Original Paper

Detection of Suicidality Among Opioid Users on Reddit: Machine Learning–Based Approach

Hannah Yao[1], MSc; Sina Rashidian[1], MSc; Xinyu Dong[1], MSc; Hongyi Duanmu[1], BSc; Richard N Rosenthal[2], MD; Fusheng Wang[1], PhD

[1]Stony Brook University, Stony Brook, NY, United States

[2]Renaissance School of Medicine at Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Fusheng Wang, PhD
Stony Brook University
2313D Computer Science
Stony Brook, NY, 11794
United States
Phone: 1 631 632 2594
Email: fusheng.wang@stonybrook.edu

Abstract

Background: In recent years, both suicide and overdose rates have been increasing. Many individuals who struggle with opioid use disorder are prone to suicidal ideation; this may often result in overdose. However, these fatal overdoses are difficult to classify as intentional or unintentional. Intentional overdose is difficult to detect, partially due to the lack of predictors and social stigmas that push individuals away from seeking help. These individuals may instead use web-based means to articulate their concerns.

Objective: This study aimed to extract posts of suicidality among opioid users on Reddit using machine learning methods. The performance of the models is derivative of the data purity, and the results will help us to better understand the rationale of these users, providing new insights into individuals who are part of the opioid epidemic.

Methods: Reddit posts between June 2017 and June 2018 were collected from *r/suicidewatch*, *r/depression*, a set of opioid-related subreddits, and a control subreddit set. We first classified suicidal versus nonsuicidal languages and then classified users with opioid usage versus those without opioid usage. Several traditional baselines and neural network (NN) text classifiers were trained using subreddit names as the labels and combinations of semantic inputs. We then attempted to extract out-of-sample data belonging to the intersection of suicide ideation and opioid abuse. Amazon Mechanical Turk was used to provide labels for the out-of-sample data.

Results: Classification results were at least 90% across all models for at least one combination of input; the best classifier was convolutional neural network, which obtained an F_1 score of 96.6%. When predicting out-of-sample data for posts containing both suicidal ideation and signs of opioid addiction, NN classifiers produced more false positives and traditional methods produced more false negatives, which is less desirable for predicting suicidal sentiments.

Conclusions: Opioid abuse is linked to the risk of unintentional overdose and suicide risk. Social media platforms such as Reddit contain metadata that can aid machine learning and provide information at a personal level that cannot be obtained elsewhere. We demonstrate that it is possible to use NNs as a tool to predict an out-of-sample target with a model built from data sets labeled by characteristics we wish to distinguish in the out-of-sample target.

KEYWORDS

opioid epidemic; opioid-related disorders; suicide; social media; machine learning; deep learning; natural language processing

Introduction

Background

The ongoing opioid crisis is characterized by an increasing number of deaths caused by opioid overdose; this number increased to such an extent that in 2017, the US Department of Health and Human Services declared a public health emergency [1]. In 2016, the US Centers for Disease Control and Prevention reported 42,000 deaths caused by opioid overdose, resulting from both prescribed and nonprescribed opioids and including intentional (suicide) and unintentional (accidental) deaths [2]. Since 2008, suicide has been the tenth leading cause of death in the United States. In 2016, suicide became the second leading cause of death for those aged 10 to 34 years [3]. The American Foundation for Suicide Prevention approximates 44,965 Americans die by suicide each year, and for each suicide, 25 people attempt suicide. These numbers may be underestimated because of the stigma associated with suicide [4].

Motivation

Individuals with chronic pain and poor mental health are often at risk of suicide; however, most current opioid overdose prevention methods neither assess suicide risk nor tailor prevention methods to personal situations [2,5,6]. Further hindering implementation of suicide screening is the lack of available data regarding death by overdose, frequently classified as *undetermined*, because of the inability to verify a death as a suicide. For example, if a deceased individual is found with a syringe by their side, it is difficult to determine whether it was an intentionally fatal dose or if the death was accidental (and without overdose intention) because of the risky route of administration by injection [6].

Individuals with opioid use disorder (OUD) are candidates for overdose due to the nature of the disorder. This study specifically focuses on suicidal ideation or intentions among individuals with a high possibility of OUD or, at minimum, opioid misuse. The Diagnostic and Statistical Manual of Mental Disorders defines OUD as a problematic pattern of opioid use leading to clinically significant impairment or distress [7]; approximately more than 2.5 million individuals are affected by OUD [8]. Diagnostic criteria include large amounts of opioids taken over a long period of time than intended, unsuccessful efforts to control opioid use, significant time investment in opioid-related activities or that necessary for recovery, impairment of daily life functionality, continued usage despite knowledge of consequences, dosage tolerance, and unavoidability of withdrawal [7].

The *New England Journal of Medicine* proposes that suicidal intention is blurred for individuals with OUD, who are repeatedly exposed to addiction-induced usage and withdrawal cycles and mood swings [2]. Ultimately, such prolonged exposure causes desensitization to risky and impulsive behaviors that can lead to death. Although the individual is aware of the consequences, the corresponding weight is muted and there is no associated conscious suicidal intent [9]. This aspect of OUD can lead to opioid overdose deaths but may not be explicitly considered in its contribution to overdose. Even in the absence of OUD, people who experience major depressive episodes,

suicidal ideation, or other use and mental disorders are likely to misuse opioids, suggesting that opioid abuse and psychopathological conditions are self-perpetuating and cyclic behaviors [10].

