

# Identification of Imminent Suicide Risk Among Young Adults using Text Messages

**Alicia L. Nobles**  
Dept. of Systems and  
Information Engineering  
University of Virginia  
aln2dh@virginia.edu

**Jeffrey J. Glenn**  
Dept. of Psychology  
University of Virginia  
jjg9ac@virginia.edu

**Kamran Kowsari**  
Dept. of Systems and  
Information Engineering  
University of Virginia  
kk7nc@virginia.edu

**Bethany A. Teachman**  
Dept. of Psychology  
University of Virginia  
bat5x@virginia.edu

**Laura E. Barnes**  
Dept. of Systems and  
Information Engineering  
University of Virginia  
lb3dp@virginia.edu

## ABSTRACT

Suicide is the second leading cause of death among young adults but the challenges of preventing suicide are significant because the signs often seem invisible. Research has shown that clinicians are not able to reliably predict when someone is at greatest risk. In this paper, we describe the design, collection, and analysis of text messages from individuals with a history of suicidal thoughts and behaviors to build a model to identify periods of suicidality (i.e., suicidal ideation and non-fatal suicide attempts). By reconstructing the timeline of recent suicidal behaviors through a retrospective clinical interview, this study utilizes a prospective research design to understand if text communications can predict periods of suicidality versus depression. Identifying subtle clues in communication indicating when someone is at heightened risk of a suicide attempt may allow for more effective prevention of suicide.

## Author Keywords

suicide; mental health; social media; depression; text messages

## ACM Classification Keywords

H.1.2 User/Machine Systems; I.5 Pattern Recognition; J.3 Life and Medical Sciences: Health; J.4 Social and Behavioral Sciences: Psychology

## INTRODUCTION

Suicide is a serious public health problem, accounting for 41,000 deaths in the United States each year. In fact,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

© 2018 ACM. ISBN 978-1-4503-5620-6/18/04...\$15.00

DOI: <https://doi.org/10.1145/3173574.3173987>

in the United States, one person attempts suicide every 38 seconds and an average of 94 individuals complete a fatal suicide attempt each day [10]. The overall suicide rate in the United States rose by 24% from 1999 to 2014, according to the National Center for Health Statistics [44]. Suicide is especially pervasive among young people, as the second leading cause of death among individuals 15-24 years of age [11], and is increasing in prevalence. Given the staggering toll of suicide, it is catastrophic that current methods for identifying those at highest risk of suicide remain woefully ineffective and are no better than chance [12].

Thus far, the majority of suicide research in the field of psychology has focused on identifying general risk factors for suicide (e.g., age, gender, psychiatric history) by primarily comparing individuals with suicidality to control participants without suicidality [25]. For example, a history of a mental disorder, including depression, is a well-established risk factor for suicide; prior research has estimated that among individuals who die by suicide, over 90% have a history of a mental disorder [6, 18, 24] and 50-60% suffer from depression or another mood disorder [30]. However, the level of risk dramatically increases as an individual progresses in suicidal thoughts and behaviors: 34% of suicide ideators go on to make a suicide plan; 72% of individuals with a suicide plan go on to make an attempt; and 26% of ideators without a plan make an unplanned attempt [29].

Suicidality, or suicidal self-harm, is any thought or behavior taken with some intention of dying and comprises a continuum of increasing severe behaviors. Suicidal ideation is defined as having thoughts of harming oneself. A suicide attempt is any non-fatal self-harming behavior performed with some intention of dying. Suicide, or completed suicide, is an intentional self-harming behavior that results in death.

Our chief method for assessing acute suicide risk remains clinicians' judgments, which, unfortunately, do not accurately predict future suicidal behaviors [31]. There are a number

of reasons why this might be, such as an inability or unwillingness among suicidal individuals to accurately assess their current level of risk or an intentional desire to conceal suicidal thoughts or intentions. For example, one study found that 78% of hospital inpatients who died by suicide denied having suicidal thoughts during the last verbal communication [4]. Thus, there is an urgent need for novel, data-driven tools to assess acute suicide risk. We need to predict not only *who*, in general, is at heightened risk for suicide, but also *when* that person is at increased risk.

The rising use of smartphones and content-sharing services such as email, blogs, crowd-source sites, and social media has resulted in a proliferation of unstructured text data. Applying text mining techniques to person-generated data, such as text messages (i.e., short message service [SMS]), may identify how communication patterns and media use change as an individual's risk state increases (e.g., from depression to suicidal ideation to suicide attempt [29]). According to Pew Research Center, 99% of Millennials use the internet and 92% own a smartphone in the United States [39]. Millennials, the first generation to be immersed in technology and social media [48], are an especially vulnerable population as suicide is a leading cause among individuals aged 15 to 34 [10, 45]. The vulnerability of these digital natives to suicide make them an ideal population to pilot a study focused on our primary research question: Can text mining identify periods of increasing risk states for suicidality (e.g., depression to suicidality) based on everyday communications?

### Contributions

Towards this question, we make the following two significant contributions in this paper:

- We describe the design and collection of SMS, collected as part of a larger multimodal dataset built specifically to identify unique patterns of communication that occur in advance of a suicide attempt. Existing research on patterns of communication and suicidality have focused predominantly on social media. However, there has been little to no research on identifying periods of imminent suicide risk from everyday SMS communication patterns.
- We present a deep neural net (DNN) to model the within-subject difference in communication patterns of individuals during periods of suicidality compared to depression based on the daily content of their SMS. Our model shows the potential to distinguish language used during periods of suicidality (i.e., suicidal ideation and non-fatal suicide attempts, both known risk factors for completed fatal suicide), from language used during periods of depression, a less severe risk factor for fatal suicide, with a sensitivity of 81% and false alarm rate of 44%. Distinguishing suicidality from depression would be an important advance given that the shift to active suicidality indicates a more serious risk state that clinicians, patients, and families need to identify to detect increasing imminence of risk.

Applying data-driven techniques, such as those presented in this paper, will help address a serious gap in suicide research by allowing for within-subject comparisons as an individual

transitions into higher risk states, providing insight into when someone is at heightened risk.

