Investigating Facial Behavior Indicators of Suicidal Ideation

Eugene Laksana[1], Tadas Baltrušaitis[2], Louis-Philippe Morency[2], John P. Pestian[3]
[1] Institute for Creative Technologies, University of Southern California, CA, USA
[2] Language Technology Institute, Carnegie Mellon University, Pittsburgh, PA, USA
[3] Cincinnati Children's Hospital Medical Center, Cincinnati, OH

***Abstract*— Suicide is the deliberate self-inflicted act with the intent to end one's own life. It reflects both profound personal suffering and societal failure. While certain suicide risk factors are well understood, predicting suicide attempts remains a very challenging problem. In this paper, we investigate non-verbal facial behaviors to discriminate among control, mentally ill, and suicidal patients. For this task, we a balanced corpus containing interviews of male and female patients with and without suicide ideation and/or mental health disorders from 3 different hospitals. In our experiments, we explored smiling, frowning, eyebrow raising, and head motion behaviors. We investigated both the occurrence of these behaviors and also how they were conducted. We found that facial behavior descriptors such as the percentage of smiles involving the contraction of the orbicularis oculi muscles (Duchenne smiles) had statistical significance between the suicidal and nonsuicidal groups. The results demonstrated that smiling behavior was the most discriminative feature set between these 3 classes. Our experiments also demonstrated that the stage of the interview in which these facial behaviors occur impacts their discriminative power.**

I. INTRODUCTION

Suicide is the deliberate self-inflicted act with the intent to end one's life. By recent WHO estimations, over 800,000 people die from suicide every year, with at least 20 times more attempted suicides [41]. Despite the high cost to individuals, families, communities, and public health suicide still remains a misunderstood and under-researched cause of death.

Suicide risk factors include family history, demographics, mental illness co-morbidities, and nonverbal behavior and cues [35], [16], [13]. Diagnosis of suicide risk is often subjective in nature, relying almost exclusively, on the opinion of individual clinicians risking a range of subjective biases. Furthermore, as depression often places an individual at higher risk of engaging in suicidal behaviors [17]. This makes it very difficult to distinguish between suicidal depressed individuals and just depressed individuals, the task we are tackling in our work.

Predicting when someone will commit suicide is extremely difficult [20], [27], but trained clinicians can identify the contributing factors to suicide risk using standardized clinical tools [3]. Such tools can, however, be cumbersome and may not reliably translate into routine interactions between clinicians, caregivers, or educators. In this paper we describe a novel method to automatically analyzing subjects' facial behavior to categorize them as either suicidal, mentally ill but not suicidal, or a control.

This work was not supported by any organization

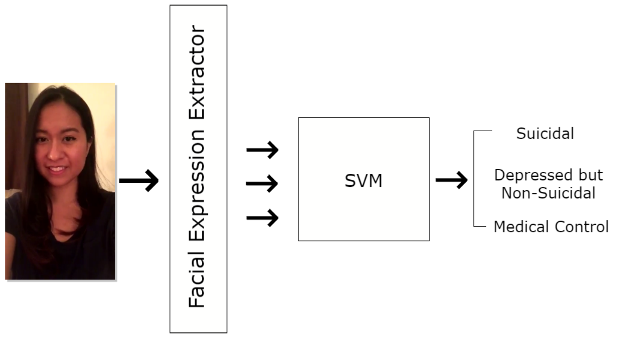


Fig. 1: This is a summary of our final infrastructure. Videos were passed through an extractor which provide statistical summaries of facial behaviors (eg. smiling, frowning, head movement, and eyebrow raising) over designated portions of each interview. These features were then used to train a support vector machine (SVM), which then predicted which of the three groups a patient belonged in.

In this paper we performed an analysis of nonverbal behaviors on a a multi-site and multi-cultural video corpus containing subjects who were either the control, suffered from depression, or were suicidal. [31] We constructed facial behavior features motivated by symptoms of depression/suicide ideation to perform (a) an assessment of these behaviors as indicators of suicidality (b) 3-way classification task using Support Vector Machines (SVMs). Based on our results, we determined that smile-related features produced the highest performance. More importantly, our experiments indicated that the question-level context in which the features are being evaluated on can be just as significant to the model's performance as selecting strong behavioral markers.

The paper is structured as follows: in Section II we discuss the related work on suicidality classification and its behavior indicators; Section III describes the dataset we used; this is followed by the description of behavioral indicators explored in our work in Section IV; we follow this by our experimental procedure in Section V and results in Section VI. We conclude and present future directions in Section VII.

