forum acusticum 2026
8-12 September · Graz, Austria

TOPOLOGICAL FEATURES FOR ENVIRONMENTAL SOUND CLASSIFICATION

Jiwon Seo[1], Sascha Grollmisch[1], Jaehyun Lee[2], Sukjin Hong[2], Jooyoung Hahn[3]

[1]*Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany*
[2]*LOAS Inc., Seoul, Republic of Korea*
[3]*Department of Software Engineering, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Prague, Czech Republic*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

In this work, we investigate whether topological features derived from persistent homology of Takens embeddings can complement conventional acoustic features for environmental sound classification on the ESC-50 dataset. A compact statistical summary of persistence diagrams is extracted from short overlapping segments, aggregated at the recording level, and combined with conventional acoustic features. Experiments show that fusing topological and acoustic descriptors consistently improves classification performance across multiple classifiers. With a Random Forest classifier, the proposed fusion achieves 48.5% accuracy, an absolute improvement of 7.6 percentage points over the baseline. With a multilayer perceptron, the improvement reaches 12.1 percentage points, yielding 55.1% accuracy. A dimensionality-matched control experiment confirms that these gains arise from informational content, not extended dimensions. A feature-importance analysis further shows that both classifiers rely on the topological descriptors. This work provides empirical evidence that persistent homology captures structural properties of environmental sounds that are complementary to conventional acoustic features.

Keywords: *topological data analysis, persistent homology, environmental sound classification, feature complementarity, Takens embedding*

1. Introduction

Environmental sound classification has become an important research topic in audio signal processing, with applications ranging from smart surveillance to context-aware systems [1]. Benchmark datasets such as ESC-50 [2] have facilitated the development and evaluation of machine learning approaches for environmental sound recognition. Conventional approaches rely on acoustic descriptors derived from time-frequency representations, such as mel-frequency cepstral coefficients (MFCCs) and related features [3], [4], [5]. More recently, large-scale pretrained audio neural networks have achieved remarkable performance on such benchmarks by learning rich representations from mel-spectrograms [6], [7], [8], substantially surpassing conventional feature-based approaches. While these models demonstrate the expressive power of mel-based representations, they rely on large amounts of external training data and substantial computational resources. In contrast, within the conventional feature-based framework, the question of whether descriptors derived from the time-domain signal structure can complement established acoustic features such as MFCCs remains open, and is the main focus of this work.

Topological Data Analysis (TDA) offers methods to analyze complex data by focusing on the topological structure of datasets [9]. One of its central tools, persistent homology, tracks how topological structures, such as connected components and loops, appear and disappear across a range of scales [10]. Applied to time-series data via Takens embeddings [11], persistent homology can reveal structural patterns that conventional signal representations do not capture [12]. This makes TDA a potential tool to capture the dynamical structure of audio signals at the time-domain level.

Several recent studies have explored TDA for audio-related tasks. Persistent homology has been applied to time-domain signals through sliding window embeddings for binary alarm sound detection [13] and to multiple representation spaces (including Takens embeddings, spectrogram surfaces, and spectrogram zeros) for vowel classification [14]. Topological de-