### BACKGROUND AND RELATED WORK

Recent data-driven efforts in suicide research have focused on mining language contained in suicide notes, clinical notes, and social media to identify individuals at increased risk of suicide. These studies have mostly focused on using approaches based on frequencies of word occurrence (i.e., how often a word appears in a document) or features based on Linguistic Inquiry Word Count (LIWC) [35], a well-validated lexicon of categories of words with psychological relevance [33].

In the clinical context, Pestian et al. [36] compared suicide notes of suicide completers to fabricated notes from non-suicidal control participants using a model with hand-crafted features, such as number of misspellings, number of paragraphs, and readability. The model outperformed psychiatry trainees and mental health professionals in identifying real suicide notes. Poulin et al. [38] compared veterans who died by suicide to two cohorts, veterans with no history of visiting mental health services and veterans with at least one psychiatric hospitalization, using machine learning with features based on the frequency of word occurrence in unstructured clinical notes contained in electronic health records (EHRs) of the United States Veterans Administration. Features that described behaviors and physical symptoms known to be markers of suicide risk, such as agitation, were most predictive of the suicidal veterans [38]. A limitation of these studies is that they use clinicians' judgment and data collected post-suicide. Furthering this work, Pestian et al. [37] compared transcripts of responses to a five-question survey administered to adolescent suicidal patients and orthopedic patients admitted to an emergency department. This study found the frequency of word occurrence in the responses to be highly predictive of suicidality. Finally, Cook et al. [8] identified heightened psychiatric symptoms of patients recently discharged from the emergency department or hospital using survey responses to a questionnaire administered through SMS. The survey included structured questions inquiring about mental and physical health and one free text question inquiring about general mood. The unstructured responses were analyzed using frequency of word occurrence.

More recent data-driven suicide research efforts have shifted focus to monitoring social media. De Choudhury et al. [9] compared individuals who transitioned from posting on online mental health subreddits to a suicide watch subreddit to individuals who only posted on mental health subreddits using LIWC, frequency of word occurrence, and hand-crafted features such as posting activity, pronoun usage, and linguistic form, providing insight into language markers indicating suicidal ideation. Braithwaite et al. [3] compared tweets of suicidal to non-suicidal participants, labeled according to a screening tool for suicidal symptoms, using a model with LIWC variables. Their study provided evidence that short messages can provide sufficient information to differentiate suicidal from non-suicidal individuals.

Previous research suggests that language, perhaps unbeknownst to the speaker, may indeed provide clues that are

indicative of suicidal intent. As can be seen, most prior work focuses on comparing an individual with suicidality to a control (i.e., non-suicidal) individual. These studies build on the body of research to identify *who* is at risk, but it remains crucial to identify *when* someone is at risk. To the best of our knowledge, this is the first study to identify periods of *known* increasing suicidal risk (i.e., depression to suicidality) *within* an individual using a *novel* data source comprised of SMS.

## DATA

We provide a description of the data collection process, as related to the data used in this paper, followed by descriptive statistics of the data. The data used in this study was collected as part of a larger effort to form a comprehensive multimodel dataset that includes personal communication (i.e., SMS, emails, and call history), social media data (i.e., Twitter and Facebook), web browsing history, and mental health history. The study protocol was approved by the Social and Behavioral Sciences Institutional Review Board (IRB) at the University of Virginia (UVa).

### Data Collection

Prior to data collection, an online survey was distributed to the Department of Psychology's undergraduate participant pool at UVa to evaluate students' communication habits using various electronic services [17].

Of the 796 students who participated in this survey, individuals highly endorsed regularly using SMS (95.1%) and email (87.7%) for writing personal messages intended for an individual or group to see, followed by Facebook (63.7%) and Twitter (31.9%). Those endorsing SMS reported sending personal messages many times a day, compared to other services, which were used once a day (e.g., email) or less than once a day (e.g., Facebook, Twitter). On a Likert scale from 1 to 5 of likelihood of using a particular service to send emotionally expressive messages, SMS had the highest likelihood (3.9), followed by Facebook (2.9), Twitter (2.7), and email (1.6).

The data collection process consisted of two phases: recruitment and the laboratory study, as described below. Figure 1 presents the steps involved in the data collection process.

#### Recruitment

Participants were recruited from the undergraduate participant pool to complete a 2-hour laboratory session. Participants received either course credit or \$40 for participating in the study. Prior to a participant being invited into the laboratory, participants were pre-screened for eligibility using two online surveys and a phone screen.

#### Online Survey Screen

An initial online survey was distributed to the undergraduate participant pool. The initial survey included the question "Have you ever had a period of sadness in the past during which you felt hopeless?" and included an option to be contacted about possible participation in studies asking about this time period in their life. Of the 2,377 students who participated in the initial survey, 1,478 (62.2%) indicated a period of past sadness.

A follow-up two-question survey was emailed to individuals who answered yes to the initial survey question and consented to be contacted (n=1,211). The second survey included the questions: "Have you ever made a suicide attempt?" and "Have you ever had thoughts of wanting to kill yourself?". Of the 871 students who participated in the follow-up survey, 593 (68.1%) indicated thoughts of killing themselves and 87 (10.0%) endorsed a past suicide attempt. Individuals who endorsed a past suicide attempt were emailed and invited to participate in a phone screen to see if they qualified for the study.

**Phone Screen.** During the phone screen, participants were provided with more information about the study and the interviewer ensured that inclusion criteria for the study were met. Inclusion criteria included: (1) confirmation of past suicidal thoughts and behaviors; (2) adult status (at least 18 years old); (3) availability and access to personal messaging data dating back to prior significant life events (i.e., suicide attempts); and (4) minimal or no self-reported current desire to die and no current suicide plan or intent (determined by a suicide risk assessment tool). Any individuals who were determined to be at "high or imminent risk" were excluded from participation and referred to clinical care. Of the 77 students who consented to a phone screen, 52 (67.5%) completed the phone screen and 42 (80.7%) qualified to complete the laboratory study.

#### Laboratory Study

The laboratory procedure included downloading the participant's communication data, an interview with the participant, and completion of questionnaires. The laboratory study was conducted in Spring and Fall 2016.

**Data Download.** For transparency, participants downloaded their own SMS data, with the assistance of the experimenter, using third-party software to extract the data. This was part of a larger data collection effort that also included call history, Facebook, Twitter, Gmail, Google Hangouts, and Google Chrome search history.

