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

ML Tlachac, Adam Sargent, Ermal Toto,
Randy Paffenroth, Elke Rundensteiner

ICMLA 2020

Paper #164

Depression Screening with Audio

Depression is…

A leading cause of disability worldwide

Among the most treatable mental disorders

Often undiagnosed due to lack of symptom recognition

Solution is…

Screen for depression with voice recordings which are prevalent and passive

Challenge is…

Audio expression cues vary within and between individuals

Topological Data Analysis (TDA)

the practice of looking at topological features to help analyze data

Recently extended from 3D data to 2D data

- TDA used to engineer features from heartbeats
- Features used to detect and classify arrythmias
- TDA mitigated bias due to individual differences

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.

Timeseries to Betti Curves

Time Series → Simplicial Complex → Filtered Complex → Persistence Barcode → Betti Curve

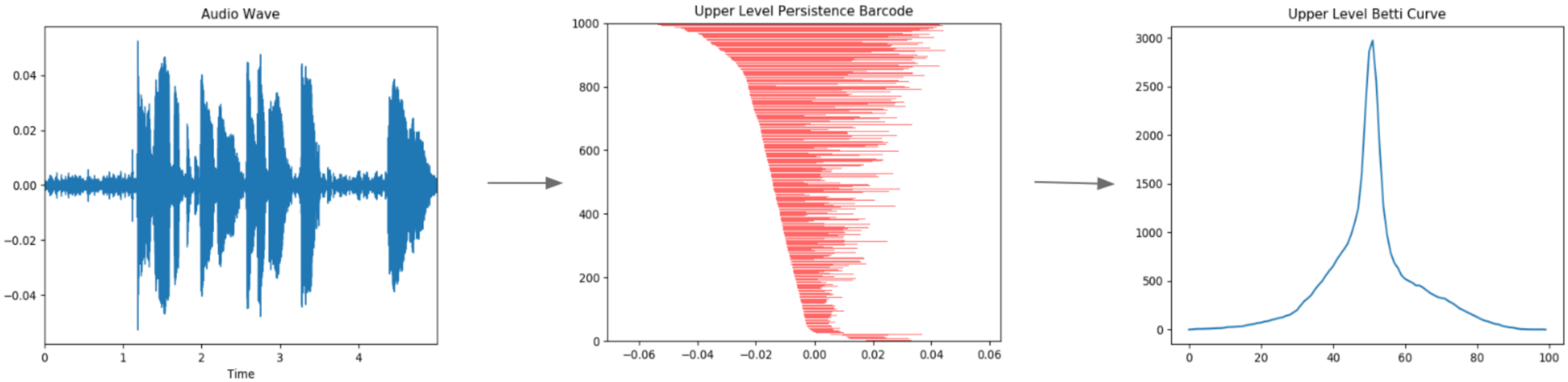
- **Upper-level filtered complexes** start at the global minimum and iteratively add more portions of the wave until the global maximum is reached
- **Sub-level filtered complexes** start at the global maximum and iteratively add more portions of the wave until the global minimum is reached

The persistence of a feature is defined by *(birth time, death time)*

- **Birth time** is the filtration level at which the feature appears
- **Death time** is the filtration level at which the feature disappears

Contributions

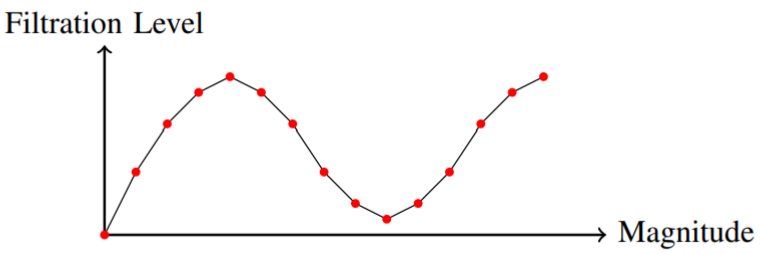
1. Constructing filtered complexes from audio waves
2. Leveraging TDA features for depression screening
3. Comparing the prediction ability of Betti curves built from scripted crowd-sourced recordings and open-ended clinical interviews

