological features' efficacy.

2.5.1. **Support Vector Machine**

A potent classification method called Support Vector Machine (SVM) maximizes the margin between classes to create an ideal separation hyperplane [4]. SVM has been frequently used in activity recognition applications because of its capacity to handle nonlinear decision boundaries through kernel functions.

The Radial Basis Function (RBF) kernel was used in this work because it is good at capturing nonlinear correlations between topological features.

2.5.2. **K-Nearest Neighbors**

An instance-based learning technique called K-Nearest Neighbors (KNN) categorizes unknown samples based on the majority class of their closest neighbors [4]. The approach has shown competitive performance in multiple HAR investigations and is easy to deploy.

2.5.3. **Random Forest**

Several decision trees are combined in the Random Forest ensemble learning technique to enhance classification performance and lessen overfitting. Bootstrap samples and randomly chosen feature subsets are used to build individual trees, and majority voting is used to decide the final forecast. Fifty decision trees made up the Random Forest classifier used in this study's tests.

2.6.Experimental Environment

The Python programming language was used to carry out each experiment. Python-based modules were used to merge topological feature extraction techniques and persistent homology computations with machine learning algorithms. The Scikit-learn framework was used to create and assess the classification models [8]. Confusion matrices and graphical visualizations were used to analyze the results in order to evaluate the efficacy of the suggested TDA-based HAR framework.

3. Experimental Results and Discussion

The experimental results from the suggested Topological Data Analysis (TDA)-based Human Activity Recognition framework are shown in this section. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) classifiers are used to examine the impact of persistence interval selection strategies on classification performance. The classification accuracy obtained from confusion matrices is the main focus of the evaluation, and the separability of the generated topological feature representations is investigated using scatter plot visualizations.

3.1. Classification Performances

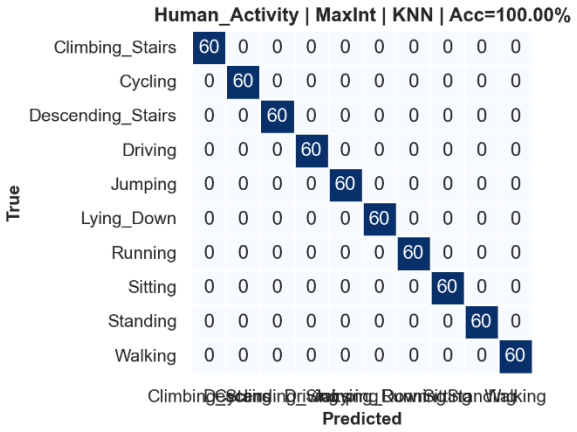
Table 3.1. summarizes the classification accuracies obtained using different sampling strategies and machine learning classifiers.

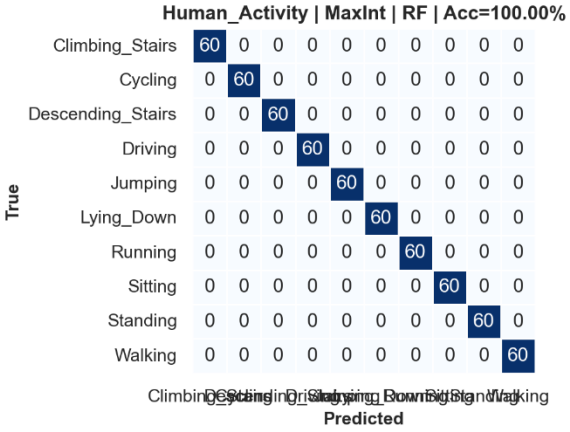
Table 3.1. Metric results for the three classifiers