II. BACKGROUND

We first discuss the work done on computational models of suicidality together with work on related topics in healthcare. We then move on to describe the work done in medical and psychology literature on visual behavioral indicators of suicidality.

A. Computational analysis

Efforts to understand suicide risks can be roughly clustered into traits or states. Trait analyses focus on stable characteristics rooted in, and measured using biological processes [6], [21]. State analyses, the topic of this research, measure dynamic characteristics like verbal and non-verbal communication, termed Thought Markers [28].

Work in Natural Language Processing have successfully identified differences in retrospective suicide notes, newsgroups, and social media [24], [16], [19]. Desmet [9] used text-based signals to identify suicide risk that range from 60% to 90%. Li et al. [22] presented a framework using machine learning to identify individuals expressing suicidal thoughts in web forums; Zhang et al. [43] used microblog data to build machine learning models that identified suicidal bloggers with approximately 90% accuracy. Pestian et al. [29] demonstrated that machine learning algorithms could distinguish between notes written by people who died by suicide and simulated suicide notes better than mental health professionals could (71% vs. 79%) [29]. In an international, shared task-setting that includes multiple groups sharing the same task definition, data set, and a scoring metric, 24 teams developed and tested computational algorithms to identify emotions in over 1,319 suicide notes written shortly before death [40]. The results showed that the fusion of multiple methods outperform single methods [30]. Suicidal thought markers have also been studied prospectively. The Suicidal Adolescent Clinical Trial [28] used machine learning to analyze interviews with 60 suicidal and control patients, classified patients into suicidal or control groups with $> 90\%$ accuracy [28].

Acoustic indicators of suicidality have also received a lot of interest from the speech analysis community [7]. Analysis of acoustic features such as pauses and vowel spacing demonstrated their uselessness in detecting suicidality [39], [34]. Yingthawornsuk et al. [42] examined spectral properties of control, depressed, and suicidal voices. They demonstrated the ability of classifying suicidal voices using interview style speech. Scherer et al. [35] used a set of 16 adolescent speakers and performed suicidality classification using Support Vector Machine (SVM) and Hidden Markov Model (HMM) classifiers.

All of the automatic classification of suicidality work has been done on acoustic and linguistic signals, and we are not aware of work using nonverbal visual behaviors. However, visual signals have been used for other health care related applications, specifically: psychosis, depression, Post Traumatic Stress Disorders, and anxiety. Tron et al. [38] used Facial Action Unit based features (activation level, length and change ratio) to classify between patients with schizophrenia and controls. Relationships between automatically detected facial Action Units and depression have been explored by Girard et al. [15]. Alghowinem et al. found eye gaze based features to be discriminative of patients with depression versus controls [1]. Finally, Stratou et al. [37] found gender differences in automatically detected Action Unit 4 (frown) in depressed patients. Our work builds on top of this work by exploring the relationships between suicidality and automatically detected facial Action Units.

B. Behavioral indicators

We are not aware of any computational work using visual indicators of suicidality, however, this is not the case for studies in medicine and psychology.

Rudd et al. [33] present warning signs of suicide identified by the American Association of Suicidology. Out of the warning signs the potentially visually identifiable ones include feelings of hopelessness, rage, anger, anxiety, agitation, and dramatic changes in mood. Mandrusiak et al. [23] survey warning signs of suicidality on various Internet sites to identify additional indicators such as feelings of sadness or indications of depression, and sudden changes in behavior. However, they find a lot of inconsistency in the reported warning signs making it difficult to apply them to our work.

A number of studies have look at the reduced presence of the so called Duchenne smile [12] as a behavioral indicator of depression and psychosis [14], [4], [32]. The Duchenne smile is defined as the combination of AU6 and AU12, rather than just AU12 and is more strongly associated with enjoyment [12]. Such distinction allows for differentiation between *felt* smiles and social ones [32], [4] Gaebel and Wölver [14] found that depressed and schizophrenic patients smiled less than controls, with a particularly large effect on the occurrence of Duchenne smiles. Our work also explores the Duchenne smile as a behavioral indicator of depression and suicidality.

