Investigating Facial Behavior Indicators of Suicidal Ideation

Eugene Laksana\textsuperscript{1}, Tadas Baltrušaitis\textsuperscript{2}, Louis-Philippe Morency\textsuperscript{2}, John P. Pestian\textsuperscript{3}

\textsuperscript{1} Institute for Creative Technologies, University of Southern California, CA, USA
\textsuperscript{2} Language Technology Institute, Carnegie Mellon University, Pittsburgh, PA, USA
\textsuperscript{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 used 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 statistically significant differences between the suicidal and nonsuicidal groups. 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 estimates, over 800,000 people die from suicide every year, with at least 20 times more attempted suicides [42]. 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. This risks a range of subjective biases. Furthermore, depression often places an individual at higher risk of engaging in suicidal behaviors [17], making 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 differentiate between suicidal, mentally ill but not suicidal, and control groups.

In this paper we performed an analysis of nonverbal behaviors on a multi-site and multi-cultural video corpus containing subjects who were either the control, suffered from depression, or were suicidal. [31] We analyzed facial behavior features motivated by symptoms of depression/suicide ideation to perform two tasks. The first is an assessment of these behaviors as indicators of suicidality. This involved null hypothesis testing and comparing the distributions of statistical summaries of the facial expressions. The second is a 3-way classification task using predictive models.

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. Performing analysis on Traits involves focuses 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 has successfully identified differences in retrospective suicide notes, news- groups, and social media [24], [16], [19]. Desmet [9] used text-based signals to identify the risk of suicide risk with 60% to 90% accuracy. Li et al. [22] presented a framework using machine learning to identify individuals expressing suicidal thoughts in web forums; Zhang et al. [44] 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...
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 looked at the reduced presence of the Duchenne smile [12] as a behavioral indicator of depression and psychosis [14], [4], [32]. The Duchenne smile is defined as a smile while the orbicularis oculi muscles contract. This feature is more strongly associated with genuine enjoyment than a "normal" smile. [12]. Such distinction allows for differentiation between felt smiles and social ones [32], [4] Gaebel and Wölfer [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.

We are not aware of any computational work using visual indicators of suicidality, however, this is not the case for studies in other mental illnesses such as depression.

III. DATASET

In this work, we used a dataset consisting of interviews with subjects from the Cincinnati Children’s Hospital Medical Center (CCHMC), the University of Cincinnati 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 conducted a verbal ubiquitous questionnaire. 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 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.

Some subjects were not asked all 5 ubiquitous questions during their interviews. We removed these videos along with those where OpenFace [2], the feature extractor that we will use for our experiments, was unable to extract at least 50% of the frames. The latter condition could occur if the patient is wearing glasses, 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.
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. However, we can break characteristics of these behaviors down into various facial expressions. 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 eyebrow 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], a 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 non-depressed patients tend to smile at similar frequency; however, their dynamics differed. [35] 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.

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 1 [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 non-Duchenne smile oftentimes serves to mask negative emotions of the patient [11].

This smiling behavior is common in patients with depression and suicide ideation [3], [8]. Due to the solemn nature of the questions asked through the ubiquitous questionnaire, we believe that the presence of non-Duchenne smiles during the interview could contain significant information strongly related to internalized negative affect.

No AU12 event in the DISFA dataset was shorter than 0.2 seconds. For this reason, 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. With this definition of the smile event, we constructed the following computational 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 frames 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 at intensity level A 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 non-Duchene to
Duchenne smiles.

3) **Slope 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 intensity level A. We defined the smile offset as the longest interval within a smile event in which AU12’s intensity started with intensity level A 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 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 pre-occupation [18]. This is particularly helpful for us because non-suicidal patients who are asked intimate questions such as "do you have hope?"

### V. Experimental Methodology

Frowning events and features were created using the same methodology introduced for smile events and their corresponding 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.

#### A. Eyebrow Raises

Depressed patients are less likely to use visual emphasis with their nonverbal communication due to psychomotor retardation. Hence, eyebrow raising behavior, which is used to highlight attentiveness, convey surprise, or express interest may not be as prominent within the depressed and suicidal groups as it would be in the control.

Eyebrow raising features were created using the same methodology introduced for smile and frown features with the exception of 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.

#### B. 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 L1-norm of the numerical derivatives of the 3-dimensional head position vector.

Our experiments were designed around these 3 research questions:

1) Which facial behavior is most predictive of the patient’s condition?
2) Which classifier is best suited to predict the patient’s condition?
3) What is the influence of the question-context from which these behavioral descriptors are being extracted from?