**Interview: Identification of Mental Health Episodes.** Participants were asked by the experimenter to identify up to three episodes for the following events in their life: (1) suicide attempts, (2) suicidal ideation (with no attempt), and (3) depression (with no suicidal ideation or attempt). Regarding suicide ideation and depression episodes, participants were asked to identify specific periods lasting two weeks during which they remember having the particular experience. Regarding suicide attempts, participants were asked to identify the exact date of each past attempt, and the two week period immediately prior to the attempt was considered the suicide attempt "episode." Each episode was set at two weeks long in a conservative effort to capture the critical period of increased ideation, planning, and intent leading up to a suicide attempt (given prior research indicating that actions precipitating suicide attempts typically occur within one week of the attempt [26]). Periods of positive mood were also identified by participants as part of the larger data collection. In addition to providing specific dates for the episodes, the participants were asked of their certainty of the identified dates. Interviews with participants were performed by a PhD student from the

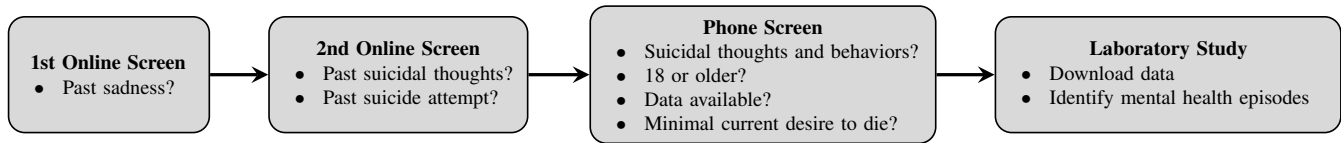


Figure 1. Schematic diagram of the data collection process.

Department of Psychology trained in conducting suicide interviews and risk assessments under the supervision of a licensed clinical psychologist.

#### Ethical Considerations

Prior research indicates that asking young adults with a history of suicide attempts about suicide does not cause an increase in psychological distress or increase suicidal thoughts or behaviors, either immediately following an assessment [14] or several years after an assessment [41]. Nonetheless, to assess any changes as a consequence of the laboratory study, participants were asked to rate their negative mood and desire to die (measured on Likert scales from 0 to 10) at the beginning and end of the laboratory study. Risk mitigation plans were in place for participants who experienced a significant increase in negative affect or suicidality. Fortunately, as expected based on prior research, no participants expressed a significant increase in their desire to die following the interview. Only one participant reported an increase in their desire to die (1 pre-interview to 2 post-interview). On average, participants' negative mood did not change significantly (3.7 pre-interview to 3.9 post-interview) and desire to die decreased slightly (0.8 pre-interview to 0.6 post-interview).

#### Description of Data

A total of 33 participants with periods of suicidality took part in the study; however, extracted SMS data was limited to only 26 of the participants. Six of the participants were omitted because of a software error during the laboratory study preventing collection of SMS data and one participant did not have data stored on their phone during the time frame of their identified mental health episodes. It should be noted that participants identified more mental health episodes than presented in this study; however, we are limited by the time frame of data currently stored on the participant's phone.

#### Participants

Table 1 presents descriptive statistics about the participants.

#### Episodes

Table 2 presents the descriptive statistics about the two-week periods identified as mental health episodes. Given the small sample size, episodes of ideation and attempt were combined into a single episode of suicidality.

#### Text Messages

A total of 1,029,481 incoming and outgoing messages were downloaded from the 26 participants. Messages were subset to include only outgoing messages ( $n=469,362$ ) and subset again to include only messages that occurred within the time frame of the episodes identified by the participants, resulting in 136,347 outgoing messages. Messages were then converted

into a daily time window for each participant, resulting in a total of 1,213 days of messages with 528 days occurring during time periods identified as depression and 685 days occurring during time periods identified as suicidality. On average, both depression and suicidality episodes contain 13 days of messages.

No. of Participants	
Total	26
Gender	
Female	22
Male	4
Age	
Mean	20.42
Std Dev	2.55
Ethnicity	
Hispanic	3
Non-Hispanic	23
Race	
White	17
Asian	4
Black	2
Multiple/Other	3

Table 1. Descriptive statistics about the participants.

No. of Episodes	
Suicidality	54
Certain	44
Uncertain	10
Average Episodes per Participant	2
Depression	40
Certain	39
Uncertain	1
Average Episodes per Participant	2

Table 2. Descriptive statistics about the episodes.

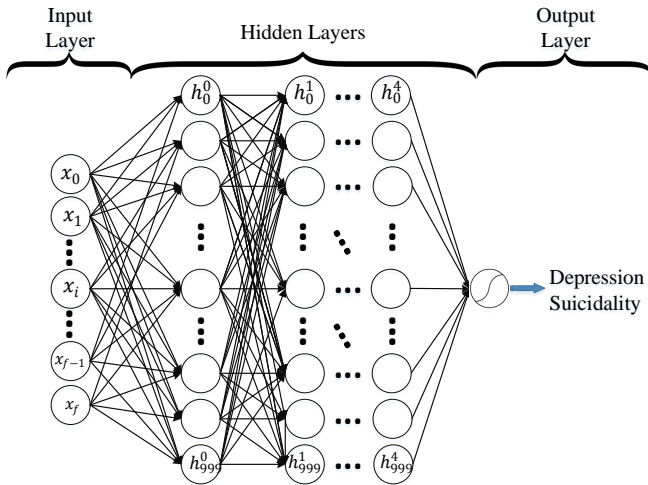
## METHODS

For our prediction task, we used supervised machine learning to build classifiers that predict a binary classification of depression versus suicidality using a day of text messages as an observation. Two sets of features were used to construct the classifiers: (1) a feature set comprised of psycholinguistic features, specifically LIWC variables, and (2) a feature set of word occurrence, specifically term frequency-inverse document frequency (tf-idf) of unigrams (i.e., an individual word).