III. Dataset

In this work, we used a dataset consisting of interviews with subjects from the Cincinatti Children's Hospital Medical Center (CCHMC), the University of Cincinatti Medical Center (UC), and the Princeton Community Hospital (PCH). The participants were assigned to one of three groups: control, mentally ill, or suicidal. Control patients are defined as patients in the Emergency Department (ED) who had no history of mental disorders or active suicidal thoughts, plans, or attempts within the previous year. Mentally ill patients are those who have met diagnostic criteria for depression but have had no active thoughts, plans, or attempts of suicide in the ED or outpatient clinics. Suicidal patients are those who have had active suicidal thoughts, made plans to die by suicide, or attempted suicide within the previous year, as either disclosed in person or found in electronic medical records. The dataset is comprised of 123 controls, 126 mentally ill patients, and 130 suicidal patients.

Each subject met an interviewer who used a set verbally conducted a Ubiquitous Questionaire (UQ). This dyadic interaction contains 5 open-ended questions: "Do you have hope?", "Do you have any fear?", "Do you have any secrets?" , "Are you angry?", and "Does it hurt emotionally?" These questions were designed to stimulate further conversation related to the patients' conditions and past

TABLE I: Dataset Demographics.

Facility-level Demographics

Hospital Name	Control	Mental Health	Suicidal	Male	Female	Age Range	Average Age
CCHMC	41	42	43	39	87	13 - 18	15.6
UC	42	42	44	61	67	19 - 70	42.6
PCH	40	42	43	48	77	18 - 66	42.1
All	123	126	130	148	231	13- 70	33.5

Gender-level Demographics

Gender	Control	Mental Health	Suicidal	Age Range	Average Age
Male	49	50	49	13 - 62	34.71
Female	74	76	81	13 - 70	32.7

experiences. Subject responses are video and audio recorded, and transcriptions with pointers to time intervals containing responses to each of the 5 questions are provided. Each video is approximately 8 minutes long. Additional demographics are provided in Table I.

IV. Visual Behaviors and Suicidality

Literature indicates a number of facial behavior patterns that are believed to be associated with suicidal ideation. Among identified behavioral cues are anxiety, deception, outbursts of anger, and crying [3], [18]. Many of these behaviors such as deception and anxiety are very difficult to detect even with current state-of-the-art Computer Vision systems, hence we focus on the easier to detect facial behaviors. This following section will describe the four facial behaviors – smiling, frowning, eyebrow raising, and head movement that we investigated as they related to depression and suicide ideation in literature and how we operationalized these markers by computationally defining them.

Smiling, frowning, and eybrow raises can be described using Action Units (AUs), from the Facial Action Coding (FACS) system [10] for movements of facial muscle groups. Head motion velocity can be computed when provided the subject's head position relative to the camera at any given time. We used OpenFace [2], an open source state-of-the-art toolbox to extract per-frame AU intensities and head pose in each video frame. Our decision to use this toolbox is largely based on the similarity between our dataset and the Denver Intensity of Spontaneous Facial Action (DISFA) corpus, which OpenFace has been tested on for AU detection. In our experiments, we took statistical summaries (averages and standard deviations) of each of the described features at either the interview or specified question-level context.

A. Smiling Dynamics

Scherer, et al. has indicated that depressed and nondepressed patients tend to smile at similar frequency; however, their dynamics differed. Hence, type of smile that a patient produces during an interview contains just as much, if not more, information regarding their mental and emotional state than just the presence of a smile itself. [35]

For instance, the contraction of the orbicularis oculi muscle during a smile event creates what is known as the Duchenne smile as seen in Figure 2 [11]. The Duchenne smile, along with the smile's onset/offset sharpness and duration, have been shown to be useful for discriminating between genuine and posed smiles [5]. These smile features are relevant because a ”false” smile oftentimes serves to mask negative affect [11].

This type of inclination is common in patients with depression and suicide ideation [3], [8]. Due to the solemn nature of the questions asked through the UQ, we believe that the presence of ”false” smiles during the interview could contain significant information strongly related to internalized negative affect.

We defined a smiling event as any continuous interval of at least 0.2 seconds consisting of nonzero AU12 (Lip Corner Pull) intensity in which AU12's intensity exceeded 1.0 (intensity level A in FACS) at least once. This is to ensure that noise from OpenFace, which can result due to patients pronouncing vowels that produce AU12, were not captured as a legitimate smile. We chose this threshold because no AU12 event in the DISFA dataset was shorter than this. With this definition of the smile event, we constructed the following descriptors:

1) Intensity, Length, and Count: Action Unit intensity is provided through OpenFace on a 5-point scale. The length of the event is described in seconds. Count is simply the total number of smiles present over the section of the interview that the facial behaviors are being extracted from.