Relevant predictors of suicide risk include chronic pain combined with depression [11,12]. Individuals with a history of depression also have a higher average daily dose and are more insistent on long-term opioid treatment for chronic pain [12], which is characteristic of OUD. Chronic pain and depression may also lead to suicidal thoughts even in originally healthy individuals [13], and both passive and active suicidal ideation are associated with mental illness, previous suicide attempts, and chronic pain [11,13]. Some suggest that opioid abuse manifests largely as a maladaptive coping strategy with physical and mental pain and is not solely biological [14,15]. Then, even if there is no conscious suicide intent, poor mental health leads the opioid-worn individual to adopt the mindset "I would not mind dying," and they can become more reckless in their behavior, resulting in a higher probability of overdose. Given these correlations, it is not unreasonable to seriously consider suicide risk among opioid users.

Those with OUD who experience thwarted belongingness may not be willing to discuss their concerns with those in real life due to social stigma [16]. Instead, they may turn to web-based means of discussion, connecting with and seeking support from those who are more likely to understand them than health professionals [17]. Most opioid studies based on social media have been conducted on Twitter. However, Twitter is limited to 280 characters, which may not provide enough space for struggling individuals to express themselves thoroughly, and it can be highly noisy due to lack of monitoring and the medley of topics [18-22].

In contrast, Reddit is a forum-based social media platform for the discussion of many topics. Each topic has its own *subreddit* for topic-specific content, and the character limit for a text post is 40,000. Users can interact directly with an original post and with other users by commenting below the original submission. According to Reddit's statistics page, it is currently the fifth most visited site in the United States, with more than 330 million monthly active users and 138,000 active communities [23]. Individuals are increasingly using mental health subreddits as a means of web-based support when offline support is unavailable. As Reddit also provides anonymity that is not available in real life, there is increased self-disclosure around stigmatic topics [24-26]. Furthermore, as subreddits are category-specific, struggling individuals can receive support and understanding from those with shared experience, who can provide therapeutic factors in the comments (eg, instillation of hope and altruism for those in opioid recovery or withdrawal) [27].

We consider the interpersonal psychological theory of suicide behavior as motivation for examining the connection between suicide and OUD [28]. In brief, this theory states that an individual will not die by suicide unless they have both the desire and the capability to act on that desire. According to the theory, suicidal desire is derived from an individual's feeling of burdensomeness and an impeded sense of belongingness

(social alienation) accompanied by the capability for suicide, which develops from repeated exposure to pain and exasperation [9]. Thwarted belongingness manifests as introspective torture in the individual with OUD, who often have ruined relationships or feels permanently labeled as *an addict* by society and cannot find understanding from those around them. The capability for suicide is acquired through active engagement with suicidal-like behavior and requires numbing out the innate sense of self-preservation [28]. This is applicable to those with OUD, many of whom have experienced overdose first hand or have witnessed their friends passing away from overdose and thus become desensitized over time.

Objectives

To the best of our knowledge, the intersection of suicidality and OUD has not been analyzed in a data-driven fashion with machine learning methods, although both areas have been examined separately and research has been conducted clinically. We seek to use subreddits in combination with each other and machine learning methods to classify for in-sample data then predict for out-of-sample data to detect suicide ideation in the context of OUD.

In this study, we (1) utilized the structure and nature of subreddits to train classifiers with the goal of extracting posts containing both OUD and suicidal ideation, (2) compared inputs and models for classification, and (3) aimed to extract suicide risk posts from an opiate context and extract opioid addiction from a suicidal context, where *context* refers to the subreddit(s) to which the data set belongs. We also asked nonspecialist workers from Amazon Mechanical Turk (MTURK) to annotate a sample of posts to obtain labels for whether a post is indicative of both suicide ideation and opioid addiction.

Related Work

Social media studies utilizing Reddit for its ability to provide massive amounts of text data have become popular. Many of these focus on current mental health concerns, such as depression and anxiety, and seek to analyze the linguistic features of post content and employ traditional methods such as N-grams, latent Dirichlet allocation topic modeling, linguistic inquiry and word count, term frequency-inverse document frequency (TF-IDF), and word embeddings to extract and analyze the emotional and mental states of Reddit users [25,27,29-31].

The subreddits r/suicidewatch and r/opiates revolve around the topics of suicidality (ie, discussions about, circumstance disclosure, emotional expressions, etc) and opioid usage (ie, drug doses, activities done while high, withdrawal anguish, etc), respectively. Analyses of specific subreddits have provided insights into the general mentality of subreddit users through their use of language. For example, a study by Park et al [29] showed that there are significant structure differences between sentences in mental health subreddits and nonmental health subreddits. They revealed differences between Reddit's r/suicidewatch and the general mental health subreddit in their readability indexes and showed increased usage of first-person singular pronouns and decreased usage of first-person plural, second-person, and third-person pronouns [29]. Kumar et al

[32] conducted a survey of r/suicidewatch to test for the Werther Effect, specifically following high-profile suicides, and found increased negative emotions, anger, self-harm, and suicidal words. They similarly reported increased usage of first-person singular pronouns along with decreased lexical diversity, longer posts, and fewer future-oriented words and more past-oriented words [32]. Yet another analysis of r/suicidewatch extracted recurring topics of discussion and then compared the results with suicide risk factors defined by domain experts; it was found that r/suicidewatch is capable of capturing less well-known dimensions of the risk factors, such as concerns about failing school, gun ownership, drug abuse, financial concerns, deceased friends, and domestic abuse [33]. The analysis of r/opiates yields dominant topics and characteristics of discussion, particularly for opioid users, such as opioid withdrawal and opioid intake routines [34].