***Corresponding author**: jiwon.seo@idmt.fraunhofer.de, jooyoung.hahn@fjfi.cvut.cz.*

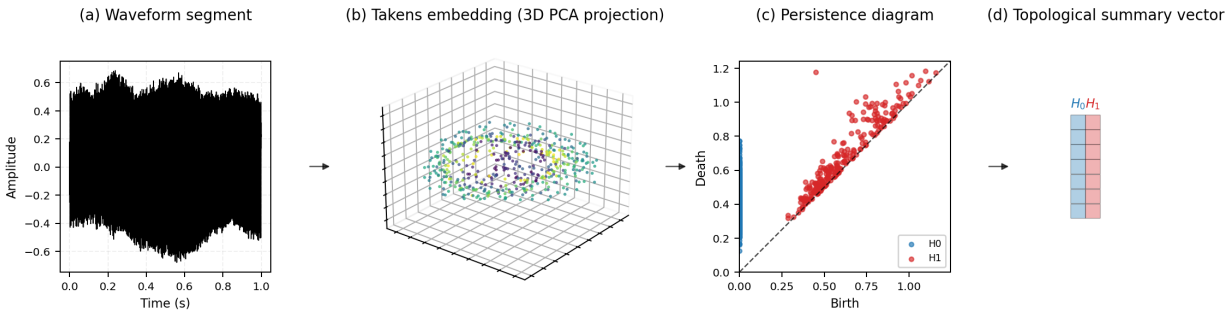

Figure 1: Topology-based feature extraction pipeline for an ESC-50 audio segment (class: *siren*). From left to right: input waveform; Takens embedding projected onto its first three principal components (PCA) for visualization, where axes correspond to PCA components and colors indicate temporal progression from $t = 0\,\mathrm{s}$ in purple to $t = 1\,\mathrm{s}$ in yellow; persistence diagram showing birth–death pairs for H_0 (connected components, shown in blue) and H_1 (loops, shown in red) as two separate diagrams overlaid in a single plot; and topological summary vector containing seven statistics per homology dimension (blue: H_0, red: H_1), yielding a 14-dimensional descriptor.

scriptors have also been derived from spectrogram representations for audio fingerprinting [15]. In the speech domain, persistent homology has been applied to attention maps and embeddings of pretrained transformer models for emotion recognition and speaker verification [16], and Betti curves extracted from waveform filtrations have been used for depression detection [17]. However, these studies have shown that topological descriptors describe meaningful structural patterns in only specific audio domains: periodicity in vowels, temporal regularity in alarms, and global structure in speech embeddings. Most relevant to our work, [13] combined persistent homology of time-domain embeddings with acoustic features such as MFCCs, and reported that this fusion substantially improved alarm detection over acoustic features alone. That study, however, addressed only a single binary task. Multi-class environmental sound classification, where the signal diversity spans quasi-periodic (e.g., siren, clock tick), impulsive (e.g., door knock, glass breaking), and stochastic sounds (e.g., rain, crackling fire), remains largely unexplored from a topological perspective.

In this work, we investigate whether a simple topological descriptor, a compact statistical summary of persistence diagrams obtained from Takens embeddings of time-domain audio, can complement conventional acoustic features, namely MFCCs, for environmental sound classification on the ESC-50 dataset. Unlike prior work that employed rich topological vectorizations or operated on spectral representations, we deliberately adopt such a summary per homology dimension, maintaining a comparable dimensionality to the acoustic feature space so that neither modality dominates the fused representation. Our goal is not to maximize absolute accuracy or to benchmark alternative vectorization schemes, but to establish whether persistent homology provides complementary information to standard acoustic features for this task.

2. Topological Feature Extraction

This section describes how each audio segment is converted into a fixed-length topological feature vector, in three steps: a Takens embedding, persistent homology, and a statistical summary of the result. An overview of the pipeline is illustrated in Fig. 1.

2.1. Takens Embedding

To capture the temporal structure of an audio signal, we construct a point cloud, a set of discrete points in a high-dimensional space, using Takens embedding [11]. Given a one-dimensional time series $x(t)$, vectors of the form

$$\mathbf{x}_t = [x(t),\, x(t+\tau),\, x(t+2\tau),\, \ldots,\, x(t+(d-1)\tau)]$$

are constructed by sliding a window along the signal, where d denotes the embedding dimension and τ the time delay. The resulting vectors form a point cloud in $\mathbb{R}^d$. This construction, also known as sliding window transformation of the signal [13], maps the temporal structure of the audio signal into a geometric representation, where recurring patterns in the signal manifest as structured geometric arrangements in the embedding space. The resulting point cloud is illustrated in Fig. 1(b).

2.2. Persistent Homology

To characterize the topological structure of the point cloud, we apply persistent homology (PH) via a Vietoris–Rips filtration [10], constructed by gradually increasing a distance threshold and connecting points whose pairwise distance falls below this threshold. As the threshold grows, topological structures evolve: connected components, each initially corresponding to a single point, progressively merge into larger ones, while loops appear and close up at larger thresholds. Each structure is recorded by its birth value, the threshold at which it first appears, and its death value, the threshold at which it merges with another structure or closes. The difference between these values, known as persistence, reflects how robustly a

structure exists across distance scales.
We compute persistence diagrams for homology dimensions H_0 and H_1. Each point in the persistence diagram represents a distinct birth–death pair recording the appearance and disappearance of a structure during the filtration, and the total number of such points varies depending on the geometric structure of the point cloud. H_0 captures connected components: it reflects how signal segments with similar temporal patterns group together in the embedding space, showing the clustering structure of the reconstructed trajectory. H_1 captures loops: it reflects closed-loop structures in the trajectory that tend to arise from periodic or quasi-periodic patterns in the audio signal, such as oscillatory waveforms or tonal components. An example persistence diagram for H_0 and H_1 is shown in Fig. 1(c).