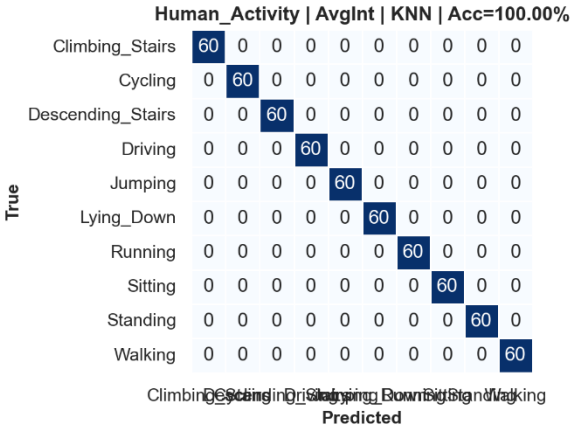
Model	Method	Accuracy	Precision	Recall	F1-Score

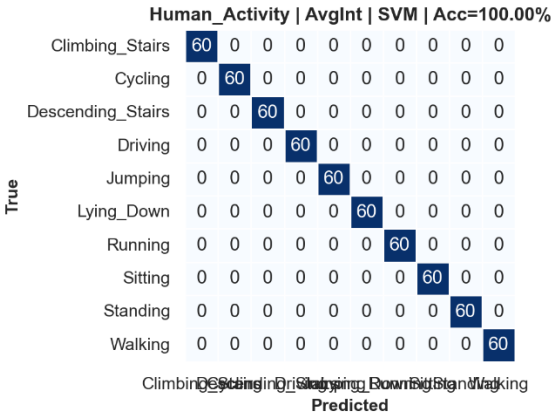
SVM	RandInt	25.56	17.33	25.56	20.63
KNN	RandInt	20.56	23.62	20.56	19.47
RF	RandInt	18.89	19.49	18.89	19.12
SVM	MaxInt	100	100	100	100
KNN	MaxInt	100	100	100	100
RF	MaxInt	100	100	100	100
SVM	AvgInt	100	100	100	100
KNN	AvgInt	100	100	100	100
RF	AvgInt	100	100	100	100

