INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Emotion Recognition from Natural Phone Conversations in Individuals With and Without Recent Suicidal Ideation

John Gideon[1], *Heather T Schatten*[2,3], *Melvin G McInnis*[4], *Emily Mower Provost*[1]

[1]Department of Computer Science and Engineering, University of Michigan, USA
[2]Department of Psychiatry and Human Behavior, Brown University, USA
[3]Psychosocial Research Program, Butler Hospital, USA
[4]Department of Psychiatry, University of Michigan, USA

{gideonjn, mmcinnis, emilykmp}@umich.edu, heather_schatten@brown.edu

Abstract

Suicide is a serious public health concern in the U.S., taking the lives of over 47,000 people in 2017. Early detection of suicidal ideation is key to prevention. One promising approach to symptom monitoring is suicidal speech prediction, as speech can be passively collected and may indicate changes in risk. However, directly identifying suicidal speech is difficult, as characteristics of speech can vary rapidly compared with suicidal thoughts. Suicidal ideation is also associated with emotion dysregulation. Therefore, in this work, we focus on the detection of emotion from speech and its relation to suicide. We introduce the Ecological Measurement of Affect, Speech, and Suicide (EMASS) dataset, which contains phone call recordings of individuals recently discharged from the hospital following admission for suicidal ideation or behavior, along with controls. Participants self-report their emotion periodically throughout the study. However, the dataset is relatively small and has uncertain labels. Because of this, we find that most features traditionally used for emotion classification fail. We demonstrate how outside emotion datasets can be used to generate more relevant features, making this analysis possible. Finally, we use emotion predictions to differentiate healthy controls from those with suicidal ideation, providing evidence for suicidal speech detection using emotion.

Index Terms: emotion recognition, suicidal speech, small data, uncertainty, neural networks

1. Introduction

Suicide is an increasingly serious public health issue, with the suicide rate increasing from 10.46 to 14.48 deaths per 100,000 between 1999 and 2017 [1]. A recent meta-analysis suggests that our ability to predict suicide is only slightly above chance levels, and has not improved over the past 50 years [2]. Early detection of suicidal ideation is crucial for prevention and intervention. However, relying on self-report of suicide risk is problematic, as the majority of patients deny suicidal ideation and intent in their last communication before their death by suicide [3, 4]. This points to the need for additional, objective monitoring strategies to better know when to intervene. Prior research has shown that individuals experiencing suicidal thoughts manifest changes in their speech [5]. This presents an opportunity for effective monitoring, as speech can be easily collected and relates to an individual's underlying condition.

Most previous work into automatically detecting suicidal or depressed speech has focused on laboratory collected datasets [5, 6, 7, 8]. However, these datasets are not necessarily representative of the variations in environment and mood present under real world conditions. Furthermore, characteristics of speech change at the sub-second scale, while thoughts of suicide can be longer in duration [9] and vary considerably [10].

Suicidal ideation is also related to the manner in which emotion is expressed and prior work has examined this link [11, 12]. Self-reports of momentary suicidal ideation have been strongly associated with negative affect among psychiatric inpatients [13]. By first detecting emotional variations from speech, it may be possible to use emotion as a predictor of suicide. This still requires emotion detection of real world speech, which is a difficult task due to the confounding factors of environment, noise, and subject differences. Furthermore, due to the sensitive nature of suicidal data, there are no publicly available datasets linking naturally recorded speech, emotion, and suicide.

In this paper, we present the Ecological Measurement of Affect, Speech, and Suicide (EMASS) Dataset. It contains recordings of natural phone conversations, as well as regular self-reports of emotion, mood, and suicidal thoughts using ecological momentary assessment (EMA) methods [14]. The participants include individuals with recent suicidal ideation or behavior, as well as psychiatric and clinical controls. The collection is ongoing, and the dataset is still relatively small. We demonstrate how outside data can be used to generate emotionally salient features. We then train a model to accurately predict a set of emotion measures, despite restrictive real-world conditions and small amounts of data. These measures were found to be indicative of suicidal ideation in prior work [13]. Finally, we show how emotion predicted from speech can be used to separate healthy controls from those with recent suicidal ideation. This system could eventually allow for the detection of the onset of suicidal ideation, making early intervention possible.