2.3. Topological Summary Vector

Persistence diagrams contain a variable number of points, determined by the topological structure of each point cloud rather than the embedding dimension, and cannot be directly used as fixed-dimensional feature vectors. Various vectorization schemes have been proposed to address this, including persistence landscapes [18], persistence images [19], and Betti curves [20], which map persistence diagrams to functional or image-based representations. In this work, we adopt simple statistical summaries of the persistence lifetimes to minimize complexity and isolate the contribution of topological information within a standard classification pipeline.
For each persistence diagram, the persistence of each point is defined as $l = \text{death} - \text{birth}$, and the following seven statistics are extracted per homology dimension:

- number of birth–death pairs in the persistence diagram (count),
- sum of persistence values,
- mean of persistence values,
- standard deviation of persistence values,
- three largest persistence values (top_1, top_2, top_3).

These statistics are designed to capture both the overall distribution of birth–death pairs and the most persistent among them, yielding a compact 7-dimensional descriptor per homology dimension. The resulting descriptor is shown in Fig. 1(d).

3. Experimental Setup

3.1. Dataset

Experiments are conducted on the ESC-50 dataset [2], which contains 2,000 environmental audio recordings organized into 50 balanced classes across five major categories: animals, natural soundscapes, human non-speech sounds, domestic sounds, and urban noises. Each recording has a duration of 5 seconds. Following the official evaluation protocol, we perform 5-fold cross-validation using the predefined dataset splits. All recordings are resampled to 16 kHz and peak-normalized prior to feature extraction.

3.2. Feature Extraction

To obtain stable topological descriptors, feature extraction is performed on short overlapping segments rather than on entire recordings. Processing a full 5 s clip as a single input would produce point clouds dominated by silence or background noise, since the target sound often occupies only a fraction of the recording. Shorter segments increase the chance that at least some windows capture the target event and yield point clouds with meaningful geometric structure. This segment-based processing is therefore a requirement for topological feature extraction, and is applied consistently to acoustic features for comparability. Each recording is divided into overlapping segments of 1 s with a hop size of 0.5 s, yielding nine segments per recording. For each segment, features are first computed at the frame level and averaged to produce a single segment-level descriptor. Segment-level descriptors are then aggregated across all segments using mean and standard deviation pooling, producing the final recording-level feature vector. This two-stage aggregation is necessary because topological features require segment-level point clouds; however, for recordings with short transient events (e.g., glass breaking), the target sound may occupy only a few segments and its contribution can be diluted by the subsequent averaging over segments.

Acoustic features.
For each segment, 12 mel-frequency cepstral coefficients (MFCC_{12}) and zero-crossing rate (ZCR) are computed following the baseline of [2], using a frame size of 400 samples (25 ms) and a hop size of 160 samples (10 ms), yielding a 13-dimensional segment-level descriptor. Mean and standard deviation pooling across segments produces a 26-dimensional recording-level feature vector. We adopt this compact acoustic baseline, denoted MFCC_{12} + ZCR, both because it provides a well-established reference point for ESC-50 and because its dimensionality is comparable to the topological descriptor, enabling a controlled comparison of information content. To match the dimensionality of the proposed fusion, we additionally compute an extended acoustic set, MFCC_{26} + ZCR, using 26 coefficients (54 dimensions after pooling).