### Features

#### Psycholinguistic Features

Each observation (i.e., day of text messages) was scored using the 2015 version of LIWC. LIWC is a language analysis



**Figure 2. Deep Neural Net Architecture.**  $x$  = input feature, where subscript  $f$  is the total number of features.  $h$  = hidden layer, where superscript indicates the hidden layer and subscript indicates the node.

software that analyzes and outputs scores for linguistic and psychological dimensions of language. For our analysis, we used the word count, mean words per a sentence, percentage of words with over six letters, 21 linguistic features (e.g., first-person pronouns), 41 psychological constructs (e.g., affect), six personal concerns (e.g., work), five informal language markers (e.g., swear words), 12 punctuation categories (e.g., exclamation marks), and four variables that summarize the language including analytical thinking, clout, authenticity, and emotional tone. With the exception of six LIWC variables, the score for each observation is expressed as the proportion of words belonging to a given category (e.g., score of 5.0 on 'Anger' indicates 5% of words in the day of messages belonged to the 'Anger' category); the other six LIWC variables include the number of words in an observation, mean words per sentence, and four summary variables that are standardized scores based on previously published research [35].

#### Word Occurrence

We first pre-processed the text by removing special characters and then measured the frequency of unigrams using tf-idf. Tf-idf measures the frequency of word occurrence contained in a single daily message (hence term frequency) and offsets this by the frequency of word occurrence in all daily messages (hence inverse document frequency). This measure adjusts for words that tend to appear more frequently in the daily messages. We chose to include stopwords in the model given this category of words could provide additional insight and prior suicide research did not indicate that such words should be excluded.

#### Prediction Task

We built several binary classifiers using the previously described feature sets and implementing 10-fold cross-validation to determine the best fit. Cross-validation was stratified to maintain the original class balance of 56%. In this paper, we present the results of our best performing classifier, a DNN

compared to a generally robust performing text classification technique, support vector machine (SVM)<sup>1</sup>. Predictive analysis was implemented in Python using the scikit-learn [34] and Keras [7] libraries.

We altered the depth and width of the DNN and present the results for the best performing DNN. The final architecture of the DNN included the input layer, five hidden layers containing 1,000 nodes each, and the output layer. The input layer consists of the text features as outlined in the Features section. The nodes in the hidden layers use a rectified linear unit (ReLU) [28, 21] activation function. The output layer is a single sigmoid activation function which identifies the binary classification. All layers are fully connected and each layer only receives input from the previous layer and outputs to the next layer. Additional details on this architecture can be found in [21]. We trained the model using the standard back-propagation algorithm with the Adam optimizer [20] and categorical cross entropy loss function. Dropout was used to avoid overfitting [43]. Figure 2 presents a schematic of the final DNN architecture.

## RESULTS

Table 3 presents the results, averaged across folds, of the SVM and DNN classifiers using the two feature sets of LIWC variables and tf-idf. Our best performing model is the DNN using the tf-idf feature set. The accuracy of the model is 70% compared to the default accuracy of a classifier that predicts the most prevalent class, which is 56%. The recall (i.e., sensitivity) of the model is 81% and the specificity is 56% with a 44% false alarm rate. Sensitivity is the percentage of the suicidal observations that were correctly classified (i.e., true positives divided by all positives). Specificity is the percentage of the depression observations that were correctly classified (i.e., true negatives divided by all negatives). The false alarm rate is the percentage of observations originating during a time period of depression (with no suicidality) that were incorrectly classified as suicidal (i.e., false positives divided by all negatives).

As shown in Table 3, the models that use a feature set of LIWC variables are less discriminatory for our data. Based on the performance and reading a subsample of the daily messages, we hypothesize that this may be due to high variation in usage of formal language and expression of emotional content among each participant. LIWC is a dictionary-based approach (i.e., this approach is limited to the words that are contained in the lexicon) whereas tf-idf allows the model to "learn" all words used by the participant. Informal language includes shorthand abbreviations, such as "u" for "you", "ikr" for "I know, right", and "tbh" for "to be honest"; variations of the same word "yay", "yayayay", and "yayayayayay"; colloquialisms; and improper sentence structure (e.g., "hehe thanks"). This is not an exhaustive list and the messages contain more obscure examples of informal language. Additionally, some of the participants' messages are more emotionally expressive and we hypothesize that this may be more transparent

<sup>1</sup>Code is shared as an open source tool at <https://github.com/BarnesLab/Identification-of-Imminent-Suicide-Risk-Among-Young-Adults-using-Text-Messages>

Model	Features	Performance Metrics									
		Accuracy		Precision		Recall		Specificity		F-1	
		M	SD	M	SD	M	SD	M	SD	M	SD
SVM	LIWC	0.53	0.03	0.56	0.02	0.81	0.05	0.17	0.03	0.66	0.03
DNN	LIWC	0.63	0.03	0.65	0.03	0.77	0.07	0.46	0.10	0.70	0.03
SVM	tf-idf	0.64	0.03	0.64	0.02	0.83	0.05	0.40	0.07	0.72	0.02
DNN	tf-idf	0.70	0.04	0.71	0.05	0.81	0.08	0.56	0.12	0.75	0.03

Table 3. Results for classification of depression versus suicidality. M = mean. SD = standard deviation.

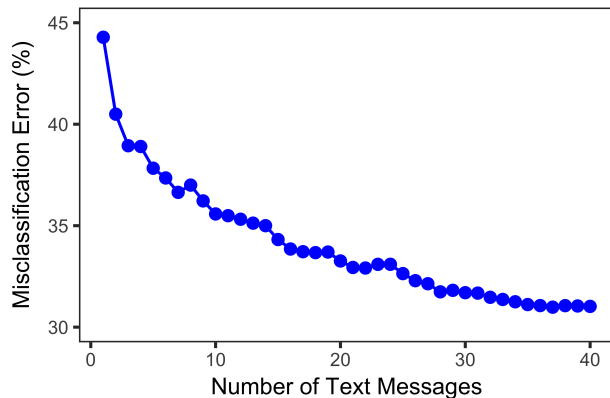


Figure 3. Misclassification error compared to daily number of text messages.

in a model that accounts for the hierarchical structure of the data (i.e., daily messages originate from a single participant).