2) Duchenne Smile Percentage: Any smiling event in which the mean of AU6 (Cheek Raiser) intensity during the duration of the smile was at least 1.0 was considered a Duchenne smile. The ratio of Duchenne smiles to total number of smiles is the Duchenne smile percentage. This allows us to measure the ratio of ”fake” to ”real” smiles.

3) Sharpness of Smile Onset/Offset: We first applied a moving average filter over the AU12 intensity signal. We defined the smile onset as the longest interval within a smile event in which AU12's intensity consistently increased and exceeded a score of 1.0. We defined the smile offset as the longest interval within a smile event in which AU12's intensity started with a score of at least 1.0 and consistently decreased. The sharpness of the onset was defined as the absolute value of the slope of the line connecting the beginning of the onset to the end of the onset as described by

Fig. 2: Duchenne (top) vs non-Duchenne (bottom) smiles with OpenFace outputs. Any score greater or equal to 1 is considered an AU activation. The Duchenne smile, defined by the co-occurrence of AUs 6 and 12, involves the contraction of the orbicularis oculi and is commonly associated with a spontaneous smile.

Schmidt, et al. [36] The sharpness of the offset was defined as the absolute value of the slope of the line connecting the beginning of the offset to the end of the offset.

B. *Frowning Behavior*

Investigations done by Heller, et al. [18] demonstrated that suicidal subjects who had reattempts produced significantly higher frowning events during their interviews than the single attempt group.

Since questions such as ”Do you have hope?” may evoke more negative affect for patients suffering from mental illness or active suicide ideation than for healthy subjects, we hypothesize that patients belonging to the control group will produce fewer frowns. Frown intensity, length, onset, and offset could contain important information related to a patient’s affective state. For example, a high intensity frown with slow onset and offset could be a subject crying, whereas lower intensity with fast onset and offset could simply be a quick expression of disgust or shock.

Frowning can also indicate a state of confusion or preoccupation [18]. This is particularly helpful for us because non-suicidal patients who are immediately asked intimate questions such as ”do you have hope?” without rapport-building, as done in these interview settings, are likely to express confusion or concern.

Frowning events and features were defined in the same way as we did for smile events and their correspond features except with AU17 (Chin Raise) instead of AU12. For the frown descriptors, we defined frown intensity, frown count, frown offset sharpness, frown onset sharpness, and frown length.

C. *Eyebrow Raises*

Raised eyebrows are commonly related to expressions of surprise. This affect may be particularly important due to the wide range of subjects in the dataset who will be answering the same, deeply intimate questionaire. Subjects who have been admitted into the emergency department due to or with prior records of depression or suicide ideation are more likely to being accustomed to questions of similar nature to that of the UQ from therapy sessions prior to being admitted. However, subjects belonging to the control group, who are not as likely to be as familiar with such intimate questions in a clinical setting, may respond with initial surprise. Thus, capturing the said expression under proper context could lead to a feature that is discriminative of the control group.

Eyebrow raising events were defined in the same way as smile and frown events were except with the mean intensity between AU1 (Inner Brow Raiser) and AU2 (Outer Brow Raiser). Descriptors that we defined for the eyebrow raise included eyebrow raise count, eyebrow raise intensity, and eyebrow raise length.

D. *Head Motion Velocity*

Over 70% of the subjects studied by Nepon et al. [26] who had reported a suicide attempt in their lifetime also claimed to suffer from an anxiety-related disorders. Behaviors related to anxious expressions and their relationship with suicide ideation are therefore worthwhile to investigate. This namely takes the form of fidgeting, looking around the room, and other indications of preoccupation. Since this current work is strictly constrained to facial expressions and head gestures, we decided to investigate head motion velocity. Therefore, if a subject participates in anxiety-driven tasks with their head, such as quickly looking around the room, high head velocity would be captured for that event. On the contrary, a subject who remains stable throughout the duration of the interview will have a relatively low head motion velocity.

We defined head motion velocity as the absolute value of the summation of the numerical derivatives for each of the components of a 3-dimension head position vector.

V. Experimental Methodology

We performed 3 sets of classification experiments in hopes of answering the following questions:

1) Which behavioral patterns are most discriminative of the 3 classes?
2) Which classifier is best suited to perform the classification task?
3) What is the influence of the question-context from which these behavioral descriptors are being extracted from?