Topological features.
For each segment, Takens embedding is applied directly to the waveform with embedding dimension $d = 32$ and time delay $\tau = 1$ sample. The embedding dimension must be sufficiently large to unfold the signal trajectory and reveal its geometric structure, while an excessively large value increases the computational cost of persistent homology; the time delay must be small enough to preserve local temporal coherence [21], [22]. Embedding vectors are generated by sliding a window of $d = 32$ samples with a stride of 16 samples (1 ms at 16 kHz) along each 1 s segment, yielding approximately 1,000 points per seg-

Table 1: Classification accuracy (%) on ESC-50 (mean ± standard deviation over the official 5-fold cross-validation) for a Random Forest (RF) and a multilayer perceptron (MLP), across three feature groups: the acoustic features ($MFCC_{12}$ + ZCR, with $MFCC_{26}$ + ZCR as a dimensionality-matched control), the topological features (PH (H_0), PH (H_1), and the concatenation PH (H_0,H_1)), and their fusion. $MFCC_{12}$ and $MFCC_{26}$ use 12 and 26 mel-frequency cepstral coefficients, respectively; ZCR is the zero-crossing rate; Dim is the feature-vector dimensionality at the recording level, after pooling the segment-level features. Bold marks the best-performing configuration.

Feature	Dim	RF (%)	MLP (%)
Acoustic features			
$MFCC_{12}$ + ZCR	26	40.9 ± 1.3	43.0 ± 2.9
$MFCC_{26}$ + ZCR	54	44.4 ± 0.9	44.5 ± 3.6
Topological features			
PH (H_0)	14	23.5 ± 1.4	26.7 ± 1.9
PH (H_1)	14	24.7 ± 1.4	24.5 ± 1.8
PH (H_0,H_1)	28	33.3 ± 2.1	36.3 ± 2.5
Feature fusion			
$MFCC_{12}$ + ZCR + PH (H_0)	40	44.9 ± 1.9	50.7 ± 3.5
$MFCC_{12}$ + ZCR + PH (H_1)	40	46.5 ± 1.6	50.0 ± 2.4
$MFCC_{12}$ + ZCR + PH (H_0,H_1)	**54**	$\mathbf{48.5 \pm 1.9}$	$\mathbf{55.1 \pm 1.9}$

ment. PH up to H_1 is computed on the resulting point clouds using Vietoris–Rips filtration with Euclidean distance [23]. Following the summary procedure described in Section 2.3, each persistence diagram is summarized into a 7-dimensional vector per homology dimension, yielding a 14-dimensional segment-level topological descriptor. Mean and standard deviation pooling across segments produces a 28-dimensional recording-level topological feature vector. We denote this recording-level vector by PH (H_0,H_1), and its two 14-dimensional single-homology halves by PH (H_0) and PH (H_1). Alternative parameter configurations (e.g., different numbers of top persistence values) were also evaluated; in all cases, classification accuracy was similar to or lower than that obtained with the selected settings.

Feature fusion.
For fusion experiments, acoustic and topological recording-level feature vectors are concatenated. Prior to classification, each feature block is standardized independently using z-score normalization, with mean and standard deviation estimated from the training folds.

3.3. Classification and Evaluation

We evaluate all feature configurations using two classifiers. A Random Forest (RF) with 1,000 trees and no depth restriction is used as the primary classifier, following the evaluation protocol of the original ESC-50 benchmark [2]. Among conventional classifiers including support vector machines and k-nearest neighbors, RF achieved the best performance in the original study and is therefore adopted here. A multilayer perceptron (MLP) with two hidden layers of 256 and 128 units, ReLU activations, batch normalization, and dropout is additionally evaluated to assess whether the observed improvements generalize across different classifier assumptions. The MLP is trained using the Adam optimizer with a learning rate of 10^{-3} for up to 100 epochs with early stopping, using an 80/20 stratified split of the training folds for validation. The same MLP architecture and training hyperparameters were fixed across all feature configurations without scenario-specific tuning. Performance is reported as mean classification accuracy over the five official ESC-50 folds. To quantify how much each feature group ($MFCC_{12}$, ZCR, PH (H_0), PH (H_1)) contributes, we compute two importance measures. First, for the RF we report the mean decrease in impurity (MDI) [24], summed within each group. Second, because the MDI is biased toward high-cardinality features [25], we report a group permutation importance [26], computed for both classifiers as the accuracy drop when each feature group is jointly permuted on the test fold, where a larger drop marks a more important group.

4. Results

Table 1 summarizes classification accuracy on ESC-50 across all feature configurations and classifiers. Note that our segment-level extraction and aggregation (Section 3.2) give an RF baseline of 40.9%, below the 44.3% clip-level baseline reported for ESC-50 [2]; our results are therefore not directly comparable to that benchmark.

