2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-7281-8470-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLA51294.2020.00056

Topological Data Analysis to Engineer Features from Audio Signals for Depression Detection*

ML Tlachac
Data Science Department
Worcester Polytechnic Institute
Worcester, MA, USA
mltlachac@wpi.edu

Adam Sargent
Computer Science and Mathematics Departments
Worcester Polytechnic Institute
Worcester, MA, USA
alsargent@wpi.edu

Ermal Toto
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA, USA
toto@wpi.edu

Randy Paffenroth
Data Science and Mathematics Departments
Worcester Polytechnic Institute
Worcester, MA, USA
rcpaffenroth@wpi.edu

Elke Rundensteiner
Data Science and Computer Science Departments
Worcester Polytechnic Institute
Worcester, MA, USA
rundenst@wpi.edu

***Abstract*—Topological Data Analysis (TDA) can be used to extract features from raw signal data. These features, in the form of Betti curves, can be leveraged by machine learning models. Betti curves have been shown to mitigate bias from individual differences when classifying signals. One of the major challenges in detecting depression from audio clips is the variance of audio expression between participants. Thus, we hypothesize that Betti curves could help mitigate this audio expression variance. In this research, we are the first to construct Betti curves from audio signals. We then leverage these Betti curves to detect depression with audio clips from open-ended clinical interviews and scripted crowd-sourced recordings. For both datasets, machine learning models built on Betti curves achieve statistically significantly higher F1, AUC, and Accuracy scores than the same models built on only state-of-the-art audio engineered features. The AUC metric improvement was 0.054 for the clinical interviews and 0.066 for the crowd-sourced audio. Thus, we demonstrate TDA can be useful in screening for depression from audio signals.**

***Index Terms*—Topology, Machine learning, Feature extraction or construction, Audio input/output, Modeling and prediction**

I. Introduction

A. Topological Data Analysis

Topological Data Analysis (TDA) is the practice of looking at topological features to help analyze the data. This is done by using persistent homology to extract and categorize these features. The categorization is accomplished by transforming data into simplicial complexes, then finding the persistence of topological features with filtration. This has mostly been used for analyzing sets of three dimensional types of data but has recently been used to analyze time series [1].

The persistent homology of a filtered complex can be represented as either a persistence barcode or a persistence diagram. There are three main ways to interpret persistence diagrams and barcodes in a way that allows for insertion into machine learning or deep learning models: persistence images, persistence landscapes, and Betti curves. Persistence images and landscapes are both interpretations of persistence diagrams, whereas Betti curves are an interpretation of persistence barcodes [1]. Unlike persistence barcodes, Betti curves can be input into traditional machine learning models [2].

Recently, a study used Betti curves to detect and categorize heart arrhythmias based upon heart beats [2]. These TDA features were leveraged to mitigate the bias due to individual differences with the goal of making the models more robust. In this study, deep learning models were built with and without TDA features extracted from electrocardiogram signals. Overall, including Betti curves improved arrhythmia detection, though the impact was greater for arrhythmia classification [2].

B. Depression Screening with Audio

The integration of audio into society makes it an ideal candidate to passively screen for health issues. This is especially true for mental health conditions, such as depression, where stigma and unrecognized symptoms often impede diagnosis [3]. While depression is the leading cause of disability worldwide [4], it is also among the most treatable mental disorders [5].

Within the healthcare domain, traditional machine learning is often preferred over deep learning given the smaller datasets [6]. While there is a lack of audio datasets with mental health labels [7], audio datasets are common and audio feature engineering has been widely studied. There are many tools for audio feature engineering, such as the state-of-the-art openSMILE feature extraction toolkit [8]. Despite tools like openSMILE, depression screening ability from audio samples remains low [7]. This is currently true even for the Distress Analysis Interview Corpus Wizard-of-Oz (DAIC-WOZ) [9], the benchmark dataset for detection depression from audio. A major challenge identified in screening for mental illnesses with audio data is that audio expression cues vary within and between individuals [10].