These findings from r/suicidewatch support the findings of analyses of real-life writings left behind by people who attempted or completed suicide. In particular, a study of word use in poetry by suicidal and nonsuicidal poets revealed shifts in first-person singular pronoun and communication words (eg, *talk*), and a linguistic study of suicide completers showed that these individuals differ in time orientation compared to control groups [35,36]. The idea of *alienation* from the interpersonal theory of suicide supports the findings of increased usage of first-person pronouns—that when one is unable to relate to society, the presence of first-person plural *we* and other words of interaction decreases [28,37].

Neural networks (NNs) have been shown to achieve excellent results in natural language processing. Unlike traditional methods, *learning text from scratch* at the character or wording level requires no knowledge of sentiment; high-level targets can be used as input [38,39]. This is especially valuable in this context because drug-related words are often slang, and expressions related to getting high or intoxicated may be falsely categorized as negative when they are not due to the inability of sentiment lexicons to learn domain specificity [40]. Furthermore, there is no need for exhaustive feature engineering because weights can be learned [41].

Kim Yoon [42] demonstrated the ability of a simple convolutional neural network (CNN) with 1 layer of convolution in sentence classification for multiple data sets, finding that the performance of a simple CNN is comparable with that of traditional methods [42]. Orabi et al [43] used CNN and recurrent neural networks (RNNs) to predict depression for Twitter data, showing that CNN performs better than RNN. Johnson and Zhang [44] successfully used word sequences to classify documents with CNN, whereas Kim and Orabi only classified short, sentence length texts [44]. Singh et al [45] compared NNs and found that CNN performed best when considering both speed and accuracy. Therefore, we used a simple CNN architecture for its speed and performance.

Methods

Data Collection

The social media platform used in this study is Reddit. We focused on the subreddits r/suicidewatch and r/depression and opioid-related forums; also, a control group was used to model healthy, nondisordered language. The subreddit r/suicidewatch acts as a *bottom net*, such that if a user expresses suicidality, they will commonly be directed to r/suicidewatch for help. The subreddit has policies that forbid trolling and discourage activism (ie, repeatedly posting hotlines) in favor of direct peer support, which gives this forum credibility [46]. Subreddit r/depression is similar in forbidding *empty encouragement* (ie, *don't worry, it gets better*). These characteristics exemplify the seriousness of these discussion boards and substantiate the reliability of the data. In contrast to these subreddits, however, r/jokes has few or no rules, only specifying that there are to be no personal attacks and stating that discussions must be lighthearted and civil; meanwhile, r/showerthoughts simply asks for original content. Posts in opioid-related forums were assumed to be indicative of individuals who struggle with opioid abuse and seek understanding on the web; our assumption is based on the fact that the act of posting about opioids on social media indicates that the presence of these drugs is sufficiently dominant in one's life. All opioid-related forums were selected if the forum was relatively active and if the focus drug of the forum was frequently discussed in r/opiates, the most active opioid subreddit. Data were collected from Reddit using the pushshift.io Python Application Programming Interface (API) [47] and Reddit's Python Reddit API Wrapper (PRAW) [48]. The IDs of submissions between June 2017 and June 2018 were obtained via pushshift.io and then passed to PRAW, which retrieved the actual submission. A submission consists of an ID, author, title, and body text/content.

We used posts from r/suicidewatch to represent suicidal language and a collection of opioid-related subreddits to model the language of opioid users (which is referred to in this study as r/opiates or *opiates data*). We assumed that all posts from r/suicidewatch were suicidal and that the language of r/depression would be most similar to that of r/suicidewatch [49]. Several control subreddits were used for language comparison based on a study by Shen et al [50] for detecting anxiety on Reddit, which implies that subreddits themselves can be used as labels. This set of diverse subreddits seeks to offset the impact of excessive first-person pronoun usage that is common when expressing negative emotions and subject-specific words [51,52]. We assumed that the majority of posts in these control subreddits were *mentally healthy*, meaning that post content (saying nothing about the user themselves) is not reflective of mental disturbances such as depression and suicidal ideation. In addition to the selected subreddits from those used by Shen et al [50], we included r/careerguidance and r/personalfinance to account for finance stressors that may be experienced by those in r/suicidewatch, and we included r/offmychest to account for the potentially rougher and more profane language of opioid users, which is indicative of strong negative emotions but unlikely to indicate suicidality. Examples of posts in some specific subreddits are shown in Table 1, and the data groups and their subreddits are listed as follows:

- *Depression*: depression.
- *Suicidewatch*: suicidewatch.
- *Control*: askdocs, askscience, books, careerguidance, fitness, frugal, jokes, lifeprotips, offmychest, parenting, personalfinance, productivity, randomkindess, relationships, showerthoughts, talesfrom retail, theoryofreddit, wholesometextposts, writing, youshouldknow.
- *Opiates*: benzodiazepines, benzorecovery, heroin, methadone, opiates, opiatesrecovery, quittingkratom, suboxone.