Using only conventional acoustic features ($MFCC_{12}$ + ZCR), the RF baseline achieves 40.9% and the MLP baseline achieves 43.0%. Topological features alone yield lower but non-trivial performance, with PH (H_0) and PH (H_1) at 23.5–26.7%. Combining H_0 and H_1, PH (H_0,H_1), achieves 33.3% (RF) and 36.3% (MLP), substantially outperforming PH (H_0) and PH (H_1) individually. This suggests that H_0 and H_1 represent complementary structural information, consistent with their topological roles (Section 2.2): H_0 describing the clustering structure of the reconstructed trajectory, and H_1 its cyclic patterns.

Fusing topological and acoustic features consistently improves performance across both classifiers. The best results are obtained with $MFCC_{12}$ + ZCR + PH (H_0,H_1), reaching 48.5% with RF and 55.1% with MLP, corresponding to absolute improvements of 7.6 and 12.1 percentage points over the respective base-

Table 2: Random Forest (RF) and multilayer perceptron (MLP) feature-group importance on ESC-50 for the proposed fusion ($MFCC_{12}$ + ZCR + PH (H_0,H_1)), over the official 5-fold cross-validation. MDI is the mean decrease in impurity, Perm. the group permutation importance (the accuracy drop when a group is jointly permuted). Values are means; standard deviations are omitted (all at most 0.02). The topological feature rows are shown in bold.

Feature	RF MDI	RF Perm.	MLP Perm.
$MFCC_{12}$	0.46	0.25	0.43
ZCR	0.04	0.02	0.06
$MFCC_{12}$ + ZCR	0.51	0.29	0.45
PH (H_0)	**0.23**	**0.11**	**0.26**
PH (H_1)	**0.26**	**0.25**	**0.24**
PH (H_0,H_1)	**0.49**	**0.36**	**0.37**

lines. The larger gain observed with MLP suggests that nonlinear interactions between acoustic and topological features contribute to additional performance improvements.

To verify that the performance gains are not attributable to the increase in feature dimensionality alone, we compare the proposed fusion $MFCC_{12}$ + ZCR + PH (H_0,H_1) against the control $MFCC_{26}$ + ZCR, both 54-dimensional. While $MFCC_{26}$ + ZCR improves over $MFCC_{12}$ + ZCR by 3.5 (RF) and 1.5 (MLP) percentage points, the proposed fusion yields substantially larger gains of 7.6 (RF) and 12.1 (MLP) percentage points. The gap between the two 54-dimensional configurations indicates that topological descriptors contribute information beyond what additional MFCCs provide.

Both classifiers rely substantially on the topological descriptors (Table 2). These descriptors hold 49% of the RF MDI, and jointly permuting them lowers accuracy by 0.36 (RF) and 0.37 (MLP); permuting all acoustic features together lowers it by 0.29 and 0.45, respectively. The higher-order MFCCs (coefficients 13–26) added by the dimensionality-matched control have a permutation importance of only 0.18 (RF) and 0.19 (MLP), about half those of the topological descriptors (0.36 and 0.37). Unlike the MDI, which distributes one total across the groups and sums to one, the permutation importance is measured independently for each group and does not sum to one.

5. Conclusion and Future Work

In this work, we investigated whether topological features from persistent homology of Takens embeddings, which reconstruct a signal's dynamics as a point cloud from its time-domain samples, complement conventional acoustic features such as MFCCs for environmental sound classification on the ESC-50 dataset. Experiments demonstrated consistent classification improvements across both a Random Forest and a multilayer perceptron classifier, with absolute accuracy gains of 7.6 and 12.1 percentage points over the acoustic baseline, respectively. A dimensionality-matched control experiment confirmed that these gains arise from the informational content of the topological descriptors rather than from the extended feature dimensionality, demonstrating that persistent homology of Takens embeddings captures structure complementary to MFCCs. A feature-importance analysis further supported this, showing that these descriptors were more important to both classifiers than the additional MFCCs of the control. In absolute terms, the accuracy remains below that of large-scale pretrained audio models; the contribution here is a consistent gain within the conventional feature pipeline rather than competitive accuracy.