This work was supported by the US Department of Education P200A150306 & P200A180088: GAANN grants and NSF III: Small #1910880.

978-1-7281-8470-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMLA51294.2020.00056

C. Approach & Contributions

Electrocardiogram and audio signals are both represented by waves. Thus, as TDA can be applied to electrocardiograms, it can also be applied to audio signals to engineer features in the form of Betti curves. Our hypothesis is that TDA features may be able to mitigate the variance of audio expression between participants and thus result in better depression screening.

In this paper, we apply TDA to classify audio data. Specifically, we explore the potential of Betti curves to screen for depression using two very different audio datasets. We accomplish this by comparing the prediction ability of machine learning models built with only Betti curves, state-of-the-art audio engineered features, and both feature sets. Our research contributions include:

1) constructing filtered complexes from audio waves,
2) leveraging topological features to improve depression screening ability from audio samples, and
3) comparing the prediction ability of Betti curves built from scripted crowd-sourced recordings and open-ended benchmark DAIC-WOZ recordings.

II. Topological Data Analysis Background

Applying TDA to timeseries allows for the extraction of features relating to the shape of the data. These features can then be used by machine learning models to classify the timeseries. The benefit of incorporating TDA features into machine learning models is that these features are theorized to be more robust than alternative features [2]. As audio recordings are impacted by the individual and the quality of the recording, feature robustness is desirable when classifying audio.

In order to extract TDA features suitable for traditional machine learning, the original data must undergo a number of transformations. First, the timeseries are considered as simplical complexes. Then the persistent homology of the filtered simplical complex are extracted. These can be represented with persistence barcodes or a persistence diagrams. As neither of these are suitable input for traditional machine learning models, Betti curves are built from the persistence barcodes. The specifics of these transformations are detailed below.

A. Simplicial Complex

A simplicial complex is defined by a pair $K(V, S)$ where V is a finite set of points, called *vertices*, of K and S is a set of non-empty subsets of V that satisfy the conditions [11]:

1) $p \in V \Rightarrow \{p\} \in S$
2) $\sigma \in S, \tau \subset \sigma \Rightarrow \tau \in S$

Each $\sigma \in S$ is known as a *simplex* of K, thus the entire structure being known as a *simplicial complex* [11]. A polytope is a generalized term for a n-dimensional shape. A simplex is a general term for the smallest regular polytope in a given dimension, such as a triangle or a regular tetrahedron. For simplicial complexes, we look at the connections between these simplices. A filtered simplicial complex, or *filtered complex*, is a nested sequence of simplicial complexes such that

$$K_0 \subset K_1 \subset K_2 \subset \cdots \subset K_n$$

For any K_i, i is known as the *filtration level* of K_i [11]. Notably, a subset of a simplicial complex can contain the same number of points as the original complex but have different connections between the points. Upper-level and sub-level filtered complexes differ by construction. Upper-level filtered complexes start at the global minimum and iteratively add more portions of the wave until the global maximum is reached. Sub-level filtered complexes complete this process in reverse: they start at the global maximum and move to the global minimum.

B. Persistent Homology

Persistent homology observes how the topological features appear and disappear over a filtered complex. The persistence of a feature is defined by a pair (i, j) such that i is the filtration level at which the feature appears and j is the filtration level at which the feature disappears [12]. These are referred to as the *birth time* and *death time*. An example of a feature is whether a component in the complex is connected to the rest of the components; a feature would start when a component is added and end when the component is connected to another component. This type of feature is a 0-dimensional hole.

The persistent homology of an upper-level or sub-level filtered complex can be represented as either a persistence barcode or a persistence diagram [13]. Both persistence representations consider the "birth" and "death" date of the topological features tracked. A persistence barcode represents the distance between a birth and death time as a line between filtration levels. This represents the persistence of each feature as a distinct interval over filtration levels, after which each interval is displayed over filtration levels. A persistence diagram is a plot of the birth dates versus the death dates.