The question-context analysis refers to evaluating features from only the frames where patients are responding to individual questions during the interview. (Table IIA). For example, performing question-context analysis on Q1 would entail only evaluating features extracted over the frames where the patient is answering the “Do you have hope?” question.

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 averaged accuracy for each of the 3 classes (control, mentally ill but not suicidal, and suicidal).

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

C. Question Context-level Evaluation

The subjects each responded to 5 ubiquitous questions. 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.

D. Selecting a Predictive Model

We compared three predictive models: a Support Vector Machine (SVM) with radial basis function kernels, Random Forest, and Multinomial Naive 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.

VI. RESULTS & DISCUSSION

A. Statistical Analysis

To better understand the correlation between patient’s conditions and their facial behaviors, we performed a set of statistical analyses.
Our experiments showed that the smile descriptors are the most useful features for making prediction between the tested groups. Furthermore, varying accuracy scores across different question for each feature demonstrated the context-sensitive nature of the behavioral markers.

<table>
<thead>
<tr>
<th>Question</th>
<th>ID</th>
<th>SVM</th>
<th>Naive Bayes</th>
<th>Random Forest</th>
<th>Majority Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 Have you hope?</td>
<td>Q2 Are you afraid?</td>
<td>Q3 Do you have any secrets?</td>
<td>Q4 Are you angry?</td>
<td>Q5 Does it hurt emotionally?</td>
<td></td>
</tr>
<tr>
<td>Majority Vote</td>
<td>Frown</td>
<td>Head velocity</td>
<td>Eyebrows</td>
<td>Smile</td>
<td></td>
</tr>
<tr>
<td>39.3</td>
<td>37.6</td>
<td>38.9</td>
<td>36.2</td>
<td>35.2</td>
<td>34.4</td>
</tr>
<tr>
<td>38.2</td>
<td>39.7</td>
<td>30.0</td>
<td>30.5</td>
<td>34.1</td>
<td>42.7</td>
</tr>
<tr>
<td>41.3</td>
<td>33.7</td>
<td>34</td>
<td>36.7</td>
<td>35.2</td>
<td>34.2</td>
</tr>
<tr>
<td>46.2</td>
<td>47.3</td>
<td>36.3</td>
<td>40.2</td>
<td>39.5</td>
<td>44.6</td>
</tr>
</tbody>
</table>

TABLE II: Averaged accuracy scores of SVM prediction models for different question-contexts and for the whole interviews.

We found 7 features to be statistically significant ($p < 0.0015\%$) when features were extracted over the entire interview. Bonferroni correction was used to establish a conservative $p$ value threshold, by correcting for 33 comparisons made (the total number of features used). We used this figure to determine statistical significance of our Kruskal-Wallis tests. These included number of smiles, sum of smiling intensity, average smiling intensity, standard deviation of smiling intensity, percentage of Duchenne smiles, number of frowns, and the sum of frowning intensity. The box plots in Figure 2 show the distribution of exemplar features’ statistical summary from specific contexts.

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 2a and 2b 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.

TABLE III: Each prediction models’ (SVM, Random forest, Naive Bayes) average accuracy over our most discriminative feature set (smiling).

<table>
<thead>
<tr>
<th>Model</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>42.4%</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>39.0%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>39.4%</td>
</tr>
<tr>
<td>Majority Vote</td>
<td>34.5%</td>
</tr>
</tbody>
</table>

B. Classification Task

Table III shows our results for each predictive model when trained using our most discriminative feature set (smiling). Our tests showed that the SVM had the highest performance (Table II), so we performed final classification using this model.

We built individual SVMs for each feature subset extracted from both the interview and question levels and performed cross validation and testing over the following parameters: $\{C: 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2, 1e3; \Gamma: 1e-4, 1e-3, 1e-2, 1e-1, 1, 1e1, 1e2 \}$. We recorded accuracy scores in Table II and compared our results with a naive majority vote baseline. Our results indicated that each facial behavior feature had above-chance performance. Smiling features, however, were our most discriminative behavioral markers.

We also found that selecting the proper context in the interview from which we perform our facial behavior analysis is important. To highlight this 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.

While performing the pairwise testings, 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 may be largely due to an inability to discriminate well between the control and mental health groups.

VII. CONCLUSION & FUTURE DIRECTION

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

1) The facial behavior features that we constructed seem to be discriminative of suicide ideation and depression within the context of this verbal ubiquitous questionaire.
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 helpful for building visual features that will be helpful for predicting 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 these are the behaviors which motivated our investigations into these behavior patterns. We should then use these predictions to help facilitate suicidality classification.

REFERENCES