2. The EMASS Dataset

The Ecological Measurement of Affect, Speech, and Suicide (EMASS) Dataset is a collection of natural smartphone speech and momentary self-ratings. The collection is ongoing, and the current snapshot includes 43 individuals, each enrolled for eight weeks. Participants were divided into four groups - healthy controls (HC), psychiatric controls (PC), and individuals that have experienced recent suicidal ideation (SI) or suicide attempts (SA). Individuals in the SI and SA groups were admitted to the hospital for thoughts or behavior related to suicide. Individuals in the PC group were admitted to the hospital for reasons other than suicide (e.g., substance use). All groups, with the exception of HC, were enrolled in the study during their psychiatric admission. Immediately following discharge, they were given a smartphone with the PRIORI app, which securely records their

http://dx.doi.org/10.21437/Interspeech.2019-1830

Table 1: *The amounts of data from different groups of subjects, including healthy controls (HC), psychiatric controls (PC), and individuals hospitalized for suicidal ideation (SI) and attempts (SA). Subjects must contain at least five calls to be included.*

(a) *The full dataset without associating calls with surveys.*

	All	HC	PC	SI	SA
Subjects	43	19	7	12	4
Calls	4078	1780	761	1208	295
Hours	402	239	51	93	15

(b) *Only those calls occurring within one hour before a survey.*

	All	HC	PC	SI	SA
Subjects	16	13	0	3	0
Calls	216	201	0	15	0
Hours	25	23	0	2	0

end of phone conversations [15]. These recordings are then encrypted and uploaded to our server for automatic analysis. Table 1a shows the number of calls collected for each subject group, for a total of 4,078 calls over 402 hours.

In addition to PRIORI, the mEMA app by ilumivu was installed on the smartphone, which presented participants with three surveys throughout the day at random times. They were also asked to initiate surveys if they experienced suicidal ideation or behavior. In these surveys, participants were asked to report on their current affect, using items from the Positive and Negative Affect Schedule (PANAS-X) [16]. Affect was rated on a five point Likert Scale for 11 different categories, which are divided into three groups, based on previous work [13]. The three groups are: **(1) Positive Emotion** - Confident, Excited, Happy; **(2) Negative Emotion** - Sad, Guilty, Worried, Shame, Hopeless; **(3) Anger/Irritability** - Anger at Others, Anger at Self, and Irritable. There are 3,359 surveys included in the dataset snapshot used for this study.

We associate phone call recordings with surveys to enable automatic prediction. Our initial experiments concluded that calls occurring after surveys were less related to the rated emotions than calls occurring before surveys. As such, call recordings are labelled with the emotion present in the closest following survey. Furthermore, we hypothesize the more time that separates a call and a survey, the weaker the certainty in the rated emotion. Because of this, we examine the impact of cutting off the training and testing data at different hours of separation. Survey separation is measured from the start of a call to the survey response. Figure 1 displays the number of calls present at different cutoffs ranging from one hour to two days.

We require at least five calls to be within the cutoff for a subject to be included in experiments (not necessarily from five unique surveys). Table 1b gives the amount of data available for a one hour cutoff. While this severely reduces the data, it increases our certainty in the results. As such, we focus analysis on these 16 subjects and 216 calls from the HC and SI groups.