Both persistence barcodes and persistence diagrams provide useful information regarding the shape of the data. Yet, unlike humans who can visually discern important information, the disjoint intervals and points representing the persistence of features do not provide much information to machines regarding how these features interact or compare to one another. Thus, to be interpreted by machines, all the features must be combined into a single data source, such as a Betti curve.

C. Betti Curve

Introduced by Yuhei Umeda [1], Betti sequences interpret the persistence of timeseries data for utilization by machines. Betti curves, visual representations of Betti sequences, have been successfully used for classification of arrhythmias from heartbeats [2]. Betti curves are created from persistence barcodes of a filtered complex. These barcodes are transformed into a single line which can be represented as a 1-dimensional array suitable for traditional machine learning input.

Specifically, a Betti curve represents the sum of lines in a persistence barcode over filtration levels. To form a Betti curve, each line in the barcode is considered as a 1 if active or a 0 if not. Then, the barcode is sampled over the filtration levels at n equally spaced points. At each point, the number of active lines in the barcode is totaled, and added to the curve. This n defines the number of components in the Betti curve. This

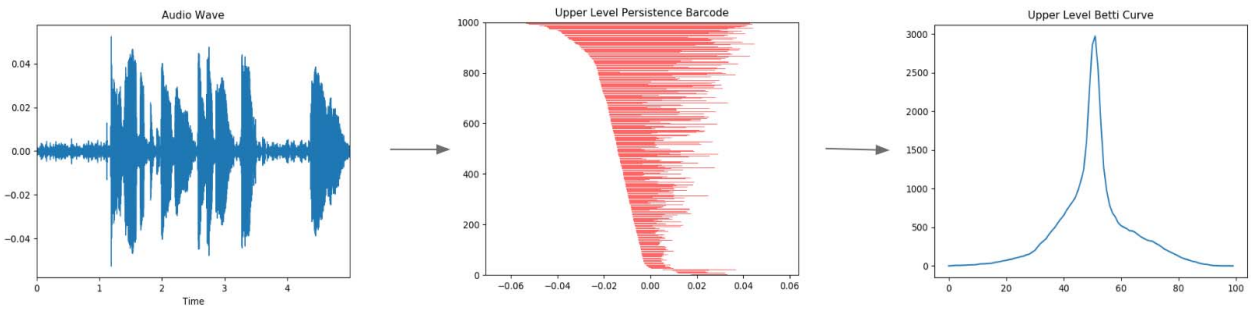

Figure 1: Example of constructing a Betti curve with 100 components from an audio wave using a persistence barcode.

leads to a curve which provides a good linear representation for the original barcode, which has the distinct advantage of being easier to construct than comparable representations.

III. Experimental Methodology & Data

A. Constructing Filtered Complex From Audio

For topological data analysis, data must first be transformed into a filtered simplicial complex, as mentioned in Section II-A. To accomplish this, we leverage Gudhi, a python library for performing topological data analysis. Gudhi uses simplicial trees to efficiently represent simplicial complexes [11]. To construct a filtered complex from sound waves, we consider each segment of the wave as a path between two vertices, then link these paths to create the full wave. Then we assign filtration levels to each segment such that the location of the 0-dimensional holes appear and disappear are the local maxima and minima of the wave. This is done by considering the magnitude of the wave at a specific point as its filtration level.

We consider both upper-level and sub-level filtration curves as the persistence captured can vary slightly based on when features appear and disappear. Filtration from minimum to maximum value will result in an upper-level curve and filtration from maximum to minimum value will result in a sub-level curve. This filtration process ensures the constructed filtered complexes capture the critical points of the sound wave.

The process of constructing filtration levels from the Simplicial complex in Figure 2 is illustrated in Figure 3. Before filtration is applied, we can observe the wave as a hole. This is example is for constructing an upper-level filtered complex so we begin with the smallest values as seen in Figure 2.