Table 1. Example posts from subreddits belonging to data groups.

Subreddit group	Title	Body text
Control (r/showerthoughts)	It becomes less and less acceptable to cry in public the older you get, despite the reasons for doing so becoming more and more valid	Kids just don't understand
Opioid related (r/opiates)	Happiness is…a big, dark shot of Heroin after being sick all week. <3	And also knowing you have enough for not only the next morning but at least a few more days while you figure out what to do next!
Depression	I won't commit suicide but I wouldn't mind dying	So much shit has been piling on and on. I feel like I am not making the people I care about proud and the only reason they talk to me is because of pity. I will not take my own life, but if a car hit me, I got terminal illness or if something fell on me. I would not be sad about me being gone
SuicideWatch	I'm waiting for the courage to end my own life	I feel like I'm close to making a serious attempt sooner than later and I'm ok with that. My impulsive behaviors have gotten worse other the last few months and some of those ways include bodily harm. In September i impulsively jumped out of a friends window and injured myself and now I am cutting myself at random for the first time in years. No one around me understands how exhausting it is to wake up everyday in my own skin. In my own head. I'm sick of the stomach pangs and guilt and crying and disappointment. Some nights I just pray I'll have the courage to end it. People die in the world all over, shouldn't matter when I go

Feature Matrix Construction

We maintained the high-level structure of the language, only lowercasing and removing URLs [42]. The text was not lemmatized, and all stop words and profanity were kept. All punctuation marks were removed except for periods, commas, exclamation marks, question marks, and apostrophes. We sought to make use of word order for text categorization [44]. As the title of a post may be a good representation of post content, the title of every post was prepended to the body text. The joined title and body content composed the NN input. All posts were either zero padded or truncated to a length of 1500 words, as we wanted to use the entire post if possible. Each post (title+body text) was preprocessed, tokenized, and encoded, preserving the word sequence, and is thus represented as a vector of words. We also reran the classification using only posts with more than 1000 words to show that differing content lengths do not impede performance. Textbox 1 shows examples at each text-processing stage.

Textbox 1. Example of text-processing stages.

Unprocessed text

- Question. What's the best otc med to OD on? I'm so over the day to day grind. I'm a failure at life and ready to check out...

Processed text

- Question. what's the best otc med to od on ? i'm so over the day to day grind . i'm a failure at life and ready to check out . . .

Tokenized

- | question | . | what's | the | best | otc | med | to | od | on | ? | i'm | so | over | the | day | to | day | grind | . | i'm | a | failure | at | life | and | ready | to | check | out | . | . | . |

As sequential text data cannot be used as is for training machine learning models, there are several techniques that can convert these data to numerical values:

1. TF-IDF: The term frequency-inverse document frequency shows how important a word is in a document with respect to its frequency in other documents. We used a 100-dimension TF-IDF from in-sample unigrams and bigrams [53].
2. Word embedding: This is a set of methods and language models for converting text data to vector space. We took advantage of 3 methods, including the Gensim Word2Vec model Global Vectors for Word Representations (GloVe), which require embeddings to be trained in advance on a large corpora of data and also a simpler version of word to ID mapping, which was performed in running time [54,55].
3. Character embedding: As out-of-dictionary words are all mapped to random vectors in pretrained word embeddings or the slightest change in slang language could end up mapping words to different vectors, adding character embedding as well showed improvements for training the machine learning models.

In this study, we took advantage of different combinations of these knowledge representation methods to evaluate whether classification performance improves additional information by alleviating vocabulary issues. Multimedia Appendix 1 shows the complete list of combinations.

Models

Our main focus was on a CNN text classifier. In addition, we implemented several traditional and deep learning models to serve as baselines for the CNN. Details are as follows.

Classic Baselines

We employed well-known traditional machine learning methods as baselines, including logistic regression (LR), random forest (RF), and support vector machines (SVMs). The implementation used Python language and was based on the scikit-learn toolkit [53].