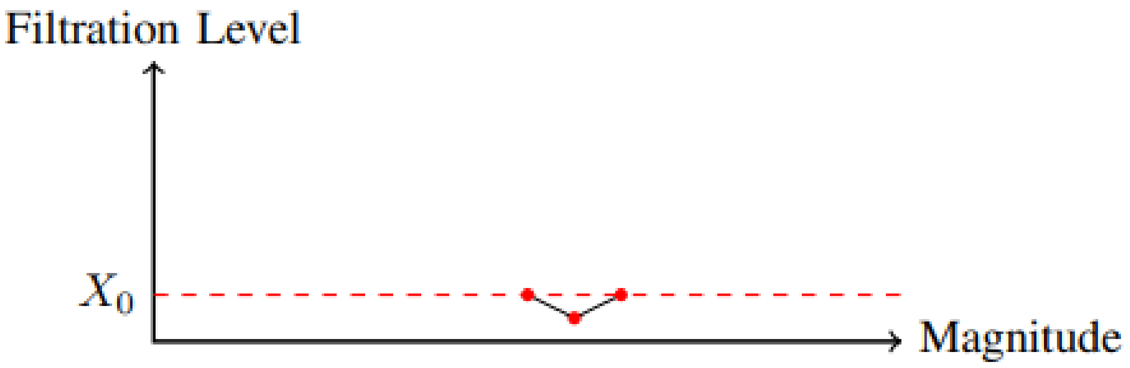


Constructing Filtered Complex From Audio

1. Consider each segment of the wave as a path between two vertices
2. Link these paths to create the full wave
3. Assign filtration levels to each segment such that the location of the 0-d holes appear and disappear are the local maxima and minima of the wave
4. These filtration levels can be converted into barcodes

Upper-level filtered complex from

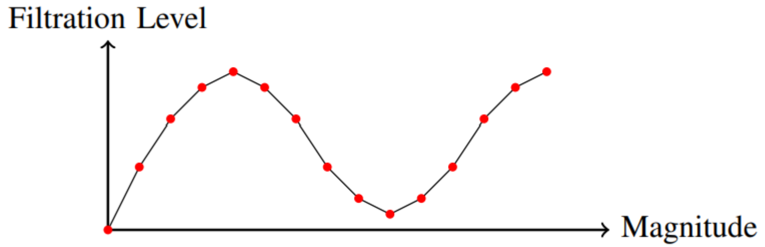


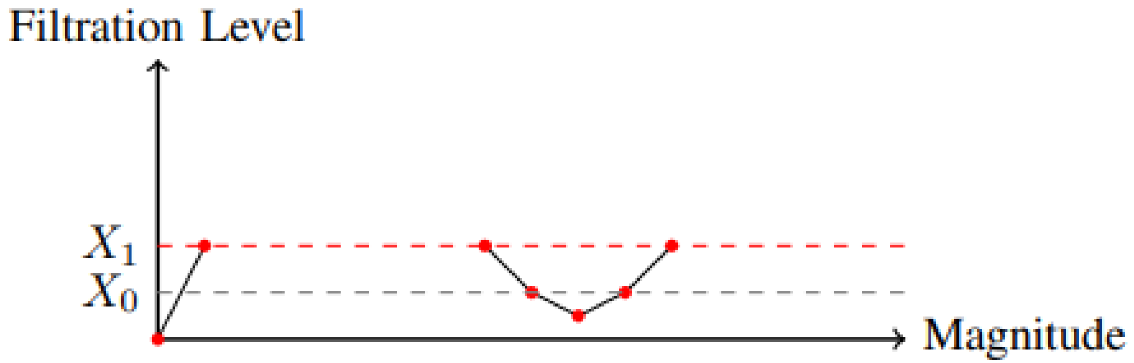


Constructing Filtered Complex From Audio

1. Consider each segment of the wave as a path between two vertices
2. Link these paths to create the full wave
3. Assign filtration levels to each segment such that the location of the 0-d holes appear and disappear are the local maxima and minima of the wave
4. These filtration levels can be converted into barcodes