At each step, we iteratively add the smallest values remaining until the maximum value is reached. In the second graph of Figure 3 the first segment of the simplicial complex on the left is added, creating the first hole. This hole is continued over multiple filtration levels until it is finally closed at the final filtration level X_4 as seen in the last graph of Figure 3.

These filtration levels are converted into barcodes and then into Betti curves, as depicted in Figure 1. The construction of Betti curves from barcodes is discussed in Section II-C. Thus, in this manner, we are able to extract persistence features appropriate for machine learning models from audio waves.

B. Data

In this research, we leverage audio from two very different datasets. The DAIC-WOZ dataset [9], [14] contains clinical interviews and is a benchmark in depression detection. Each of the 138 participants respond to a varied series of questions from a virtual interviewer. The second dataset includes audio recordings from the crowd-sourced Moodable [15] and EMU [16] databases,further called Moodable/EMU. This data was collected on Mechanical Turk, a crowd-sourcing platform, from 2017-2019 under IRB [15]. Participants were prompted to read: "The quick brown fox jumped over the lazy dog" or "That which we call a rose by any other name would smell as sweet". 230 read the former prompt and 60 read the latter prompt.

Each recording in Moodable/EMU is labeled with a Patient Health Questionnaire-9 (PHQ-9) score. The PHQ-9 is a depression screening survey with nine questions [5]. As each question has options from 0 to 3, PHQ-9 scores range from 0 to 27. Participants in DAIC-WOZ are labeled with PHQ-8 scores. As seen in Figure 4, the datasets have different PHQ distributions. The PHQ-8 is interpreted the same as the PHQ-9 but lacks the ninth question. As score of at least 10 indicates depression [5], our goal is to detect participants with PHQ$\geq$ 10.

C. Data Preprocessing & Feature Engineering

Betti curves are cumulative. As such, for Betti curves to capture the relevant differences in audio rather than length of the clip, the signals need to be the same length. As DAIC-WOZ contained longer audio recordings, we extracted the first five seconds of consecutive audio for each of the 135 participant with that quantity of consecutive audio. To retain all 290 Moodable/EMU participants, we construct Betti curves only from the first two seconds of each provided audio sample.

We arbitrarily select 100 filtration levels which result in Betti curves with 100 components for every audio clip. We construct Betti curves from both upper-level and sub-level filtration curves, further referred to as Betti_u and Betti_s, respectively.

We apply openSMILE [8] to the audio, resulting in 2268 features. Each clip is also labeled with gender. Thus, we have 2469 potential features for every audio clip. The five feature sets we utilize in the experiments are summarized in Table I. Given the quantity of features, we use principal component analysis (PCA), a feature reduction method. PCA successively calculates

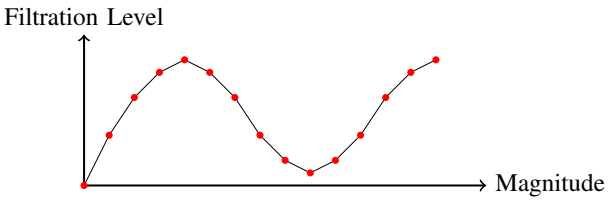

Figure 2: Simplicial complex from a simplified sound wave.