NN Baselines

We implemented FastText (FAST), RNN, and attention-based bidirectional RNN (ATTENTION) in their original architectures for comparison [56-58]. FAST uses a shallow NN and word representations constructed from bag-of-n-grams, which are averaged into the text representation and input to a linear classifier. RNN considers historical information and shares weights across time by allowing previous outputs to be reused as inputs. ATTENTION builds on RNN by introducing the ability to *attend* to specific subinputs rather than all available information, thereby improving the decisions made at each step.
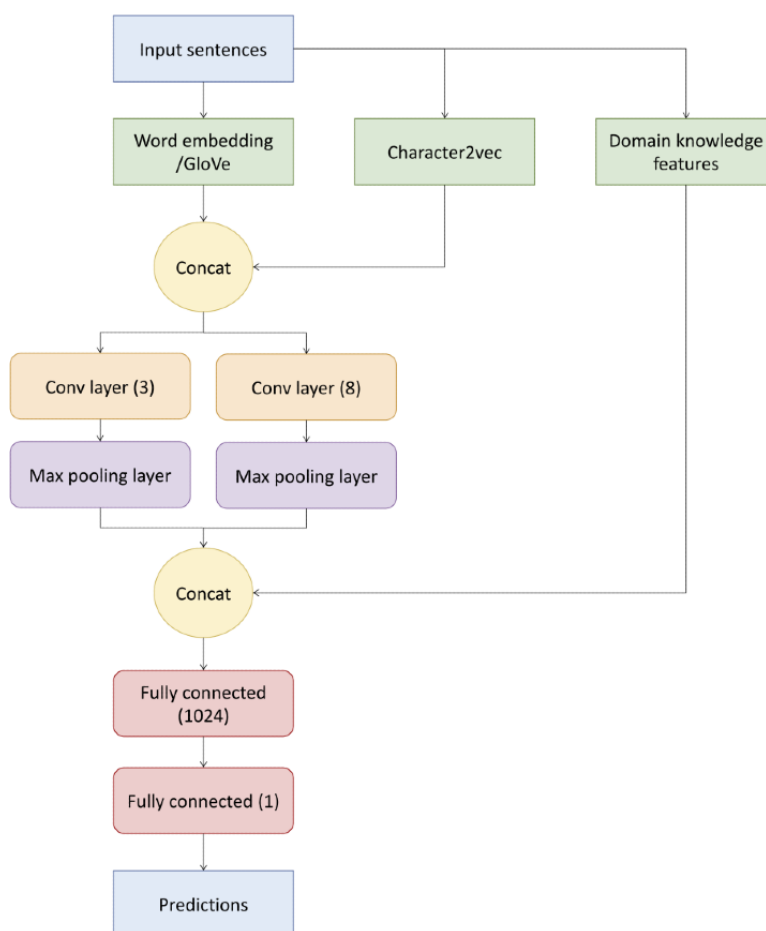
CNN Model

We implemented a CNN based on Kim Yoon's CNN architecture and used the Keras NN library in Python [59]. The major difference between this implementation and Kim's is that random initialization was used instead of word2vec because both perform equally, max pooling was used instead of global pooling, and filter sizes of 3 and 8 were used instead of 3, 4, and 5. A filter size of 8 was shown to work better with longer documents and would slide over a window size of 8 words at a time [60]. The activation layer following every convolutional layer was the rectified linear unit, $f(x)=\max(0,x)$, which maps nonnegatives between the input and feature map. In terms of integrating additional information, domain knowledge features and character2vec (char2vec), into the original CNN system, we simply used concatenation. As the char2vec feature matrix has the same size as the word embedding features and similar logical meaning, we concatenated them together before the main convolution part. Domain knowledge features, a large vector showing the information extracted from the linguistics perspective, were concatenated before the final fully connected layers part. We used a batch size of 512, epoch count of 6, and embedding layer that was learned at the training time. The

embedding layer had a dropout of 0.5 to prevent overfitting and a dimension size of 100. Loss was calculated using cross-entropy. Figure 1 shows our CNN architecture.

Figure 1. Overview of the convolutional neural network architecture.

Classification and Prediction

Our hypothesis is from an intuitive standpoint. We considered 2 cases: (1) there exist posts with suicide sentiments among the opiates data set and (2) there exist posts indicative of opioid addiction among the suicidewatch data set. We sought to extract posts that contain both suicidal sentiments and opioid abuse. These 2 cases are referred to as *C1* and *C2* and are represented by their respective models.

Case 1 (C1) is a model that will be trained on data from r/suicidewatch and the control subreddits for positive language, and it will learn to distinguish between suicidal and positive language. As r/opiates contain mixed sentiments, if we feed the out-of-sample data from r/opiates into C1, then C1 should be able to classify between suicide risk posts and nonsuicide risk posts. Posts from r/opiates that are classified as *suicidal* will then contain both suicidality and OUD aspects. In addition, we coarsely assessed the quality of posts predicted as *suicidal* by constructing a word2vec from predictions.

Case 2 (C2) is a model that is trained on data from r/depression and r/opiates. The subreddit r/depression is the most similar to r/suicidewatch which is available. C2 will learn to distinguish

between posts containing drug addiction and no drug addiction. If we feed out-of-sample data from r/suicidewatch into C2, then C2 should be able to classify between posts containing opioid usage and nonopioid usage. Posts from r/suicidewatch that are classified as having opioid usage will then contain both suicidality and OUD aspects.

For classification, the data sets were split with an 8:2 ratio for training and testing. The data for training were split again with an 8:2 ratio for train versus train validation. To evaluate the prediction ability, we sampled 500 posts ad hoc from r/opiates and 500 posts from r/suicidewatch to be annotated. To enforce a lower bound on sample quality and reduce variance among the posts, we only sampled from posts between 30 and 500 words. As class balance is a concern within a sample of 500, a maximum of 250 of 500 posts were selected if they contained a keyword or a key phrase of interest. Thus, the 8:2 train/test split over the entire data set was used for in-sample testing, and then 500 samples were extracted from the 20% of posts that belong to the test split for out-of-sample testing. Table 2 summarizes the C1 and C2 procedures and is illustrated in Multimedia Appendix 2, which demonstrates the workflow for C1.