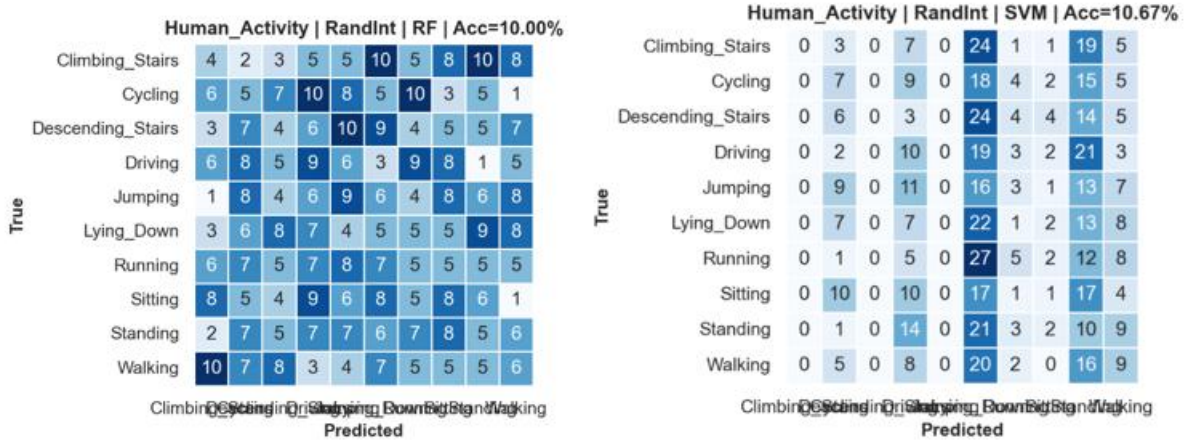
3.2. Confusion Matrix Analysis









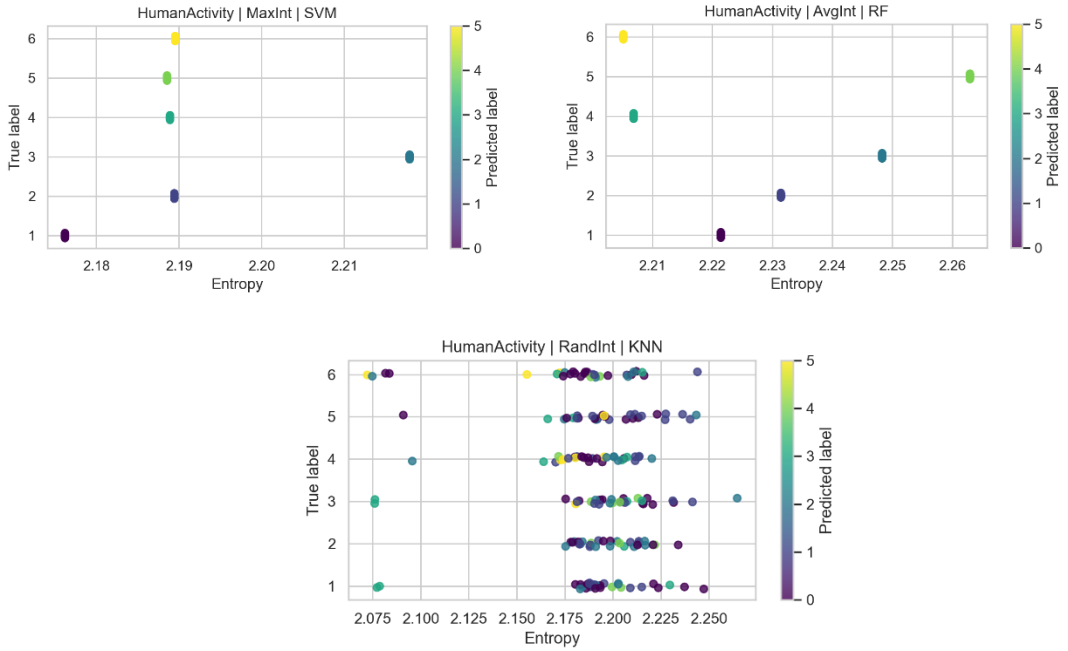


Confusion matrices were used to offer a thorough evaluation of classifier predictions. Significant misclassification between activity classes was noted under RandInt sampling, suggesting insufficient preservation of discriminative topological information.

Confusion matrices for MaxInt and AvgInt sampling, on the other hand, showed flawless diagonal structures, indicating that each activity sample was placed in the appropriate class. The capacity of the chosen persistence intervals to produce highly separable topological representations is confirmed by the lack of off-diagonal elements.

These results suggest that the quality of persistence interval selection has a greater impact on the efficacy of the suggested framework than does the classifier selection.

3.3. Scatter Plot Visualization



In order to examine the degree of separation between activity classes and to depict the distribution of the retrieved topological feature vectors, scatter plots were created.

The scatter plots for RandInt sampling showed significant class overlap, which accounts for all classifiers' comparatively low classification accuracies. The learning algorithms' efficacy was diminished by the lack of distinct cluster boundaries.

On the other hand, scatter plots for MaxInt and AvgInt showed distinct groups with little activity overlap. The improved separability seen in these representations validates the discriminative power of the chosen topological descriptors and supports the confusion matrix results.

3.4. Discussion

The results of the experiment show that the effectiveness of TDA-based Human Activity Recognition is significantly influenced by persistence interval selection. Inappropriate interval selection can significantly impair classification performance, even if persistent homology offers a strong framework for capturing structural features of sensor data.

The poor performance associated with RandInt suggests that random sampling fails to preserve meaningful topological structures necessary for distinguishing activity classes. In contrast, MaxInt and AvgInt effectively exploit stable persistence information, enabling perfect classification across all evaluated machine learning models.

The key finding of this study is that the quality of the topological representation has a bigger impact on recognition performance than the particular classifier used. Once representative topological features are obtained, both distance-based and ensemble-based learning algorithms can accurately discriminate among activities.

These findings underline the significance of choosing suitable persistence interval sampling techniques for building topological descriptors and demonstrate the promise of Topological Data Analysis as an efficient feature extraction paradigm for smartphone-based Human Activity Recognition.