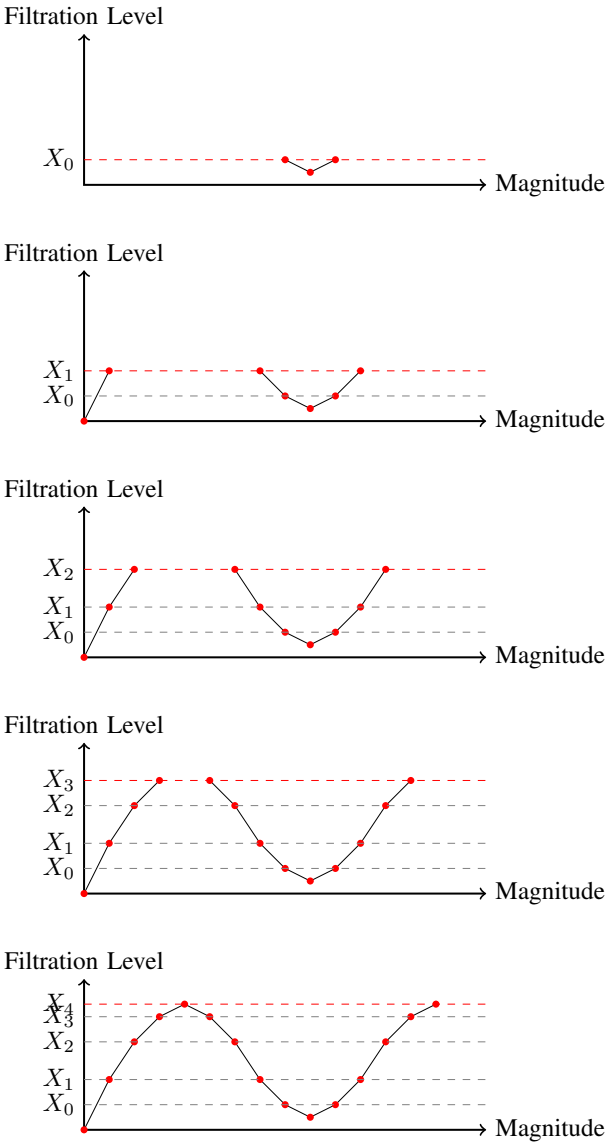

Figure 3: Adding filtration levels to construct a simplicial complex. Starts with just the first filtration level, X_0. At each iterative step, another filtration level is added until the simplicial complex is completed by the fifth filtration level, X_4.

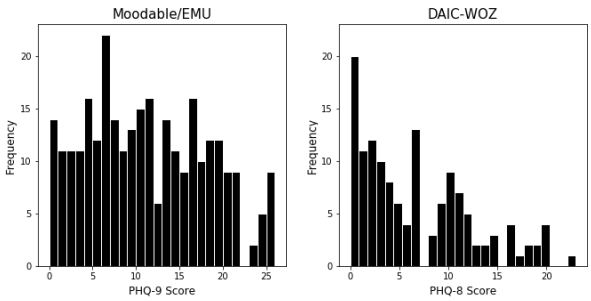

Figure 4: Distribution of participant PHQ scores.

Set Name	Features	nF
Smile	openSMILE & gender	2269
Smile+TDA$_u$	openSMILE & Betti$_u$ & gender	2369
Smile+TDA$_s$	openSMILE & Betti$_s$ & gender	2369
TDA$_u$	Betti$_u$ & gender	101
TDA$_s$	Betti$_s$ & gender	101

Table I: The five sets of features with number of features nF extracted from each audio dataset for use in the experiments.

principal components that explain the maximum amount of variance from the original features [17]. We consider at most 100 principal components for each feature set in Table I.

D. Machine Learning Methodology & Evaluation

We compare the depression screening ability of the feature sets with three common machine learning methods: support vector classifier (SVC), k-nearest neighbor (kNN), and Random Forest (RF). These models were all implemented in Scikit-learn [18] with default parameters. The models are trained on the principal components for each of the five feature sets in Table I with the goal of predicting whether the participant has PHQ$\geq$ 10. Thirty percent of data is reserved to test the model. Downsampling is applied to balance the training data. To ensure result robustness, each experimental procedure is repeated 100 times with different train-test splits.