Table 2. Summary of the 2 models for the 2 cases.

Case	C1	C2
Training and classification		
Goal	Distinguish between suicidal and non-suicidal language; predict for suicidality among opioid users	Distinguish between language of opioid usage and depressed but nonopioid using; predict for opioid usage among suicidal individuals
Data set	51,366 posts from r/suicidewatch and control subreddits	59,940 posts from opioid relevant subreddits and r/depression
Vocabulary size	70,082	64,078
Training	Trained and validated on data from r/suicidewatch and control subreddits	Trained and validated on data from opioid relevant subreddits and r/depression
Prediction on out-of-sample data		
Predicts on	Data from r/opiates	Data from r/suicidewatch
Prediction goal	Predicts for posts containing suicide risk in r/opiates	Predicts for posts containing opioid abuse in r/suicidewatch
Total posts prediction data that are between 30 and 500 words	23,740 posts from r/opiates	21,719 posts from r/suicidewatch
Sample max 250 from data for prediction containing these keywords for MTURK[a]	Commit suicide, suicide, suicidality, suicidal, want to die, want to do, want to overdose	Benzodiazepines, benzos, cocaine, codeine, fentanyl, heroin, hydrocodine, hydrocodone, hydromorphone, hydros, kratom, methadone, morphine, narcotic, narcotics, opiates, opioid, oxycodone, oxycontin, oxycottin, oxycotton, oxymorphone, suboxone
Sample count containing keywords	234	231

[a]MTURK: Amazon Mechanical Turk.

Annotations With Amazon Mechanical Turk

The sampled posts were annotated by crowdsourced workers from MTURK [61]. Each post was annotated by 3 workers, and the final label for a post was determined by the majority rule.

For C1, the MTURK task was titled "Suicidal sentiment detection." The description was "May the user be at risk of suicide or intentional overdose?" with 2 options: "Yes, risk of suicide" and "No risk of suicide." Task keywords included *depression, drug abuse, drugs, mental, mental health, opiates, overdose, suicide, suicide ideation, suicidewatch,* and text. Eligible MTURK workers were required to have masters qualifications and were awarded US $0.04 per annotation. We required no other qualifications. Assuming that the average individual lacks experience in both mental health and drug misuse [62], we defined common slang terms from r/opiates (eg, PAWS stands for postacute withdrawal syndrome) and provided simple instructions based on *Clues to Suicide* by Edwin Shneidman, who is the founder of the American Association of Suicidology and has laid groundwork in the field [63].

For C2, the MTURK task was titled "Opioid addiction among suicidal individuals." The description was "Does the post imply opioid addiction?" with 2 options: "Yes, implies opioid addiction" and "No opioid addiction." Task keywords and qualifications were the same as in the C1 MTURK tasks. We provided a list of opioid names in the instructions.

Results

Comparison of Models and the Impact of Feature Matrices

Metrics were calculated for accuracy, precision, recall, F_1 score, and area under the curve. The complete results can be found in Multimedia Appendix 1. Table 3 provides a subset of the results of comparing the CNN with the traditional baselines using the F_1 score in training the classifiers. Table 4 provides the results of comparing the NNs. At the end of this phrase, the models had been trained and were ready to be used for prediction of out-of-sample data.

XSL•FO

RenderX

Table 3. F_1 scores achieved for the given different combinations of input for classification.

Model	LR[a]	RF[b]	SVM[c]	CNN[d]
C1: r/suicidewatch versus positive control group				
TF-IDF[e]	0.902	0.904	0.915	0.685
word2vec	0.928	*0.927*[f]	*0.943*	0.961
TF-IDF + word2vec	*0.940*	0.921	0.941	*0.963*
TF-IDF + GloVe	0.927	0.829	0.886	0.923
TF-IDF + word2vec + char2vec	0.914	0.790	0.856	0.962
C2: r/depression versus r/opiates				
TF-IDF	0.889	0.800	0.811	0.729
word2vec	0.852	*0.846*	*0.881*	0.961
TF-IDF + word2vec	*0.894*	0.815	0.880	0.965
TF-IDF + GloVe	0.860	0.494	0.765	0.814
TF-IDF + word2vec + char2vec	0.858	0.581	0.741	*0.966*

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]CNN: convolutional neural network.

[e]TF-IDF: term frequency-inverse document frequency.

[f]The best results achieved by the model are in italics.

Table 4. Comparison of text classification neural network baselines with word embedding and convolutional neural network.

Model	FAST[a]	RNN[b]	ATTENTION[c]	CNN[d]
C1: r/suicidewatch versus positive control group				
Accuracy	0.950	0.944	0.939	*0.954*[e]
Precision	0.958	0.953	0.934	*0.968*
Recall	0.957	0.951	*0.965*	0.953
F_1 score	0.957	0.952	0.949	*0.961*
C2: r/depression versus r/opiates				
Accuracy	0.971	0.957	0.969	0.971
Precision	0.964	0.964	0.967	0.970
Recall	0.958	0.923	0.951	*0.962*
F_1 score	0.961	0.943	0.959	*0.966*

[a]FAST: FastText.

[b]RNN: recurrent neural network.

[c]ATTENTION: attention-based bidirectional recurrent neural network.

[d]CNN: convolutional neural network.