The content of daily messages compared to the prediction assigned by the DNN tf-idf classifier revealed that the total number of characters in incorrectly classified daily messages was 28% less, on average, than the total number of characters in the correctly classified daily messages. On average, days that were incorrectly classified contained 1,748 characters compared to days that were correctly classified with 2,427. Assuming the standard limitation of SMS length of 160 characters, that is 11 messages compared to 15 messages for incorrectly classified and correctly classified days, respectively. Figure 3 shows that classification error is reduced as the length of characters contained in the daily content is increased. Seventeen percent of the days that were incorrectly classified contain the equivalent of less than one text (i.e., less than 160 characters) and frequently reference common expressions such as “Thanks”, “No problem”, “Will you pick me up?”, and “I’m busy. Call back later.” These expressions are likely not discriminatory for our prediction task. Misclassification was stratified across participants indicating that no single participant contributed to a boost in performance. In order to protect the privacy of the participants, more specific, explicit examples of messages are excluded from this paper.

To the best of our knowledge this is the first study analyzing the relationship between everyday SMS communications and suicidality. Furthermore, SMS data differs considerably from previous research applying text mining for suicide prevention in social media and clinical documentation, given language from SMS data is drawn from one-on-one conversations intended to be private. We hypothesize that since SMS may be

targeted to many audiences (e.g., friends versus family), the linguistic style may shift and vary in usage of formal language, brevity, and expression of emotions depending on who the individual is texting.

### LIMITATIONS

Selection bias was reduced by using a case-crossover design (i.e., the individual case serves as his/her own control), but we are limited by studying only participants in the university undergraduate community, limiting generalizability. Also, the results may be affected by recall bias if participants inaccurately recalled the time periods of episodes. To help address this concern, participants were allowed to look back through their calendars, social media, and SMS when selecting the dates of each episode. Additionally, participants were asked how certain they were when identifying the date(s) of the episode (i.e., ranging from very certain that the recalled date is correct to the date may be incorrect by more than two weeks).

Another limitation of the study is use of third-party software to extract SMS data. SMS data is limited to the data currently stored on the phone and the format of the data varies based on the software.

Our prediction task is limited by the usage of a DNN which does not allow straightforward, meaningful feature extraction hindering the ability to rank features based on their ability to predict depression versus suicidality.

Finally, the financial and time commitment for the data collection was substantial. Crowdsourcing an online data collection process, such as [32], may be an option in the future to reduce study costs and labor.

### DISCUSSION

The current study utilized a novel laboratory-based research design to identify suicide risk based on text messaging patterns. This is the first study, to our knowledge, to examine the association between private SMS data and suicidal behaviors. A supervised DNN classifier based on daily SMS content achieved strong precision and recall (i.e., sensitivity) for correctly assigning days during which participants experienced suicidal thoughts/behaviors versus days during which participants reported being depressed but indicated no suicidality.

### Implications for Human-Computer Interaction in the Clinical Setting

This research may be useful for monitoring during periods of distress, and during and following treatment for mental health problems to identify high-risk states, provide real-time

alerts of escalating risk states, and provide appropriate interventions. Monitoring of patients requires a substantial commitment from the provider and health system [46] but passive monitoring through ubiquitous technologies, such as the smartphone, could reduce this burden.

#### *Utility in Various Clinical Contexts*

This research has potential utility in the context of a support system that aids in prevention and intervention to reduce suicidal behavior and mortality, including as part of an aftercare plan following hospitalization and in outpatient settings. Research has shown that the risk of suicide is especially high following discharge from psychiatric hospitalization [40]. There is some evidence to suggest that short-term programs can be effective in reducing suicide attempts following discharge. For example, the Norwegian multidisciplinary chain-of-care networks provide follow-up care after hospital care to those who attempt suicide, and this follow-up care has been shown to reduce treatment dropout rates and repeat attempts [23]. Another intervention study found that sending supportive letters to randomly selected patients discharged following a suicide attempt significantly decreased the rate of suicide death among patients [27]. However, such crisis management programs are very labor- and resource-intensive. Therefore, tools that are both effective and practically feasible so that they can be widely disseminated are needed during this critical period to monitor risk.

As part of an aftercare plan, patients could be given the option of installing a passive screening system on their smartphone to collect text data and produce a predictive model of acute risk. If the predictive model indicated a patient was at acute risk, the clinician could provide an intervention or supportive responses. These types of systems may be helpful for monitoring response to treatment and helping to prevent relapse. The benefit of such systems is that the data help determine objective level of risk so that the entire burden does not fall on the individual or their clinician. This system incorporates many elements of treatment that have been shown to work for suicide prevention, including: (1) allowing patients to understand how their thoughts are incorporated into a model of suicidality indicating increasing risk, (2) a focus on treatment compliance (i.e., treatment is only effective if the patient is engaged), (3) patients taking personal responsibility for their treatment, (4) patients are guaranteed easy access to crisis services, and (5) patient writing being a part of their treatment experience [42]. This type of system is an example of linking an individual's everyday and clinical data with digital technology to create a nuanced view of human disease [19].

Such passive monitoring tools could also be useful to monitor individuals not currently in crisis but who may be receiving ongoing outpatient treatment. Such treatment may take place for an hour once a week, leaving a lot of time between sessions during which risk cannot be directly assessed. Even for individuals at general risk of suicide, based on a collection of risk factors, it is not feasible for clinicians to assess for suicide risk in an ongoing basis. A passive monitoring tool using text messages could provide a means to more continuously assess suicide risk, detecting any unexpected spikes in risk levels.

#### *Forms of Intervention*

Although the possible clinical applications of this work could help address a major public health problem, the development of a predictive tool would raise a number of practical and ethical issues, such as what types of interventions to undertake and who would be notified when a high-risk alert is triggered.

When the general risk level is low (i.e., not in the context of recent acute suicidality), one possible intervention is a supportive text message sent to the individual prompting them to respond yes or no if they wish to receive additional outpatient or emergency support. If the response is yes, it would be important that the protocol be clearly established by the treatment team with the individual beforehand regarding what next steps would be taken (e.g., phone call). If the response is no, the individual could be automatically sent a general online intervention or a personalized resource, such as an existing suicide safety plan serving as a reminder of the individual's coping and means restriction strategies when feeling suicidal. Given a past intervention study indicating that sending supportive letters to recent attempters decreased the rate of reattempt [27], such a system could serve as a digital equivalent and may itself have therapeutic value.