We evaluate the machine learning models by considering the average $F1$, AUC, and Accuracy (Acc) of the 100 experiments with the same experimental parameters. These metrics are calculated with the number of true positives tp, false positives fp, false negatives fn, and true negatives tn. The $F1$ score, Eq. 1, is the balance between precision= $tp/(tp+fp)$ and recall= $tp/(tp+fn)$. AUC determines classification ability by calculating the area under the ROC curve which is formed by plotting the true positive rate and false positive rate. Accuracy is the portion of correctly classified instances.

$$F1 = \frac{2(precision)(recall)}{precision + recall} \tag{1}$$

By repeating with 100 train-test splits, we generate 100 $F1$, AUC, and Acc scores for each experimental procedure. Thus, we can use a t-test to determine if the metrics for two different experimental procedures are statistically significantly different.

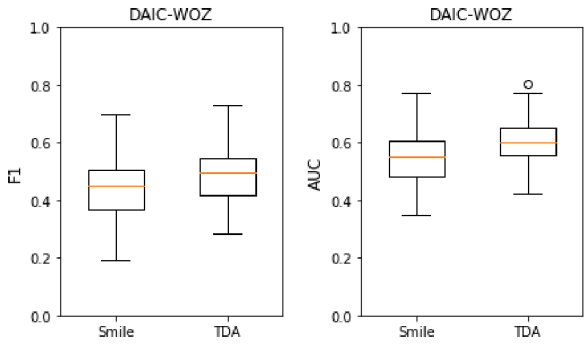

Figure 5: Distribution of $F1$ and AUC scores for best kNN models for the DAIC-WOZ dataset. The $Smile$ models leverage 25 principal components while the TDA_s models leverage 10.

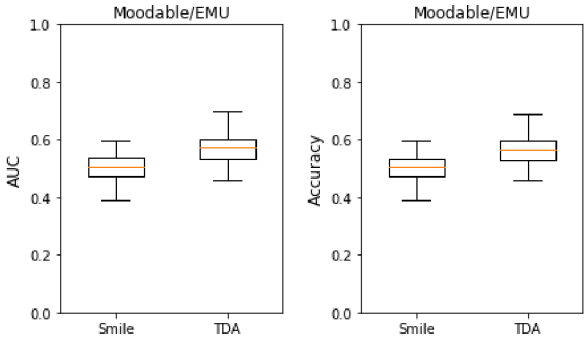

Figure 6: Distribution of AUC and Accuracy scores for best SVC models for the Moodable/EMU dataset. The $Smile$ models leverage 25 principal components while the TDA_s models leverage 5. While the distributions for AUC and Accuracy are very similar, the median is closer to the AUC upper quartile than the Accuracy upper quartile for TDA_s.

E. Tools and Availability

Code and results are available at github.com/mltlachac/TDA. Moodable/EMU audio are available at arcgit.wpi.edu/toto/MoodableEMU. Project updates are at emutivo.wpi.edu.

IV. Results

We compare the average $F1$, AUC, and Accuracy of the 100 models with the same experimental procedure. We train the three machine learning methods with principal components of five feature sets for two audio datasets. For each of these combinations, we compare the models built with the number of principal components that yielded the highest average metric.

First, we compare the depression screening ability of the models built with TDA_u and TDA_s feature sets, as described in Table I. For both audio datasets, models built with TDA_s performed better than models built with TDA_u for every machine learning method and metric. Thus, our analysis proceeds with the feature sets containing sub-level curves.

We compare the performance of three remaining feature sets for DAIC-WOZ and Moodable/EMU in Tables II and III,

Table II: Results for machine learning experiments on DAIC-WOZ with sub-level curves. For each line, the average metric is shown for models built with the number of principal components that yielded the highest average. Significance in comparison to models only built with $Smile$ features are indicated with $^{*}p < 0.05$,$^{**}p < 0.01$, $^{***}p < 0.001$.