Several directions remain for future work. First, richer topological vectorization schemes, such as persistence images [19] and persistence landscapes [18], may yield more expressive representations and further improve classification performance. Second, combining topological features with pretrained audio representations [6], [7], [8] might be a promising direction, as such embeddings already capture rich acoustic structure and topological descriptors may provide complementary information at the signal dynamics level. Third, the sensitivity of persistent homology to structural change suggests applications to other tasks, such as industrial anomaly detection.

Acknowledgements

This work was conducted as part of a joint research project between Fraunhofer IDMT and LOAS Inc., funded by the Korea Technology and Information Promotion Agency for SMEs (TIPA) under Grant No. RS-2025-25440174.

References

[1] A. Bansal and N. K. Garg, "Environmental sound classification: A descriptive review of the literature," *Intelligent Systems with Applications*, vol. 16, p. 200 115, 2022.

[2] K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in *Proceedings of the 23rd ACM International Conference on Multimedia*, Brisbane, Australia: ACM Press, Oct. 13, 2015, pp. 1015–1018.

[3] B. Logan, "Mel frequency cepstral coefficients for music modeling," in *International Society for Music Information Retrieval Conference*, 2000.

[4] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," Ircam, Analysis/Synthesis Team, 1 pl. Igor Stravinsky, 75004 Paris, France, Tech. Rep., 2004.

[5] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, pp. 279–283, 2017.

[6] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "Panns: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Au-*

dio, Speech, and Language Processing, vol. 28, pp. 2880–2894, 2020.

[7] S. Chen et al., "BEATs: Audio pre-training with acoustic tokenizers," in *Proceedings of the 40th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 202, PMLR, 2023, pp. 5178–5193.

[8] W. Chen, Y. Liang, Z. Ma, Z. Zheng, and X. Chen, "EAT: Self-supervised pre-training with efficient audio transformer," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, International Joint Conferences on Artificial Intelligence Organization, 2024, pp. 3807–3815.

[9] G. E. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, pp. 255–308, 2009.

[10] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, 2000, pp. 454–463.

[11] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, ser. Lecture Notes in Mathematics, D. Rand and L.-S. Young, Eds., vol. 898, Berlin: Springer, 1981, pp. 366–381.

[12] J. A. Perea and J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 799–838, Jun. 2015.

[13] T. Fireaizen, S. Ron, and O. Bobrowski, "Alarm sound detection using topological signal processing," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 211–215.

[14] G. Bonafos, P. Pudlo, J.-M. Freyermuth, S. Tronçon, and A. Rey, "Topological data analysis of human vowels: Persistent homologies across representation spaces," *Speech Communication*, vol. 178, p. 103 363, 2026.

[15] W. Reise, X. Fernández, M. Dominguez, H. A. Harrington, and M. Beguerisse-Díaz, "Topological fingerprints for audio identification," *SIAM Journal on Mathematics of Data Science*, vol. 6, no. 3, pp. 815–841, 2024.

[16] E. Tulchinskii et al., "Topological data analysis for speech processing," in *Proc. Interspeech 2023*, 2023, pp. 311–315.

[17] M. Tlachac, A. Sargent, E. Toto, R. Paffenroth, and E. Rundensteiner, "Topological data analysis to engineer features from audio signals for depression detection," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2020, pp. 302–307.

[18] P. Bubenik, "Statistical topological data analysis using persistence landscapes," *Journal of Machine Learning Research*, vol. 16, no. 3, pp. 77–102, 2015.

[19] H. Adams et al., "Persistence images: A stable vector representation of persistent homology," *Journal of Machine Learning Research*, vol. 18, no. 8, pp. 1–35, 2017.

[20] F. Chazal and B. Michel, "An introduction to topological data analysis: Fundamental and practical aspects for data scientists," *Frontiers in Artificial Intelligence*, vol. 4, 2021.

[21] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.

[22] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1–2, pp. 43–50, 1997.

[23] U. Bauer, "Ripser: Efficient computation of Vietoris–Rips persistence barcodes," *Journal of Applied and Computational Topology*, vol. 5, no. 3, pp. 391–423, 2021.

[24] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformatics*, vol. 8, p. 25, 2007.

[26] A. Fisher, C. Rudin, and F. Dominici, "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously," *Journal of Machine Learning Research*, vol. 20, no. 177, pp. 1–81, 2019.