This work raises important questions about who would be alerted if risk were to escalate. An alert system could be implemented by the individual's therapist or as part of a larger treatment team, such as primary care physicians. An individual's personal support network, such as family members, could also be notified, which may be especially useful for adolescents. However, such a system raises a number of legal and ethical issues. The field would need to answer questions related to mandated reporting and involuntary hospitalization. For example, would a clinician be legally and ethically mandated to intervene as they would if a patient endorsed active suicide intent in person? What is the most appropriate action for someone who denies having suicidal thoughts, plans, or intent but whose text messages indicate elevated risk? For a predictive monitoring tool to be effectively implemented, these and other questions would need to be addressed.

#### *Balancing Benefits and Risks*

In this study, the models achieved better sensitivity than specificity, meaning there were more false positive than false negatives. Similar to diagnostic medical tests that produce certain levels of false positives and false negatives, decisions would need to be made regarding the most appropriate threshold for what would be considered "elevated risk" deserving of intervention. For example, is it preferable to flag more individuals but with less certainty of risk (producing more false positives) or fewer individuals but with greater certainty of risk (producing more false negatives)?

The decision on how to tune the model to balance the trade-off between sensitivity and specificity may depend on the specific clinical issue being addressed and the resources required for such an intervention. For example, in an outpatient setting, if the alert is meant to serve as a preventive measure and reminder of suicide safety skills already discussed between the individual and clinician, a lower detection threshold may be

warranted because of the relatively low cost of a text message intervention.

A strength of this study as a basis for a future tool is that we were able to specifically detect suicidality separate from depression; this is important because depressive symptoms often accompany suicidality but they are also part of many other psychological difficulties that do not typically escalate to the level of suicidality. Thus, detecting only that a person is experiencing negative mood, without the further identification specific to increased suicide risk, would not allow for the targeted interventions needed to reduce suicidal behavior and mortality. To further increase precision, it may even be possible to tune an algorithm to how a given individual uses language over time. This would be an exciting future step to build on the current results based on a normalized sample.

While many open questions remain about the best ways to increase predictive accuracy, the current results point to the promise of behavioral tools like these to one day significantly reduce the rates and public health burden of suicide.

#### *Fusion with Other Data Streams*

While the models produced a high number of false positives, we hypothesize that text messaging data could be fused with both clinical data and other smartphone data streams to improve performance of risk models. For example, recent work has used longitudinal electronic health record data for early detection of mental health disorders [50, 49] and future suicide attempt [47, 1]. Recent advances have made it possible to passively monitor how human behavior unfolds in people's natural settings by leveraging sensors embedded in personal smartphones (e.g. GPS, accelerometers, etc.) [16, 13, 22]. This work has shown great promise for in the development of digital biomarkers that characterize mental health status [5, 15]. These other heterogeneous data streams characterizing an individual's behaviors and clinical risk are likely to be important contextual factors in assessing suicide risk, and integrating these contextual details will be an important direction for future work.

#### **CONCLUSIONS**

Our study was designed to investigate temporally sensitive patterns in communication that predict acute suicidal thoughts and behaviors. By comparing communication patterns during periods immediately preceding a suicide attempt and periods of high ideation versus depression but non-suicidal periods of their life, we aim to isolate specific communication that characterizes acute suicide risk. This research provides evidence that language changes as an individual transitions from depression to suicidality (i.e., suicidal ideation and non-fatal suicide attempts), indicating an increasing level of suicide risk. Future research should explore whether individualizing the models produces better performance that is calibrated to an individual's specific expression of language during increasing risk states.

Although depression is a risk factor for suicide, research indicates that only 2-8% of individuals with a mood disorder will go on to kill themselves [2]. Therefore, depression in itself is not clinically useful for identifying high-risk individuals.

Further, even if known risk factors are used to indicate high risk individuals, they cannot tell us *when* such individuals are at particularly elevated risk. Employing data-driven techniques, such as those developed in this study, could identify *when* individuals are at heightened risk and help direct appropriate resources to these individuals. As Nock et al. [29] report, "the biggest shortcoming in suicide research to date" is "the inability to dramatically decrease rates of suicidal behavior and mortality despite decades of research and associated commitment of resources." This research may enable new ways to identify not just *who* is at risk for a suicide attempt, but also *when* a given person increases in their risk state and acutely needs services.

#### **ACKNOWLEDGMENTS**

The authors thank Eric Rochester, Scott Bailey, and Abbie Starns for their valuable assistance wrangling and organizing the data for this study. This study was supported by the UVa Presidential Fellowships in Data Science and Hobby Postdoctoral and Predoctoral Fellowships in Computational Science, and A.L. Nobles was supported by NIH 5T32LM012416.

#### **REFERENCES**

1. Yuval Barak-Corren, Victor M. Castro, Solomon Javitt, Alison G. Hoffnagle, Yael Dai, Roy H. Perlis, Matthew K. Nock, Jordan W. Smoller, and Ben Y. Reis. 2017. Predicting Suicidal Behavior From Longitudinal Electronic Health Records. *Am J Psychiatry* 174, 2 (Feb 2017), 154–162. DOI : <http://dx.doi.org/10.1176/appi.a.jp.2016.16010077>
2. George W Blair-West, Chris H Cantor, Graham W Mellsop, and Margo L Eyson-Annan. 1999. Lifetime suicide risk in major depression: sex and age determinants. *J Affect Disord* 55, 2 (Oct 1999), 171–178.
3. Scott R Braithwaite, Christophe Giraud-Carrier, Josh West, Michael D Barnes, and Carl Lee Hanson. 2016. Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality. *JMIR Ment Health* 3, 2 (May 2016), e21. DOI : <http://dx.doi.org/10.2196/mental.4822>
4. Katie A Busch, Jan Fawcett, and Douglas G Jacobs. 2003. Clinical correlates of inpatient suicide. *J Clin Psychiatry* 64, 1 (Jan 2003), 14–19.
5. Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 1293–1304. DOI : <http://dx.doi.org/10.1145/2750858.2805845>
6. Jonathan T. Cavanagh, Alan J. Carson, Michael Sharpe, and Stephen M. Lawrie. 2003. Psychological autopsy studies of suicide: A systematic review. *Psychol Med* 33, 3 (Apr 2003), 395–405.
7. François et al. Chollet. 2015. Keras. <https://github.com/fchollet/keras>. (2015).