Upper-level filtered complex from

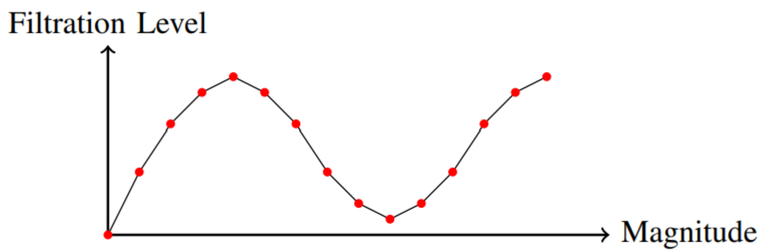


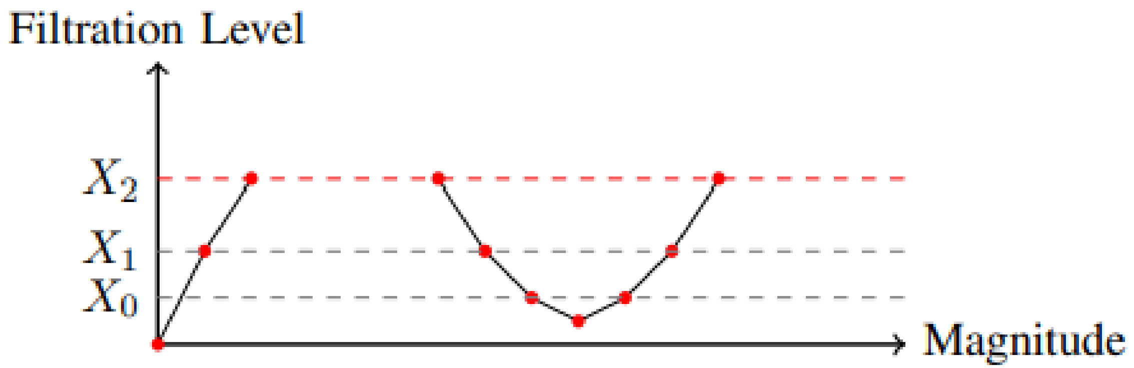


Constructing Filtered Complex From Audio

1. Consider each segment of the wave as a path between two vertices
2. Link these paths to create the full wave
3. Assign filtration levels to each segment such that the location of the 0-d holes appear and disappear are the local maxima and minima of the wave
4. These filtration levels can be converted into barcodes