[e]The best score for each model is italicized.

From Table 3, we can observe that word2vec contributed to the CNN performance. Another interesting observation is that the classification performance of all models was affected when a pretrained GloVe was used as input. This is possibly due to the local information introduced by using specific subreddits as categories. The introduction of GloVe and character embeddings impeded the performance of RF, which implies that RF experiences more difficulty in learning implicit semantic features than LR and SVM. In contrast, NNs performed well with semantic-based inputs.

Figure 2 provides a visualization of the word embeddings that the NNs trained from scratch. Color intensity and shade from light purple to hot pink indicate increasing emphasis placed on the respective words. CNN, RNN, and ATTENTION share many important words, whereas FastText appears to focus on

XSL•FO

RenderX

different words given its bag-of-N-grams approach but nonetheless achieved high classification performance.

Figure 2. Visualization of word importance as determined from the resultant word embeddings learned from scratch by the neural network models for classification of suicidal vs non-suicidal text. CNN: convolutional neural network; LSTN: long short-term memory; RNN: recurrent neural network.

CNN

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

Fast Text

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

RNN

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

Attention bidirectional LSTM-RNN

been clean for over months and i'm about to throw it all away . i just can't take it anymore the loneliness , depression , anxiety , and suicidal thoughts . on the outside i look like i'm doing good and everyone keeps complimenting me . you're doing a great job username keep it up . and i keep responding that i'm not doing fine , that tears would be pouring down my face multiple times a day , if i could cry . but i can't cry anymore for some reason , no matter how sad i get . i've thought about suicide every day for the past month and a half . i didn't get sober to keep being miserable . i have no fucking friends anymore , no one i can vent to . i got into a rehab romance against all advice and now the girl is dating someone else . i invested so much emotion into her fucking stupid , i know and now i'll never be able to have a relationship with her . i don't know why i'm even writing this . i guess there's a small part of me that is still clinging to sobriety . i'm not going to get high tonight but i've already made plans to do so tomorrow . i'm on the shot so i won't be able to do dope for a while but i'm really obsessing about shooting coke again . i really don't want to go back out and put myself through all that bullshit again but i also can't stand feeling this way for one minute longer . .

Evaluating Predictions Using MTURK Annotations

For prediction, we used traditional baselines with word2vec input and NNs. To assess the prediction performance, MTURK was used to obtain labels with 3 annotations per post. C1 predicted for suicidality among opioid data. Annotations were performed by 25 unique workers, with an average of 4 minutes spent on each post. The pairwise agreement among the 3 annotations was 0.632, 0.628, and 0.644. In total, 120 out of 500 posts were ruled as yes, suicide risk. C2 predicted opioid addiction among suicidal posts. Annotations were performed by 92 unique workers, with an average of 11.3 minutes (C1: SD 6.667 minutes; C2: SD 14.633 minutes) spent on each post. The pairwise agreement among annotations was 0.524, 0.604,

and 0.544. In total, 261 out of 500 posts were ruled as *yes, implies opioid addiction*. Table 5 shows the model performance using the majority rule MTURK annotations as the ground truth. Table 5 also shows the differences in performance depending on the ratio of labels in the data set. To construct a set of data with 500 rows with a certain percentage of suicide risk posts, we sampled using scikit-learn with replace as true.

Table 5. Count and accuracy for model predictions using Amazon Mechanical Turk labels.

Model	LR[a]	RF[b]	SVM[c]	FAST[d]	RNN[e]	ATTENTION[f]	CNN[g]
Count and accuracy for C1 (predicts for suicide risk among opiates data)							
Predicted number of suicide risk	24	12	11	97	93	98	103
All data	0.768	0.750	0.744	0.59	0.608	0.576	0.536
Suicide risk only	0.2	0.1	0.092	0.783	0.791	0.75	0.833
Nonsuicide risk only	0.947	0.959	0.963	0.529	0.55	0.521	0.432
Count and accuracy for C2 (predicts for opioid addiction among suicidewatch data)							
Predicted number of opioid addiction	88	92	105	92	158	110	127
All data	0.524	0.518	0.538	0.54	0.588	0.552	0.562
Opioid addiction	0.230	0.237	0.273	0.251	0.414	0.295	0.334
Nonopioid addiction	0.892	0.869	0.869	0.9	0.806	0.874	0.847

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]FAST: FastText.

[e]RNN: recurrent neural network.

[f]ATTENTION: attention-based bidirectional recurrent neural network.

[g]CNN: convolutional neural network.

Observations Regarding C1: Detecting Suicide Ideation in the Context of Opioid Addiction

For both C1 and C2, we noticed that the model metrics changed drastically when the ratio of posts per label was altered. For C1, traditional baselines performed better when there was a mixture of suicidal and nonsuicidal-labeled posts because NNs produce a high number of false positives. However, when only examining suicidal posts, NNs achieved better accuracy. Thus, depending on the domain and data set, it may be better to use models that are superior at the extremes or remain within a relatively less variable range. For our case of predicting suicide risk posts among the out-of-sample opiate data, it is better to err on the safe side than to miss someone who is at risk of suicide. CNN would be a good choice if the sample consisted of many suicide risk posts. The predictive power of the baselines would not be desirable when predicting for posts indicative of suicide risk, as false negatives would be costly in the context of real life.