We performed person independent 10-fold stratified cross testing. To validate the hyper-parameters we used 10-fold stratified cross-validation on the training data. Since this dataset has a balanced class distribution, we evaluated our models' performances using the average accuracy for each of the 3 classes.

For input feature sets, we used smiling, frowning, eyebrow movement, and head motion behaviors as described in Section IV.

A. Selecting a Discriminative Model

To select the best model for our experiments, we compared classification results from an SVM with radial basis function kernels, Random Forest, and Multinomial Naïve Bayes. To determine the best model, we first selected the top performing feature subset by comparing the average classification scores over each model using the 4 facial behaviors listed in Section IV. Features were extracted from each question level and over the whole interview, and classification scores were averaged. We chose the model which had the highest performance accuracy on our highest performing feature set.

Our tests showed that the SVM had the highest performance (Table III,) so we performed tests using this model. We built an individual SVMs for each feature subset extracted from both the interview and question levels and performed cross validation and testing over the following parameters: {C: 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2, 1e3; Gamma: 1e-4, 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2 }.

Model	Average Score
SVM	**42.4**
Naive Bayes	39.0
Random Forest	39.4

TABLE II: Each discriminative models' (SVM, Random forest, Naïve Bayes) average score over our most discriminative feature set (smiling).

B. Filtering the Interviews

Some subjects were not asked all 5 ubiquitous questions during their interviews. We removed these videos along with those where OpenFace was unable to extract at least 50% of the frames. The latter condition could occur if the patient is wearing glasses (thus making it difficult to extract AU4), or if their head posture is away from the camera for the majority of the session. This filtering led to 333 subjects for further analysis.

C. Question Context-level Evaluation

The remaining subjects after we performed our filtering each responded to 5 ubiquitous questions (UQ). We compared the results of experiments for when the features were extracted over the course of the entire interview and at question-level granularity to evaluate the importance of question-level context.

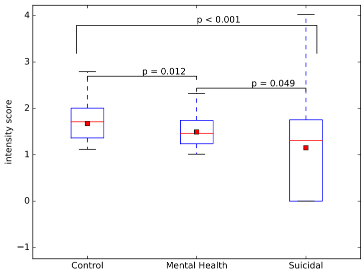
VI. Results & Discussions

A. Statistical Investigations

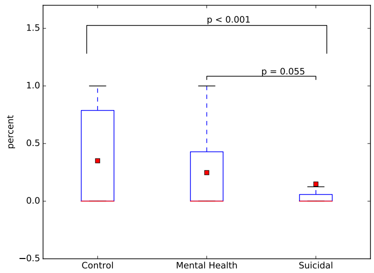
We were interested in the statistical significance between patient conditions and the behavior indicators.

To assess this, we performed nonparametric analysis of variance testing using the Kruskal-Wallis test. We found every behavioral descriptor except for head movement velocity was statistically different ($p < 0.05\%$) for the three groups for at least one context. 13 of these features (primarily smiling and frowning related descriptors) were overall statistically significantly different. 11 of them (also primarily smiling and frowning features) were statistically different for the three groups without having to take into consideration the question-level context.

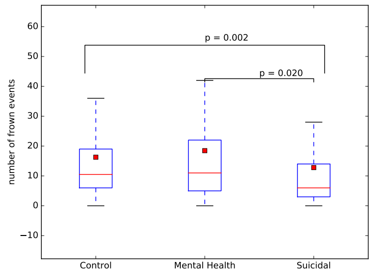
The box and whisker plots in Figure 3 show the distribution of select features' statistical summary from specific context. Each of the visualized features are statistically different for at least 5 of the 6 contexts we tested on (the 5 questions and over the full interview.) Since these features are not Gaussian distributed, we tested for statistical significance using the Mann-Whitney U test – a nonparametric test of the null hypothesis [25].



(a) Average Smile Intensity (Anger)



(b) Duchenne Smile Percentage (Anger)



(c) Number of Frowns (Hope)

Fig. 3: Box and whisker plots capturing the distribution of statistical summaries for selected behavioral descriptors. These diagrams show statistical significance when p <0.05% according to the nonparametric Mann-Whitney U test.

ID	Question
Q1	Do you have hope?
Q2	Are you afraid?
Q3	Do you have any secrets?
Q4	Are you angry?
Q5	Does it hurt emotionally?