Upper-level filtered complex from

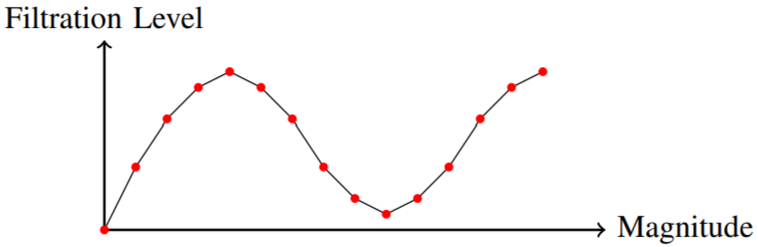


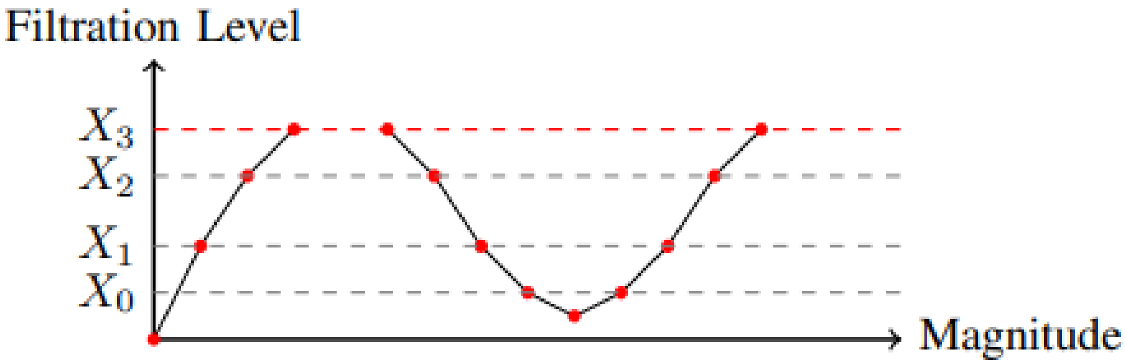


Constructing Filtered Complex From Audio

1. Consider each segment of the wave as a path between two vertices
2. Link these paths to create the full wave
3. Assign filtration levels to each segment such that the location of the 0-d holes appear and disappear are the local maxima and minima of the wave
4. These filtration levels can be converted into barcodes