Method	Metric	Smile	Smile + TDA$_s$	TDA$_s$
SVC	F1	0.479	0.469	0.441**
kNN	F1	0.436	0.434	**0.487*****
RF	F1	0.452	0.457	0.451
SVC	AUC	0.595	0.588	0.557**
kNN	AUC	0.549	0.543	**0.603*****
RF	AUC	0.572	0.579	0.576
SVC	Acc	0.587	0.583	0.548***
kNN	Acc	0.553	0.536	**0.593*****
RF	Acc	0.572	0.579	0.580

Table III: Results for machine learning experiments on Moodable/EMU with sub-level curves. For each line, the average metric is shown for models built with the number of principal components that yielded the highest average. Significance in comparison to models only built with $Smile$ features are indicated with $^{*}p < 0.05$,$^{**}p < 0.01$, $^{***}p < 0.001$.

Method	Metric	Smile	Smile+TDA$_s$	TDA$_s$
SVC	F1	0.535	0.550	0.535
kNN	F1	0.512	0.509	0.543***
RF	F1	0.530	0.541	**0.558****
SVC	AUC	0.503	0.517*	**0.569*****
kNN	AUC	0.500	0.506	0.538***
RF	AUC	0.532	0.539	0.545
SVC	Acc	0.506	0.521*	**0.563*****
kNN	Acc	0.500	0.504	0.536***
RF	Acc	0.531	0.537	0.543

respectively. The performance of models that are statistically significantly different than their $Smile$ counterparts based on t-tests are marked. Models built with TDA_s yield the highest average $F1$, AUC, and Accuracy for both datasets. For DAIC-WOZ, the highest average $F1 = 0.487$, $AUC = 0.603$, and $Acc = 0.593$. For Moodable/EMU, the highest average $F1 = 0.558$, $AUC = 0.569$, and $Acc = 0.563$. All of these averages are statistically significant improvements over the same machine learning method with only the $Smile$ features.

For DAIC-WOZ, kNN models built with the first 10 principal components of TDA_s perform the best for all metrics. For Moodable/EMU, an SVC model built with the first 5 principal components performs best for AUC and Acc, though a RF model built with the first 20 principal components performs best for $F1$. The distribution of metrics for the aforementioned kNN and SVC models with different train-test splits are depicted in Figures 5 and 6 with their $Smile$ counterparts. For DAIC-WOZ, building the kNN model with TDA$_s$ instead of $Smile$

features statistically significantly increases $F1$ by 0.051, AUC by 0.054, and Acc by 0.040. For Moodable/EMU, building the SVC model with TDA_s instead of $Smile$ features statistically significantly increases the AUC by 0.066 and Acc by 0.057.

For certain methods and metrics, models built with $Smile$ and $Smile + TDA_s$ exceed the performance of models built with just TDA_s. Only in the instance of SVC on the DAIC-WOZ data do models built with $Smile$ perform statistically significantly better than TDA_s. However, the kNN models built with TDA_s outperform the SVC models built with $Smile$.

V. Discussion

A. Betti Curves vs openSMILE

We were expecting that combining Betti curves with state-of-the-art openSMILE features would achieve the highest results, similar to the arrhythmia study [2]. However, the models built with just the sub-level Betti curves and gender achieved the highest average $F1$, AUC, and Acc for both audio datasets. Ideally, future studies will experiment on how to combine the benefits of traditional audio engineered and TDA extracted audio features. This could be achieved through altering the length of voice clips, number of Betti curve components, feature selection techniques, and/or machine learning methods.

B. Upper Level vs Sub Level Betti Curves

We consider both upper-level and sub-level Betti curves as the captured persistence can vary based on when the topological features appear and disappear during construction. For all experiments, TDA_s had higher average metrics than TDA_u. Thus, we conclude sub-level Betti curves are best for depression detection from audio and suggest sub-level curves for future experimentation in this domain. Whether sub-level Betti curves are better for audio tasks or just this task is worth further study.

C. DAIC-WOZ vs Moodable/EMU

We tested the prediction ability of Betti curves on two very different datasets. Despite this, models built with only sub-level Betti curves and gender achieved the highest $F1$, AUC, and Acc when screening for depression for both datasets. We achieved a higher average $F1$ when screening for depression with Moodable/EMU audio clips. However, the AUC and Acc were higher for DAIC-WOZ. Thus, the lower $F1$ score for the DAIC-WOZ dataset likely originated from the lack of depressed participants, as seen in Figure 4. Additionally, $F1$ increased more for DAIC-WOZ than Moodable/EMU while AUC and Acc increased more for Moodable/EMU than DAIC-WOZ.