3. Features

Due to the relatively small dataset, we focus on three knowledge based, feature sets - eGeMAPS, Rhythm Statistics, and Emotion Statistics. eGeMAPS is a state-of-the-art emotion recognition feature set [17] and Rhythm Statistics have been used effectively in prior work in mood recognition [18]. We compare the

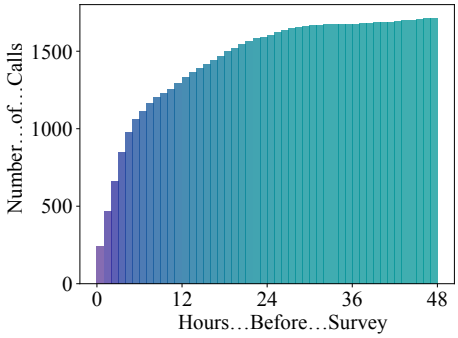


Figure 1: *Cumulative histogram of the hours from calls to the following survey. There are a total of 239 calls with a survey within one hour afterwards (216 from subjects with at least five).*

efficacy of these features with Emotion Statistics - features generated by a deep learning system trained on existing emotion corpora. Our hypothesis is that Emotion Statistics will outperform the other two, due to the small amount of data available for training emotion recognition in the EMASS corpus.

All call recordings are segmented using the ComboSAD algorithm, introduced in [19] and adapted for contiguous segments in [20]. The algorithm estimates the presence of speech using six signals - harmonicity, clarity, prediction gain, periodicity, perceptual spectral flux, and energy. These are then combined using principle component analysis (PCA) and then grouped into segments ranging from 2-30 seconds. All calls must at least contain at least three speech segments to ensure enough data for accurate feature extraction.

3.1. eGeMAPS

The eGeMAPS feature set was introduced in [17] and is extracted using OpenSMILE [21]. Low level descriptors (LLDs) are extracted for frequency, energy, amplitude, and spectral parameters in each segment and result in 23 values per frame.

3.2. Rhythm Statistics

Segments are subdivided into two second subsegments using a sliding window with a one second step size. Seven-dimensional representations of rhythm are then extracted for each subsegment, following the work by Tilsen and Arvaniti [18].

3.3. Emotion Statistics

We extract segment-level emotion using our previously trained Multiclass Adversarial Discriminative Domain Generalization (MADDoG), introduced in [22]. This allows us to use outside data to generate a set of features indicative of emotion fluctuations. The MADDoG model is trained to recognize dimensional emotion (either activation or valence) and consists of three main parts. The **Feature Encoder** uses 40 dimensional log Mel Filter Banks (MFBs), extracted using the Kaldi speech recognition toolkit [23] (frame length of 25 ms, frame shift of 10 ms). It consists of a convolutional neural network (CNN) followed by pooling to produce a segment-level representation. The **Emotion Classifier** is a dense neural network (DNN) trained to recognize three bins of dimensional emotion (low, mid, or high activation/valence) using the segment-level representation. The **Critic** is a DNN that is adversarially trained to ensure that the segment-level representation is similar across different datasets. The critic has a separate output for each training dataset.

Training alternates between two steps: (1) Freeze all

Table 2: *The results for each feature set on calls within one hour before surveys. AUCs are averaged across iterations, subjects, emotions. The best hour cutoff is determined for each feature set. The AUC error is the standard deviation across subjects.*

Features	Best Cutoff Hr.	AUC
eGeMAPS	8	0.53 ± 0.12
Rhythm Statistics	16	0.54 ± 0.09
Emotion Statistics	**24**	**0.63 ± 0.10**

weights besides the Critic and estimate the Wasserstein Distance [24] between datasets; (2) Freeze the Critic and train the remainder of the model to recognize emotion, while also minimizing the Critic's estimate of the Wasserstein Distance. This causes the segment-level representations for each dataset to iteratively get closer to one another, eventually "meeting in the middle". This has the intended effect of learning emotion in a more generalized manner so that the model can be used across yet unseen datasets. For more information, please refer to [22].

This training method allows us to leverage emotional speech from three other datasets - IEMOCAP [25], MSP-Improv [26], and PRIORI Emotion [27]. We then train two separate MADDoG models for activation and valence using the three combined datasets. Segments from the EMASS dataset are then input to the models, resulting in three bins of emotion for both activation and valence, or six values per segment.