Upper-level filtered complex from

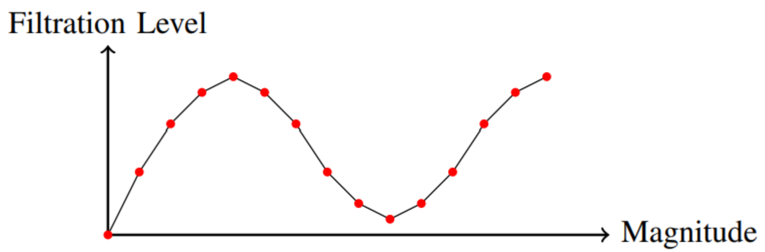


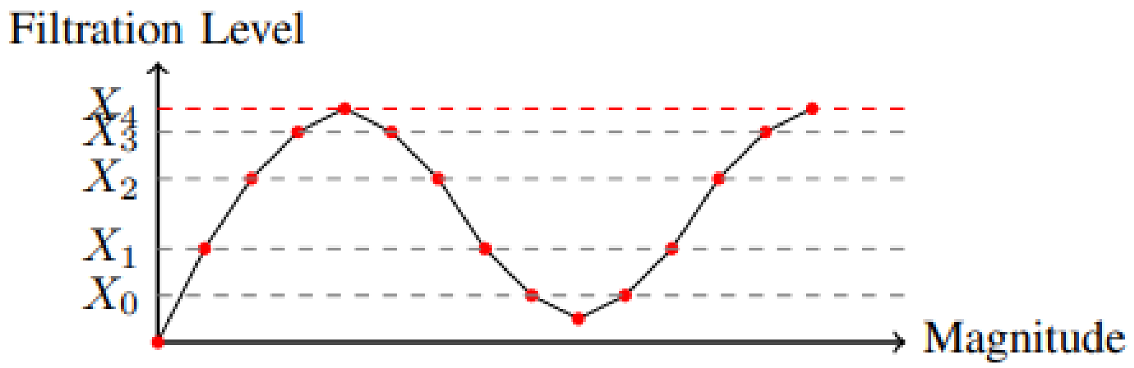


Constructing Filtered Complex From Audio

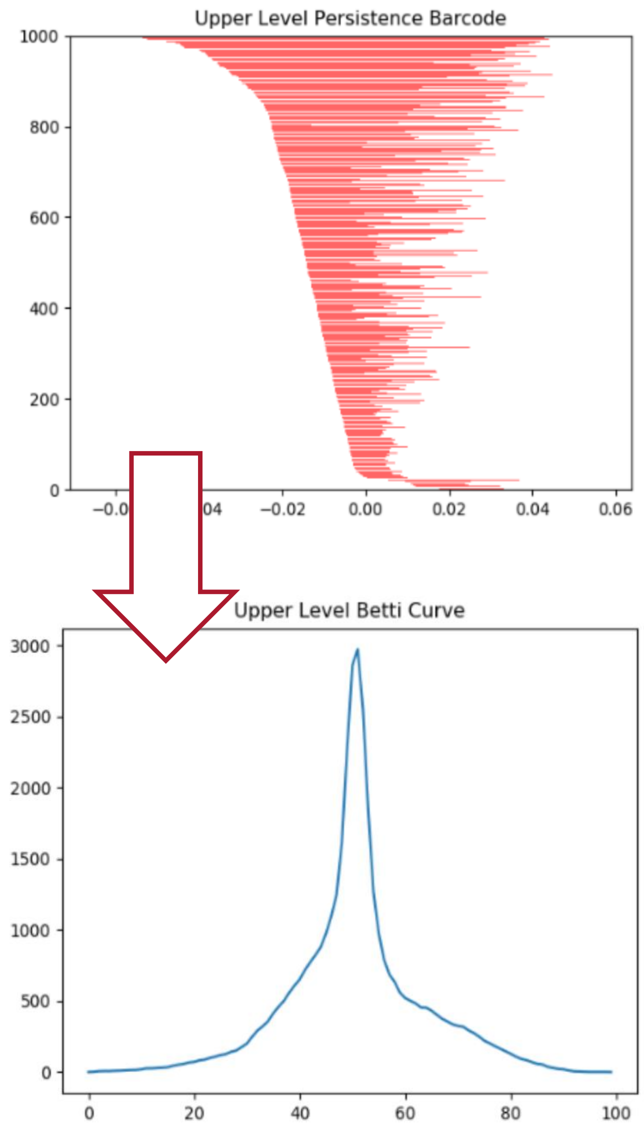
1. Consider each segment of the wave as a path between two vertices
2. Link these paths to create the full wave
3. Assign filtration levels to each segment such that the location of the 0-d holes appear and disappear are the local maxima and minima of the wave
4. These filtration levels can be converted into barcodes