4. Conclusion

This study investigated the effectiveness of Topological Data Analysis (TDA) for smartphone-based Human Activity Recognition using persistent homology-derived features and machine learning classifiers. The suggested approach selected typical persistence intervals using three distinct sampling strategies—RandInt, MaxInt, and AvgInt—after extracting topological features from the sensor data using persistent homology. Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) classifiers built in Python were then used to classify the generated topological feature representations.

Persistence interval selection has a considerable impact on categorization performance, according to experimental results. For SVM, KNN, and Random Forest, RandInt sampling yielded comparatively low recognition accuracies of 25.56%, 20.56%, and 18.89%, respectively. On the other hand, for every classifier that was tested, both the MaxInt and AvgInt sampling methods produced flawless classification accuracy. These results show that the method used to choose persistence intervals has a significant impact on the discriminative power of TDA-based features. The scatter plot visualizations and confusion matrix analysis provided additional evidence that representative interval selection produces highly separable feature spaces, which facilitate precise human activity detection.

The finding that the efficacy of the extracted topological representations has a bigger influence on recognition performance than the classifier selection itself is a significant finding of this study. Various machine learning techniques can effectively differentiate between the activity types once appropriate persistence intervals are chosen. This illustrates Topological Data Analysis's potential as a reliable feature extraction framework for Human Activity Recognition and shows how it can capture structural features that traditional statistical descriptors might not be able to effectively capture.

References

1. Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. Discrete & Computational Geometry, 28(4):511–533, 2002.
2. Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. Proceedings of the Eurographics Symposium on Point-Based Graphics, pages 157–166, 2004.
3. Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. Discrete & Computational Geometry, 33(2):249–274, 2005.
4. Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.
5. Gunnar Carlsson. Topology and data. Bulletin of the American Mathematical Society, 46(2):255–308, 2009.
6. Afra Zomorodian. Fast construction of the Vietoris–Rips complex. Computers & Graphics, 34(3):263–271, 2010.
7. Herbert Edelsbrunner and John Harer. Computational Topology: An Introduction. American Mathematical Society, 2010.
8. Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

9. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pages 437–442, 2013.
10. Clement Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The GUDHI library: Simplicial complexes and persistent homology. Mathematical Software – ICMS 2014, Lecture Notes in Computer Science, 8592:167–174, 2014.
11. Peter Bubenik. Statistical topological data analysis using persistence landscapes. Journal of Machine Learning Research, 16:77–102, 2015.
12. Steve Oudot. Persistence Theory: From Quiver Representations to Data Analysis. American Mathematical Society, 2015.
13. Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The Structure and Stability of Persistence Modules. Springer Briefs in Mathematics, 2016.
14. Jorge L. Reyes-Ortiz, Luca Oneto, Albert Samà, Xavier Parra, and Davide Anguita. Transition-aware human activity recognition using smartphones. Neurocomputing, 171:754–767, 2016.
15. Mathieu Carrière, Marco Cuturi, and Steve Oudot. Sliced Wasserstein kernel for persistence diagrams. Proceedings of the 34th International Conference on Machine Learning, pages 664–673, 2017.
16. Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. Advances in Neural Information Processing Systems Workshops, pages 1–10, 2017.
17. Nina Otter, Mason A. Porter, Ulrike Tillmann, Peter Grindrod, and Heather A. Harrington. A roadmap for the computation of persistent homology. EPJ Data Science, 6(17):1–38, 2017.
18. Mohammad Saadatfar, Hiizu Takeuchi, Yasuaki Hiraoka, Noriyuki Francois, and Itaru Sakaguchi. Pore configuration landscape of granular crystallization. Nature Communications, 8:15082, 2017.
19. Henry Adams and Gunnar Carlsson. Persistence images: A stable vector representation of persistent homology. Journal of Machine Learning Research, 18(8):1–35, 2017.
20. Raul Monserrat, Jose A. Carrasco-Ochoa, Jose Fco. Martinez-Trinidad, and Omar R. Zaiane. A topological data analysis-based classifier for supervised learning. Pattern Recognition Letters, 2021.