(a) Question to ID Mapping

	Q1	Q2	Q3	Q4	Q5	Interview	Avg
Majority Vote							34.5
Frown	**39.3**	37.6	38.9	36.2	35.2	34.4	36.93
Head velocity	38.2	39.7	30.0	30.5	34.1	**42.7**	35.85
Eyebrows	**41.3**	33.7	34	36.7	35.2	34.2	35.87
Smile	46.2	**47.3**	36.3	40.2	39.5	44.6	**42.35**

(b) Accuracy Scores

TABLE III: Averaged accuracy scores of SVMs over 3 classes. Our experiments showed that the smile descriptors we constructed are the most useful features for making prediction boundaries between the tested groups. Furthermore, these results demonstrated the context-sensitive nature of the behavioral markers.

Pestian et al. [31] found that subjects in the suicidal and mental health groups laughed significantly less than those in the control group when asked if they were angry. Box plots in Figures 3a and 3b reflect this; they indicate that subjects in the mental health and suicidal groups smile with less intensity and had lower percentages of detected Duchenne smiles than those in the control group.

Frowning behavior also seemed to occur less frequently among the suicidal group than it did in the mental health and control group. One way of interpreting this is that frowning in the context of this dataset expresses confusion or pre-occupation more than it does sadness. Polarized expressions such as crying tend to be followed up with the subject hiding their face or turning away from the camera. As consequence, OpenFace cannot accurately detect the presence of AU17 within these context. We believe that nonsuicidal patients being asked questions related to hope or anger are more likely to express a confused frown than a suicidal patient would.

While applying pairwise testing, we found that while many of our features are statistically different between the control vs suicidal and mental health vs suicide groups, few were statistically different between the control and mental health groups. This could indicate that confusions within the 3-way classification could be primarily due to an inability to discriminate well between the control and mental health groups.

B. Discriminative Models

Since the SVM reported the highest accuracy score on the top feature subset, smiling (Table II), we used this model to report our results.

We reported accuracy scores in Table III and compared our results with a naïve majority vote baseline. Our results indicated that each facial behavior feature had above-chance performance. Smiling features were our most discriminative behavioral markers. Furthermore, selecting the proper context in the interview from which we perform our facial behavior analysis is important. To highlight the last point, neglecting question-context altogether by evaluating facial expressions over the entire interview resulted in a performance loss of 3% and 4.2% with smiling and frowning features respectively.

We have also built gender independent classifiers. However, the results from these experiments were very similar to the ones reported in this paper.

VII. CONCLUSION & FUTURE DIRECTION

From our experiments, we are able to draw the following four conclusions:

1) The facial behavior features that we constructed are discriminative of suicide ideation and depression within the context of this verbal UQ.
2) Smiling-related behavioral descriptors have the highest performance relative to that of frowning, eyebrow raising, and head velocity.
3) The context from which facial behavior descriptors are being extracted are of great significance for building discriminative features for suicidality and depression.
4) Our behavioral descriptors can discriminate well between suicidal and nonsuicidal patients but not necessarily among all 3 classes.

Follow-up work should focus on extending our feature descriptors to first being able to perform tasks such as posed smile, deception, or anxiety detection. Since our features were motivated by these communication markers, it could be helpful to use these behaviors as features instead of using descriptors which may indicate their presence. As an example, building a posed smile detector using our current smile features would enable us to use the percentage of false smiles or even the presence of deception in the interviews in place of the Duchenne smile percentage feature.

We found that many of our behavioral markers are statistically different between the control/suicidal and control/mental health groups. However, only a few are statistically different between the control and mental health groups. We believe that the lack of discriminative features between the control and mental health groups led to the 3-way classifier's performance loss. Although this is an issue for 3-way classification, the confusion between the control and mental health groups should not influence binary classification between the suicidal and non-suicidal group. Future work can focus on building a discriminative model for the latter task.

REFERENCES

[1] Sharifa Alghowinem, Roland Göcke, Jeffrey F Cohn, Michael Wagner, Gordon Parker, and Michael Breakspear. Cross-Cultural Detection of Depression from Nonverbal Behaviour. *FG*, 2015.

[2] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. OpenFace: an open source facial behavior analysis toolkit. In *IEEE WACV*, 2016.

[3] Aaron T Beck, Maria Kovacs, and Arlene Weissman. Assessment of suicidal intention: the scale for suicide ideation. *Journal of consulting and clinical psychology*, 47(2):343, 1979.