Upper-level filtered complex from





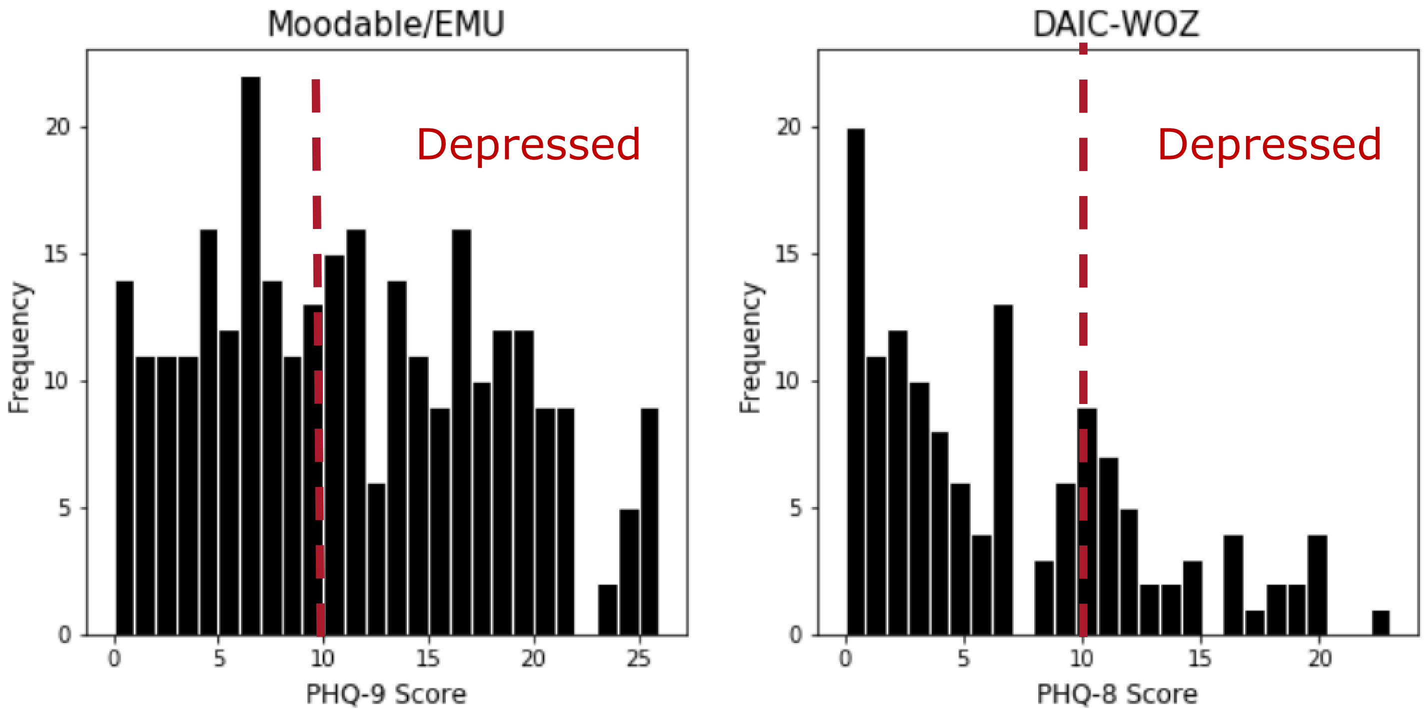
Constructing Betti Curves From Barcodes



1. Each line in the persistence barcode is considered as a 1 if active or a 0 if not
2. Barcode is sampled over the filtration levels at *n* equally spaced points
3. At each point, the number of active lines in the barcode is totaled and added to the curve
4. *n* becomes the number of components in the Betti curve

Data

- Moodable/EMU
  - 290 crowd-sourced participants
  - Read a sentence
  - We use the first 2 seconds
- DAIC-WOZ
  - 135 participants
  - Responded to clinical interview questions
  - We use the first 5 seconds



Sets of Features for Machine Learning

1
gender
feature

2268
openSMILE*
features

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

100
upper-level
Betti curve
features

100
sub-level
Betti curve
features

*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.

Machine Learning Experiments

- Feature Selection
  - Principal component analysis (PCA) with up to 100 principal components
- Machine Learning Methods
  - Support Vector Classifier (SVC)
  - k-Nearest Neighbor (kNN)
  - Random Forest (RF)
- Evaluation Metrics
  - F1 score
  - AUC
  - Accuracy (Acc)