8. Benjamin L. Cook, Ana M. Progovac, Pei Chen, Brian Mullin, Sherry Hou, and Enrique Baca-Garcia. 2016. Novel Use of Natural Language Processing (NLP) to Predict Suicidal Ideation and Psychiatric Symptoms in a Text-Based Mental Health Intervention in Madrid. *Comput Math Methods Med* 2016 (Sep 2016), 1–8. DOI : <http://dx.doi.org/10.1155/2016/8708434>
9. Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In *Proceedings of the 2016 SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2098–2110. DOI : <http://dx.doi.org/10.1145/2858036.2858207>
10. Emory University. 2017. Emory Cares 4 U: Suicide Statistics. (2017). Retrieved October 24, 2016 from <http://www.emorycaresforyou.emory.edu/resources/suicidestatistics.html>.
11. U.S. Centers for Disease Control and Prevention. 2014. Web-based Injury Statistics Query and Reporting System (WISQARS). (2014). Retrieved March 31, 2014 from <http://www.cdc.gov/injury/wisqars>.
12. Joseph C. Franklin, Jessica D. Ribeiro, Kathryn R. Fox, Kate H. Bentley, Evan M. Kleiman, Xieyining Huang, Katherine M. Musacchio, Adam C. Jaroszewski, Bernard P. Chang, and Matthew K. Nock. 2017. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychol Bull* 43, 2 (Feb 2017), 187–232. DOI : <http://dx.doi.org/10.1037/bul0000084>
13. Yusong Gao, Ang Li, Tingshao Zhu, Xiaoqian Liu, and Xingyun Liu. 2016. How smartphone usage correlates with social anxiety and loneliness. *PeerJ* 4 (Jul 2016), e2197. DOI : <http://dx.doi.org/10.7717/peerj.2197>
14. Madelyn S Gould, Frank A Marrocco, Marjorie Kleinman, John Graham Thomas, Katherine Mostkoff, Jean Cote, and Mark Davies. 2005. Evaluating Iatrogenic Risk of Youth Suicide Screening Programs: A Randomized Controlled Trial. *JAMA* 293, 13 (Apr 2005), 1635–43. DOI : <http://dx.doi.org/10.1001/jama.293.13.1635>
15. Yu Huang, Jiaqi Gong, Mark Rucker, Philip Chow, Karl Fua, Matthew S. Gerber, Bethany Teachman, and Laura E. Barnes. 2017. Discovery of Behavioral Markers of Social Anxiety from Smartphone Sensor Data. In *Proceedings of the 1st Workshop on Digital Biomarkers (Digital Biomarkers '17)*. ACM, New York, NY, USA, 9–14. DOI : <http://dx.doi.org/10.1145/3089341.3089343>
16. Yu Huang, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A Teachman, and Laura E Barnes. 2016. Assessing social anxiety using gps trajectories and point-of-interest data. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 898–903.
17. Institutional Review Board for Social and Behavioral Sciences at University of Virginia. 2014. Psychology Pool. (2014). [http://www.virginia.edu/vpr/irb/sbs/resources/\\_guide/\\_pools/\\_existing/\\_psych.html](http://www.virginia.edu/vpr/irb/sbs/resources/_guide/_pools/_existing/_psych.html).
18. ET Isometsä. 2001. Psychological autopsy studies – a review. *Eur Psychiatry* 16, 7 (Nov 2001), 379–385.
19. Sachin H Jain, Brian W Powers, Jared B Hawkins, and John S Brownstein. 2015. The digital phenotype. *Nat Biotechnol* 33, 5 (May 2015), 462–463. DOI : <http://dx.doi.org/10.1038/nbt.3223>
20. Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014). Retrieved from <https://arxiv.org/abs/1412.6980>.
21. Kamran Kowsari, Donald E Brown, Mojtaba Heidarysafa, Kiana Jafari Meimandi, , Matthew S Gerber, and Laura E Barnes. 2017. HDLTex: Hierarchical Deep Learning for Text Classification. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on*. IEEE.
22. Neal Lathia, Veljko Pejovic, Kiran K Rachuri, Cecilia Mascolo, Mirco Musolesi, and Peter J Rentfrow. 2013. Smartphones for large-scale behavior change interventions. *IEEE Pervasive Computing* 12, 3 (2013), 66–73.
23. J John Mann, Ann Haas, Lars Mehlum, and Michael Phillips. 2005. Suicide Prevention Strategies. *JAMA* 294, 16 (Oct 2005), 2064–2074.
24. Mauri J Marttunen, Hillevi M Aro, and Jouko K Lönnqvist. 1993. Adolescence and suicide: A review of psychological autopsy studies. *Eur Child Adolesc Psychiatry* 2, 1 (Jan 1993), 10–18. DOI : <http://dx.doi.org/10.1007/BF02098826>
25. Alexis M. May and E. David Klonsky. 2016. What Distinguishes Suicide Attempters From Suicide Ideators? A Meta-Analysis of Potential Factors. *Clin Psychol-Sci Pr* 23, 1 (Mar 2016), 5–20. DOI : <http://dx.doi.org/10.1111/cpsp.12136>
26. Alexander J Millner, Michael D Lee, and Matthew K Nock. 2017. Describing and Measuring the Pathway to Suicide Attempts: A Preliminary Study. *Suicide Life Threat Behav* 47, 3 (Jun 2017), 353–369. DOI : <http://dx.doi.org/10.1111/sltb.12284>
27. Jerome A. Motto and Alan G. Bostrom. 2001. A Randomized Controlled Trial of Postcrisis Suicide Prevention. *Psychiatr Serv* 52, 6 (Jun 2001), 828–833. DOI : <http://dx.doi.org/10.1176/appi.ps.52.6.828>
28. Jinseok Nam, Jungi Kim, Eneldo Loza Mencía, Iryna Gurevych, and Johannes Fürnkranz. 2014. Large-scale multi-label text classification – revisiting neural networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2014)*. Springer Berlin Heidelberg, Berlin, Heidelberg, 437–452. DOI : [http://dx.doi.org/10.1007/978-3-662-44851-9\\_28](http://dx.doi.org/10.1007/978-3-662-44851-9_28)