VI. Conclusion

Our experiments suggest TDA features may be useful in screening for depression from audio. Machine learning models built with sub-level Betti curves achieved the highest average $F1$, AUC, and Accuracy scores for both the crowd-sourced Moodable/EMU dataset and the clinical interview DAIC-WOZ dataset. These metrics are statistically significantly higher than the same models built with state-of-the-art openSMILE features. We believe the TDA features are more robust to the variance of audio expression between participants which is one of the main challenges in depression detection from audio. When detecting depression with audio, sub-level Betti curves performed better than upper-level Betti curves. This is the first paper to extract Betti curves from audio and we hope our findings inform future experiments. Future work includes exploring the impact of audio clip length and number of components.

Acknowledgment

We thank Prof Agu, Dogrucu, Perucic, Isaro, Ball, Resom, Assan, Flannery, Gao, Wu, Caltabiano, Thant, Pingal, Seifu, and Taye for their contributions to Moodable/EMU. We thank the DSRG research community at WPI for their support.

References

[1] Y. Umeda, "Time series classification via topological data analysis," *Information and Media Technologies*, vol. 12, pp. 228–239, 2017.

[2] M. Dindin, Y. Umeda, and F. Chazal, "Topological data analysis for arrhythmia detection through modular neural networks," in *Canadian Conference on Artificial Intelligence*. Springer, 2020, pp. 177–188.

[3] R. M. Epstein, P. R. Duberstein, M. D. Feldman *et al.*, ""i didn't know what was wrong:" how people with undiagnosed depression recognize, name and explain their distress," *Journal of general internal medicine*, vol. 25, no. 9, pp. 954–961, 2010.

[4] World Health Organization, "Depression," 2018. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[5] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[6] D. B. Dwyer, P. Falkai, and N. Koutsouleris, "Machine learning approaches for clinical psychology and psychiatry," *Annual review of clinical psychology*, vol. 14, pp. 91–118, 2018.

[7] N. Cummins, S. Scherer, J. Krajewski *et al.*, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.

[8] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 835–838.

[9] J. Gratch, R. Artstein, G. M. Lucas *et al.*, "The distress analysis interview corpus of human and computer interviews." in *LREC*, 2014, pp. 3123–28.

[10] C.-N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artificial Intelligence Review*, vol. 43, no. 2, pp. 155–177, 2015.

[11] J.-D. Boissonnat and C. Maria, "The simplex tree: An efficient data structure for general simplicial complexes," *Algorithmica*, vol. 70, no. 3, pp. 406–427, 2014.

[12] H. Edelsbrunner and J. Harer, "Persistent homology-a survey," *Contemporary mathematics*, vol. 453, pp. 257–282, 2008.

[13] P. Bubenik. (2017) Introduction to topological data analysis. [Online]. Available: https://people.clas.ufl.edu/peterbubenik/files/intro_tda_worksheet.pdf

[14] D. DeVault, R. Artstein, G. Benn *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, ser. AAMAS '14. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2014, p. 1061–1068.

[15] A. Dogrucu, A. Perucic, A. Isaro *et al.*, "Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data," *Smart Health*, pp. 100–118, 2020.

[16] M. L. Tlachac and E. Rundensteiner, "Screening for depression with retrospectively harvested private versus public text," *IEEE Journal of Biomedical and Health Informatics*, 2020.

[17] Scikit-learn Developers, "Decomposing signals in components (matrix factorization problems)," 2019. [Online]. Available: https://scikit-learn.org/stable/modules/decomposition.html

[18] F. Pedregosa, G. Varoquaux, A. Gramfort *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning research*, vol. 12, pp. 2825–2830, 2011.