Each experiment is repeated 100 times with different train-test splits

Results

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

Models with Highest Average Metrics

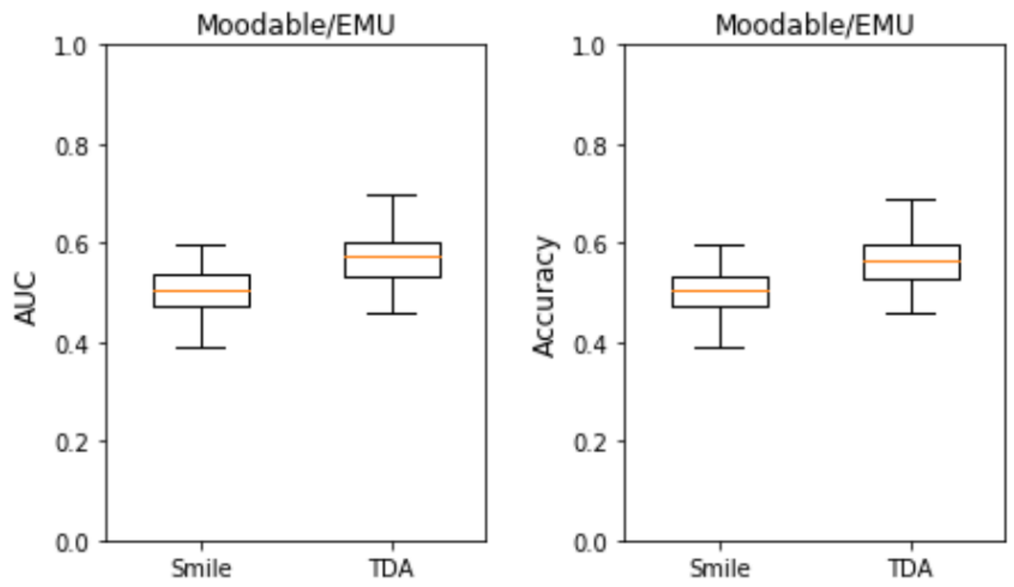


Figure 6: Distribution of AUC and Accuracy scores for best SVC models for the Moodable/EMU dataset. The $Smile$ models leverage 25 principal components while the TDA_s models leverage 5. While the distributions for AUC and Accuracy are very similar, the median is closer to the AUC upper quartile than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

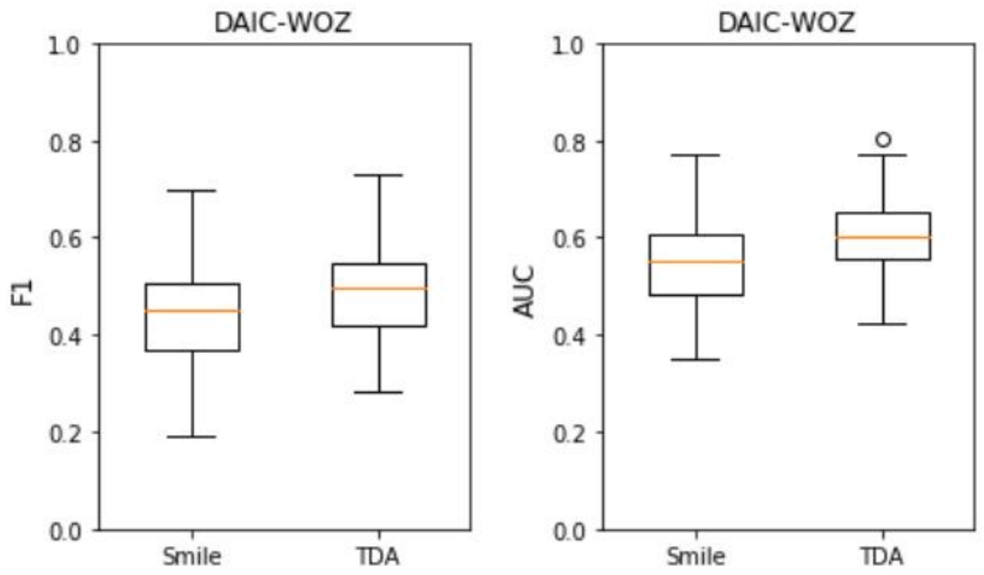


Figure 5: Distribution of $F1$ and AUC scores for best kNN models for the DAIC-WOZ dataset. The $Smile$ models leverage 25 principal components while the TDA_s models leverage 10.

Conclusion

Takeaways

- TDA features may be **useful** in screening for depression from audio
- Models built with TDA features achieved **higher metrics** than models built with state-of-the-art audio engineered features
- **Sub-level Betti curves** performed better than upper-level Betti curves for this task

Future Directions

- Experiment with altering the length of voice clips and number of Betti curve **components**
- Experiment with different feature selection techniques and machine learning methods
- Explore why **sub-level curves** performed better than upper-level curves
- Apply to similar domain, such as **emotion classification**

Questions?

mltlachac@wpi.edu

www.linkedin.com/in/mltlachac

www.github/mltlachac

www.emutivo.wpi.edu

Acknowledgements:

- US Department of Education P200A150306 & P200A180088: GAANN grants for funding
- NSF III: Small \#1910880 for funding
- Prof. Agu, Dogrucu, Peruic, Isaro, Ball, Gao, Flannery, Resom, Assan, Wu, and Thant at WPI for their contributions to Moodable/EMU
- DSRG community at WPI for their support