29. Matthew K. Nock, Guilherme Borges, Evelyn J. Bromet, Christine B. Cha, Ronald C. Kessler, and Sing Lee. 2008. Suicide and Suicidal Behavior. *Epidemiol Rev* 30, 1 (July 2008), 133–154. DOI: <http://dx.doi.org/10.1093/epirev/mxn002>
30. M. K. Nock, A. J. Millner, C. A. Deming, and C. R. Glenn. 2014. *Depression and suicide* (3rd ed.). The Guilford Press, New York, NY, Chapter 24, 448–467.
31. Matthew K Nock, Jennifer M Park, Christine T Finn, Tara L Deliberto, Halina J Dour, and Mahzarin R Banaji. 2010. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychol Sci* 21, 4 (Apr 2010), 511–517. DOI: <http://dx.doi.org/10.1177/0956797610364762>
32. OurDataHelps. 2016. OurDataHelps.org: The world needs more Data Donors. (2016). Retrieved from [ourdatahelps.org](http://ourdatahelps.org).
33. Michael J. Paul and Mark Dredze. 2017. Social Monitoring for Public Health. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 9, 5 (Aug 2017), 1–183. DOI: <http://dx.doi.org/10.2200/S00791ED1V01Y201707ICR060>
34. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12 (2011), 2825–2830.
35. J.W. Pennebaker, R.J. Booth, R.L. Boyd, and M.E. Francis. 2015. Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates ([www.LIWC.net](http://www.LIWC.net)). (2015).
36. John Pestian, Henry Nasrallah, Pawel Matykiewicz, Aurora Bennett, and Antoon Leenaars. 2010. Suicide Note Classification Using Natural Language Processing: A Content Analysis. *Biomed Inform Insights* 2010, 3 (Aug 2010), 19–28. <http://www.ncbi.nlm.nih.gov/pubmed/21643548>
37. John P. Pestian, Jacqueline Grupp-Phelan, Kevin Bretonnel Cohen, Gabriel Meyers, Linda A. Richey, Pawel Matykiewicz, and Michael T. Sorter. 2016. A Controlled Trial Using Natural Language Processing to Examine the Language of Suicidal Adolescents in the Emergency Department. *Suicide Life Threat Behav* 46, 2 (Apr 2016), 154–159. DOI: <http://dx.doi.org/10.1111/sltb.12180>
38. Chris Poulin, Brian Shiner, Paul Thompson, Linas Vepstas, Yinong Young-Xu, Benjamin Goertzel, Bradley Watts, Laura Flashman, and Thomas McAllister. 2014. Predicting the Risk of Suicide by Analyzing the Text of Clinical Notes. *PLoS One* 9, 1 (Jan 2014), e85733. DOI: <http://dx.doi.org/10.1371/journal.pone.0085733>
39. Jacob Poushter. 2016. *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies*. Pew Research Center, Washington, D.C.
40. Ping Qin and Merete Nordentoft. 2005. Suicide risk in relation to psychiatric hospitalization: evidence based on longitudinal registers. *Arch Gen Psychiatry* 62, 4 (Apr 2005), 427–432.
41. Sarah K Reynolds, Noam Lindenboim, Katherine Anne Comtois, Angela Murray, and Marsha M Linehan. 2006. Risky Assessments: Participant Suicidality and Distress Associated with Research Assessments in a Treatment Study of Suicidal Behavior. *Suicide Life Threat Behav* 36, 1 (Feb 2006), 19–34. DOI: <http://dx.doi.org/10.1521/suli.2006.36.1.19>
42. M. David Rudd. 2015. Core competencies, warning signs, and a framework for suicide risk assessment in clinical practice. The Oxford Handbook of Suicide and Self-Injury. *J Clin Psychol* 76 (2015), 323–352. Issue E896. DOI: <http://dx.doi.org/10.1093/oxfordhb/9780195388565.013.0018>
43. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J Mach Learn Res* 15 (Jun 2014), 1929–1958.
44. Sabrina Tavernise. 2016. U.S. Suicide Rate Surges to a 30-Year High. (2016). Retrieved from <http://www.nytimes.com/2016/04/22/health/us-suicide-rate-surges-to-a-30-year-high.html>.
45. U.S. Centers for Disease Control and Prevention: National Center for Health Statistics. 2016. Leading Causes of Death. (2016). Retrieved October 24, 2016 from <http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>.
46. M Valenstein, D Eisenberg, JF McCarthy, KL Austin, D Ganoczy, HM Kim, K Zivin, JD Piette, M Olfson, and FC Blow. 2009. Service implications of providing intensive monitoring during high-risk periods for suicide among VA patients with depression. *Psychiatr Serv* 60, 4 (Apr 2009), 439–44. DOI: <http://dx.doi.org/10.1176/appi.ps.60.4.439>
47. Colin G. Walsh, Jessica D. Ribeiro, and Joseph C. Franklin. 2017. Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clin Psychol Sci* 5, 3 (Apr 2017), 457–469. DOI: <http://dx.doi.org/10.1177/2167702617691560>
48. David L. Williams, Victoria L. Crittenden, Teeda Keo, and Paulette McCarty. 2012. The use of social media: an exploratory study of usage among digital natives. *J Publ Aff* 12, 2 (May 2012), 127–136. DOI: <http://dx.doi.org/10.1002/pa.1414>
49. Haoyi Xiong, Jinghe Zhang, Yu Huang, Kevin Leach, and Laura E. Barnes. 2017. Daehr: A Discriminant Analysis Framework for Electronic Health Record Data and an Application to Early Detection of Mental Health Disorders. *ACM Trans. Intell. Syst. Technol.* 8, 3, Article 47 (Feb 2017), 21 pages. DOI: <http://dx.doi.org/10.1145/3007195>

50. Jinghe Zhang, Haoyi Xiong, Yu Huang, Hao Wu, Kevin Leach, and Laura E Barnes. 2015. MSEQ: Early Detection of Anxiety and Depression via Temporal Orders of Diagnoses in Electronic Health Data. In *2015*

*IEEE International Conference on Big Data (Big Data 2015)*. 2569–2577. DOI:  
<http://dx.doi.org/10.1109/BigData.2015.7364054>