3.4. Call-Level Statistics

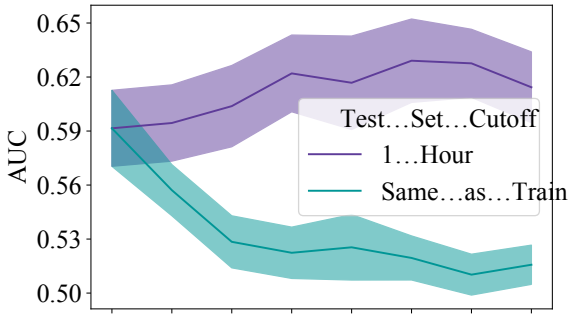
We apply 31 statistics across the concatenated segments to produce call-level features, as in [20]. These include the mean, standard deviation, skewness, kurtosis, minimum, maximum, and range of the signal. We perform linear regression on the signal and use the fit parameters and error as statistics. We then extract the various percentiles and percentile differences and calculate the percentage of the signal above different thresholds.

4. Emotion Modeling

We compare all three feature sets using a DNN, trained to classify one of the 11 emotion measures in the EMASS dataset. The selected target emotion for each experiment is converted from a five point Likert Scale to a fuzzy binary scale for classification. The purpose of this scheme is to eventually allow for our system to distinguish between baseline and atypical emotion. The emotion baseline is estimated with the median subject rating, which produces a baseline of 1, 2, or 3 for each of 11 emotions. Each emotion scale is binarized with a fuzzy value of 0.5 between baseline and atypical values, as follows:

- Baseline of 1 or 3: $1 \rightarrow 0.0 \quad 2 \rightarrow 0.5 \quad (3,4,5) \rightarrow 1.0$
- Baseline of 2: $(1,2) \rightarrow 0.0 \quad 3 \rightarrow 0.5 \quad (4,5) \rightarrow 1.0$

Each experiment begins by randomly dividing the subjects into five sets for cross validation. One of the sets is reserved for testing, ensuring a subject-independent analysis. Each of the remaining subjects has their data randomly divided between training and validation, with 1/5 of their data used for validation. This process is repeated 100 times for each subject, resulting in 100 splits. We calculate the standard deviation of the target emotion within the two folds and use the split that maximizes their product. This ensures enough emotion variability in each fold. We found that applying Z-normalization was beneficial only for the eGeMAPS feature set and normalize it based on train data. All experiments are repeated a total of 100 times with different fold assignments to achieve more stable results.



Cutoff Hr.	1	2	4	8	16	24	36	48
Median Hr.	0.43	0.96	1.78	2.63	3.27	3.71	3.93	4.03
Subjects	16	20	24	25	25	25	27	28
Calls	216	448	838	1151	1405	1574	1661	1703

Figure 2: *Mean AUC over all emotions, subjects, and iterations using emotion statistic features at different cutoffs. The error bands show the standard deviation between iterations. The table displays the amount of data at each cutoff.*

We use a DNN for classification with four hidden layers (widths of 1024, 512, 256, and 256) using a RReLU activation function, found to work best in [28]. The output layer employs a sigmoid activation function and is trained with binary cross entropy loss. This loss is weighted by the inverse of the count of each emotion value (0.0, 0.5, 1.0) in the training set. The Adam optimizer [29] is used with a learning rate of 0.0001 and default parameters. This DNN was found to outperform random forest and support vector machines (SVMs) in early experiments and is the focus of this paper. Training is performed over ten epochs with batches selected to contain all of one subject's data. This ensures the model focuses on learning within-subject variations versus cross-subject biases. We determine the stopping epoch by maximizing Pearson's correlation of the actual emotion and predictions across all data in the validation set.

The test predictions are then estimated using the held-out subjects and selected model. For each test subject, we calculate an Area Under the Receiver Operating Characteristic Curve (AUC) as our performance measure. AUC represents the ability of a system to correctly rank pairs of instances and has a chance rating of 0.5 and ideal rating of 1. Subjects must have at least one negative instance (0.0) and one positive instance (1.0) to be able to calculate a valid AUC. Instances with the fuzzy value of 0.5 are not used to calculate test AUC. Because of this, each emotion experiment will have a different set of test subjects that have enough data for AUC calculation (see Table 3).