[4] Yulia E. Chentsova-Dutton, Jeanne L. Tsai, and Ian H. Gotlib. Further Evidence for the Cultural Norm Hypothesis: Positive Emotion in Depressed and Control European American and Asian American Women. *Cultur Divers Minor Psychol*, 16(2):284–295, 2010.

[5] Jeffrey F Cohn and Karen L Schmidt. The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(02):121–132, 2004.

[6] Alessandra Costanza, Isabella D'Orta, Nader Perroud, Sandra Burkhardt, Alain Malafosse, Patrice Mangin, and Romano La Harpe. Neurobiology of suicide: do biomarkers exist? *International journal of legal medicine*, 128(1):73–82, 2014.

[7] Nicholas Cummins, Stefan Scherer, Jarek Krajewski, Sebastian Schnieder, Julien Epps, and Thomas F. Quatieri. A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, 71:10 – 49, 2015.

[8] Udo Dannlowski, Anette Kersting, U-S Donges, Judith Lalee-Mentzel, Volker Arolt, and Th Suslow. Masked facial affect priming is associated with therapy response in clinical depression. *European archives of psychiatry and clinical neuroscience*, 256(4):215–221, 2006.

[9] Bart Desmet. *Finding The Online Cry for Help*. PhD thesis, Universiteit Gent, Gent, Belgium, 5 2014.

[10] P. Ekman and W. V. Friesen. *Manual for the Facial Action Coding System*. Palo Alto: Consulting Psychologists Press, 1977.

[11] Paul Ekman. Facial expressions of emotion: New findings, new questions. *Psychological science*, 3(1):34–38, 1992.

[12] Paul Ekman, Richard J. Davidson, and Wallace V. Friesen. The Duchenne Smile: Emotional Expression and Brain Physiology II. *Journal of Personality and Social Psychology*, 58(2):342–353, 1990.

[13] William J. Fremouw, Maria de Perczel, and Thomas E. Ellis. *Suicide risk: Assessment and response guidelines*. 1990.

[14] Wolfgang Gaebel and Wolfgang Wölwer. Facial expressivity in the course of schizophrenia and depression. *European Archives of Psychiatry and Clinical Neuroscience*, 254(5):335–342, 2004.

[15] Jeffrey M Girard, Jeffrey F Cohn, Mohammad H Mahoor, S Mohammad Mavadati, Zakia Hammal, and Dean P. Rosenwald. Nonverbal Social Withdrawal in Depression : Evidence from manual and automatic analyses. *Image and Vision Computing*, 32(10):641–647, 2013.

[16] José M Gómez. Language technologies for suicide prevention in social media. *Processing in the 5th Information Systems Research Working Days (JISIC 2014)*, page 21.

[17] Keith Hawton, Carolina Casañas I Comabella, Camilla Haw, and Kate Saunders. Risk factors for suicide in individuals with depression: A systematic review. *Journal of Affective Disorders*, 147(1-3):17–28, 2013.

[18] Michael Heller, Véronique Haynal-Reymond, André Haynal, and Marc Archinard. Can faces reveal suicide attempt risks.

[19] Yen Pei Huang, Tiong Goh, and Chern Li Liew. Hunting suicide notes in Web 2.0 - Preliminary findings. *Proceedings ISM Workshops 2007 9th IEEE International Symposium on Multimedia - Workshops*, pages 517–521, 2007.

[20] M. Large and C. Ryan. Suicide risk assessment: myth and reality. *International Journal of Clinical Practice*, 68(6):679–681, 2014.

[21] H Le-Niculescu, DF Levey, M Ayalew, L Palmer, LM Gavrin, N Jain, E Winiger, S Bhosrekar, G Shankar, M Radel, et al. Discovery and validation of blood biomarkers for suicidality. *Molecular psychiatry*, 18(12):1249–1264, 2013.

[22] Tim M H Li, Ben C M Ng, Michael Chau, Paul W C Wong, and Paul S F Yip. Collective Intelligence for Suicide Surveillance in Web Forums. In *Intelligence and Security Informatics*, pages 29–37. 2013.

[23] Michael Mandrusiak, M David Rudd, Thomas E Joiner, Alan L Berman, Kimberly A Van Orden, and Tracy Witte. Warning Signs for Suicide on the Internet : A Descriptive Study. *Suicide and Life-Threatening Behavior*, 36(3):263–271, 2006.