5. Results

In this section, we explore speech emotion classification using different feature sets, survey cutoffs, and emotion measures.

We first examine the effect of feature set choice and focus on only testing with calls within one hour of a survey. We employ varying amounts of data in the training set and allow for a cutoff of 1, 2, 4, 8, 16, 24, 36, or 48 hours. Table 2 gives the performance of each feature set averaged across all subjects, emotions, and iterations using its best performing training cutoff. We find that the Emotion Statistics perform substantially better than the others, which are close to chance. This is likely due to the Emotion Statistics already containing estimates of emotion at the segment level. This allows the model to over-

Table 3: *Results on emotion measures using emotion statistic features with a 24 hour cutoff. Only calls with a survey within one hour afterwards are used in testing. The amount of subjects and non-fuzzy calls available to calculate AUCs are shown.*

Emotion	Subjects	Calls	AUC
Confident	7	66	0.64 ± 0.19
Excited	7	92	0.51 ± 0.19
Happy	7	87	0.63 ± 0.21
Sad	6	65	0.54 ± 0.08
Guilty	4	59	0.66 ± 0.25
Worried	4	50	0.62 ± 0.23
Shame	3	45	0.57 ± 0.12
Hopeless	2	21	0.72 ± 0.04
Anger at Others	**8**	**89**	**0.78 ± 0.16**
Anger at Self	5	65	0.60 ± 0.23
Irritable	3	24	0.69 ± 0.34

come the lack of data, which makes classification difficult for other feature sets. Because of this, the following analyses only focus on the Emotion Statistics feature set.

Figure 2 shows our analysis of performance at different training set cutoffs, averaged over subjects, iterations, and emotion measures. While a larger cutoff allows for more training calls, it lowers the certainty in training labels. However, because subjects usually participate in three surveys per day, the median separation between calls and surveys is still only 4.03 hours even with a 48 hour cutoff. There are diminishing returns for the amount of added data with each increase in cutoff (Figure 1). When testing on the newly added data at each increase in cutoff, we find decreased performance. However, if we only test on the 216 calls within one hour of a survey, we see performance increase with a maximum at 24 hours. This provides the most data for classification without overly diluting the labels.

We lastly examine the model's capability to detect different types of emotion using the Emotion Statistics feature set, a 24 hour training cutoff, and the 216 test calls within one hour of a survey. Table 3 presents the AUC for each emotion, averaged over all iterations and subjects with enough data for testing. Due to the lack of data, it is difficult to draw conclusions about individual measures. One exception is "Anger at Others", which has the most subjects (8), highest AUC (0.78), and a relatively small standard deviation between subjects (0.16). In total, 8/11 emotions have an AUC of at least 0.6, demonstrating an overall trend in the model's ability to capture emotion in natural speech.

6. Suicidal Ideation Analysis

In this section, we explore the relationship between suicidal ideation and emotion estimated from speech. We focus on HC and SI subjects, as they are the groups with the most data (19 and 12 subjects, respectively). Emotion measures are extracted from the 2,988 HC and SI calls using the previously trained DNNs. Each estimate is only taken from models where the subject was unused during training. We exclude measures of Excited, Sad, and Shame, as they were previously predicted with less than 0.6 AUC. We calculate the within-subject standard deviation of each emotion to gauge each emotion's variability.

Figure 3 shows the overall variability of the two groups across the different emotions. We consistently find that subjects with SI have lower levels of emotional variability. We then use those emotions with significant differences (Guilty, Hopeless, Anger at Others, Anger at Self, Irritable) to classify HC versus SI. We average each of the five emotion standard deviations for subjects and use this as an estimate. Using this method, we attain a performance of 0.79 AUC.

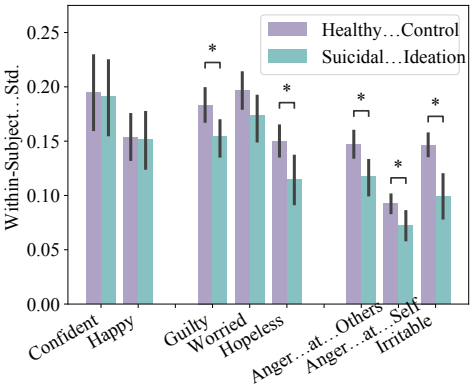


Figure 3: *The within-subject standard deviation of emotions. * Designates a significant difference (t-test, $p<0.05$).*

These findings, though preliminary, are inconsistent with existing research on affective instability and suicide. One study found that heightened affective instability was associated with suicidal behaviors [30], while another found no link [31]. However, individuals in these samples were recruited based on a borderline personality disorder diagnosis and were not compared to healthy controls, unlike the present analysis. It is possible that self-reported affect differs from more objective measures (i.e. speech), or that our experiments consider too few subjects.

7. Conclusions

In this work, we introduced the EMASS dataset, which allows for an investigation into the relationship between speech from natural conversations, emotion fluctuations, and suicidal ideation. We successfully detected emotion using the still relatively small EMASS dataset by first generating features representative of emotion dynamics on outside data. This shows that the MADDoG algorithm is capable of training sufficiently general representations for use in real-world applications. Finally, we examined how emotion fluctuations detected from speech are able to distinguish subjects with recent suicidal ideation from healthy controls, linking speech, affect, and suicidality.

Collection of the EMASS dataset is currently ongoing and extracted features will be made available through the NIH Data Archive. While this paper focused on the momentary ratings of affect, the participant surveys also include questions related to mood and suicidal ideation. Furthermore, the EMASS dataset includes weekly clinical assessments, which could give a more reliable indication of subject progression. Future work will aim to detect the onset of suicidal ideation and explore this interplay between self-assessed and clinician-assessed mood. The emotion dysregulation features introduced in this paper will be key to making this future analysis possible.

8. Acknowledgements

This work was supported by the NSF (CAREER-1651740), NIMH (R34MH100404, R01MH108610, R01MH112674), the Heinz C Prechter Bipolar Research Fund and the Richard Tam Foundation at the University of Michigan. The data collection effort was reviewed and approved by the IRBs of Butler Hospital and the University of Michigan (HUM00052163).

9. References

[1] Centers for Disease Control and Prevention NCfIPaC, "Web-based injury statistics query and reporting system (wisqars)," [online] Available from URL: www.cdc.gov/injury/wisqars.

[2] J. C. Franklin, J. D. Ribeiro, K. R. Fox, K. H. Bentley, E. M. Kleiman, X. Huang, K. M. Musacchio, A. C. Jaroszewski, B. P. Chang, and M. K. Nock, "Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research." *Psychological Bulletin*, vol. 143, no. 2, p. 187, 2017.

[3] E. T. Isometsa, M. E. Heikkinen, M. J. Marttunen, M. M. Henriksson *et al.*, "The last appointment before suicide: is suicide intent communicated?" *The American journal of psychiatry*, vol. 152, no. 6, p. 919, 1995.

[4] K. A. Busch and J. Fawcett, "A fine-grained study of inpatients who commit suicide," *Psychiatric Annals*, vol. 34, no. 5, pp. 357–364, 2004.

[5] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.

[6] S. Scherer, J. Pestian, and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 709–713.

[7] M. E. Larsen, N. Cummins, T. W. Boonstra, B. O'Dea, J. Tighe, J. Nicholas, F. Shand, J. Epps, and H. Christensen, "The use of technology in suicide prevention," in *2015 37th annual international conference of the IEEE engineering in Medicine and biology society (EMBC)*. IEEE, 2015, pp. 7316–7319.

[8] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker, "Detecting depression: a comparison between spontaneous and read speech," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 7547–7551.