[24] Pawel Matykiewicz, Wlodzislaw Duch, and J Pestian. Clustering semantic spaces of suicide notes and newsgroups articles. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 179–184. Association for Computational Linguistics, 2009.

[25] Patrick E McKnight and Julius Najab. Mann-whitney u test. *Corsini Encyclopedia of Psychology*, 2010.

[26] Josh Nepon, Shay-Lee Belik, James Bolton, and Jitender Sareen. The relationship between anxiety disorders and suicide attempts: findings from the national epidemiologic survey on alcohol and related conditions. *Depression and anxiety*, 27(9):791–798, 2010.

[27] Joel Paris. Predicting and preventing suicide: Do we know enough to do either? *Harvard Review of Psychiatry*, 14(5):233–240, 2006.

[28] John P Pestian, Jacqueline Grupp-Phelan, Kevin Bretonnel Cohen, Gabriel Meyers, Linda A Richey, Pawel Matykiewicz, and Michael T Sorter. A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. *Suicide and life-threatening behavior*, 2015.

[29] John P Pestian, Pawel Matykiewicz, and Jacqueline Grupp-Phelan. Using natural language processing to classify suicide notes. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 96–97. Association for Computational Linguistics, 2008.

[30] John P Pestian, Pawel Matykiewicz, and Michelle Linn-Gust. What's in a note: construction of a suicide note corpus. *Biomedical informatics insights*, 5:1, 2012.

[31] John P Pestian, Michael Sorter, Brian Connolly, K Bretonnel Cohen, Jeffry T Gee, Louis-philippe Morency, and Stefan Scherer. A Machine Learning Approach to Identifying the Thought Markers of Suicidal Subjects : A Prospective Multicenter Trial.

[32] Babette Renneberg, Katrin Heyn, Rita Gebhard, and Silke Bachmann. Facial expression of emotions in borderline personality disorder and depression. *Journal of Behavior Therapy and Experimental Psychiatry*, 36(3):183–196, 2005.

[33] M David Rudd, Alan L Berman, Thomas E Joiner, Matthew K Nock, Morton M Silverman, Michael Mandrusiak, Kimberly Van Orden, and Tracy Witte. Warning Signs for Suicide: Theory, Research, and Clinical Applications. *Suicide and Life-Threatening Behavior*, 36(3):255–262, 2006.

[34] Stefan Scherer, Louis-Philippe Morency, Jonathan Gratch, JP Pestian, and CA Playa Vista. Reduced vowel space is a robust indicator of psychological distress: A crosscorpus analysis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.

[35] Stefan Scherer, John Pestian, and Louis-philippe Morency. Investigating the speech characteristics of suicidal adolescents. In *ICASSP*, pages 709–713, 2013.

[36] Karen L Schmidt and Jeffrey F Cohn. Dynamics of facial expression: Normative characteristics and individual differences. In *ICME*. Citeseer, 2001.

[37] Giota Stratou, Stefan Scherer, Jonathan Gratch, and Louis-Philippe Morency. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender. *Journal on Multimodal User Interfaces*, 9:17–29, 2015.

[38] Talia Tron, Abraham Peled, and Alexander Grinsphoon. Automated Facial Expressions Analysis in Schizophrenia : a Continuous Dynamic Approach. In *Pervasive Computing Paradigms for Mental Health (MindCare)*, pages 1–10, 2015.

[39] Verena Venek, Stefan Scherer, Louis-Philippe Morency, Albert Rizzo, and John Pestian. Adolescent suicidal risk assessment in clinician-patient interaction: A study of verbal and acoustic behaviors. In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pages 277–282. IEEE, 2014.

[40] Ellen M Voorhees, Donna K Harman, et al. *TREC: Experiment and evaluation in information retrieval*, volume 1. MIT press Cambridge, 2005.

[41] World Health Organization. Preventing suicide: A global imperative. Technical report, 978 92 4 156477 9, 2014.

[42] T. Yingthawornsuk, H. Kaymaz Keskinpala, D. M. Wilkes, R. G. Shiavi, and R. M. Salomon. Direct acoustic feature using iterative EM algorithm and spectral energy for classifying suicidal speech. In *INTERSPEECH*, volume 2, pages 749–752, 2007.

[43] Lei Zhang, Xiaolei Huang, Tianli Liu, Ang Li, Zhenxiang Chen, and Tingshao Zhu. Using linguistic features to estimate suicide probability of chinese microblog users. In *Human Centered Computing*, pages 549–559. Springer, 2015.