[9] T. K. Witte, K. K. Fitzpatrick, T. E. Joiner Jr, and N. B. Schmidt, "Variability in suicidal ideation: a better predictor of suicide attempts than intensity or duration of ideation?" *Journal of affective disorders*, vol. 88, no. 2, pp. 131–136, 2005.

[10] T. K. Witte, K. K. Fitzpatrick, K. L. Warren, C. Schatschneider, and N. B. Schmidt, "Naturalistic evaluation of suicidal ideation: variability and relation to attempt status," *Behaviour Research and Therapy*, vol. 44, no. 7, pp. 1029–1040, 2006.

[11] A. Kranzler, K. B. Fehling, M. D. Anestis, and E. A. Selby, "Emotional dysregulation, internalizing symptoms, and self-injurious and suicidal behavior: Structural equation modeling analysis," *Death studies*, vol. 40, no. 6, pp. 358–366, 2016.

[12] K. C. Law, L. R. Khazem, and M. D. Anestis, "The role of emotion dysregulation in suicide as considered through the ideation to action framework," *Current Opinion in Psychology*, vol. 3, pp. 30–35, 2015.

[13] M. F. Armey, L. Brick, H. T. Schatten, N. R. Nugent, and I. W. Miller, "Ecologically assessed affect and suicidal ideation following psychiatric inpatient hospitalization." *General hospital psychiatry*, 2018.

[14] N. Bolger, A. Davis, and E. Rafaeli, "Diary methods: Capturing life as it is lived," *Annual review of psychology*, vol. 54, no. 1, pp. 579–616, 2003.

[15] Z. N. Karam, E. M. Provost, S. Singh, J. Montgomery, C. Archer, G. Harrington, and M. G. Mcinnis, "Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4858–4862.

[16] D. Watson and L. A. Clark, "The PANAS-X: Manual for the positive and negative affect schedule-expanded form," 1999.

[17] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

[18] S. Tilsen and A. Arvaniti, "Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 628–639, 2013.

[19] S. O. Sadjadi and J. Hansen, "Unsupervised speech activity detection using voicing measures and perceptual spectral flux," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 197–200, 2013.

[20] J. Gideon, E. M. Provost, and M. McInnis, "Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 2359–2363.

[21] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[22] J. Gideon, M. G. McInnis, and E. Mower Provost, "Barking up the Right Tree: Improving Cross-Corpus Speech Emotion Recognition with Adversarial Discriminative Domain Generalization (ADDoG)," *arXiv e-prints*, p. arXiv:1903.12094, Mar 2019.

[23] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, no. EPFL-CONF-192584. IEEE Signal Processing Society, 2011.

[24] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv preprint arXiv:1701.07875*, 2017.

[25] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Language resources and evaluation*, vol. 42, no. 4, p. 335, 2008.

[26] C. Busso, S. Parthasarathy, A. Burmania, M. AbdelWahab, N. Sadoughi, and E. M. Provost, "Msp-improv: An acted corpus of dyadic interactions to study emotion perception," *IEEE Transactions on Affective Computing*, no. 1, pp. 67–80, 2017.

[27] S. Khorram, M. Jaiswal, J. Gideon, M. McInnis, and E.-M. Provost, "The priori emotion dataset: Linking mood to emotion detected in-the-wild," *Interspeech 2018*, pp. 1903–1907, 2018.

[28] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[30] S. Yen, M. T. Shea, C. A. Sanislow, C. M. Grilo, A. E. Skodol, J. G. Gunderson, T. H. McGlashan, M. C. Zanarini, and L. C. Morey, "Borderline personality disorder criteria associated with prospectively observed suicidal behavior," *American Journal of Psychiatry*, vol. 161, no. 7, pp. 1296–1298, 2004.

[31] P. S. Links, R. Eynan, M. J. Heisel, A. Barr, M. Korzekwa, S. McMain, and J. S. Ball, "Affective instability and suicidal ideation and behavior in patients with borderline personality disorder," *Journal of personality disorders*, vol. 21, no. 1, pp. 72–86, 2007.