Given that the models tended to predict at the extremes for C1, we speculated that the models would serve well as weak learners to an ensemble model. We explored a simple ensemble model using scikit-learn's AdaBoost Classifier and LR, SVM, RNN, and CNN as weak learners. We used MTURK-labeled data of size 500 as the ground truth and the weak learner's probabilistic predictions as the input data. The data were split into training and testing sets at an 8:2 ratio. AdaBoost was run 5 times with different data shuffles and random states, and the average accuracy was taken.

In Figure 3, the top graph illustrates the performance of selected models as we increased the percentage of suicide risk labels from opiate data in C1 from 10% to 90%. The line plot shows the performance of AdaBoost with those models as weak learners. AdaBoost achieved respective accuracies of 0.89, 0.84, 0.82, 0.8, 0.8, 0.86, 0.9, 0.91, and 0.94 for the respective intervals. The bottom graph in Figure 3 displays the model contribution at each step. From the graph, combining the models that predict at the extremes results in a more robust classifier. This is an advantageous approach because the ensemble model now has knowledge about the out-of-sample data, which the weak learner models did not have.

XSL•FO

RenderX

Figure 3. The top bar graph shows changes in prediction performance in C1 depending on data category ratios. The overlaying line plot shows accuracy achieved from using the four models as weak learners in AdaBoost. The bottom graph shows weak learner contribution.
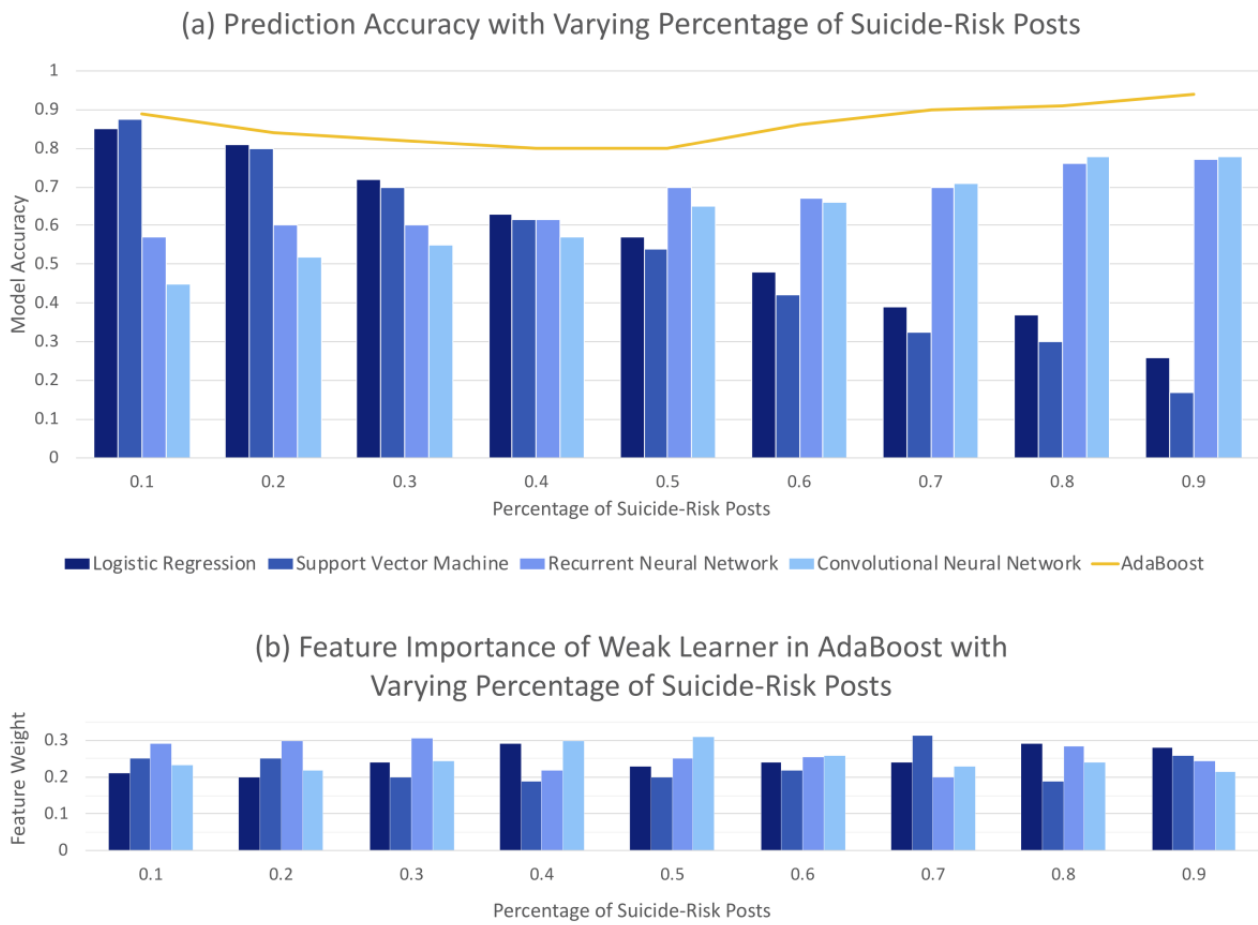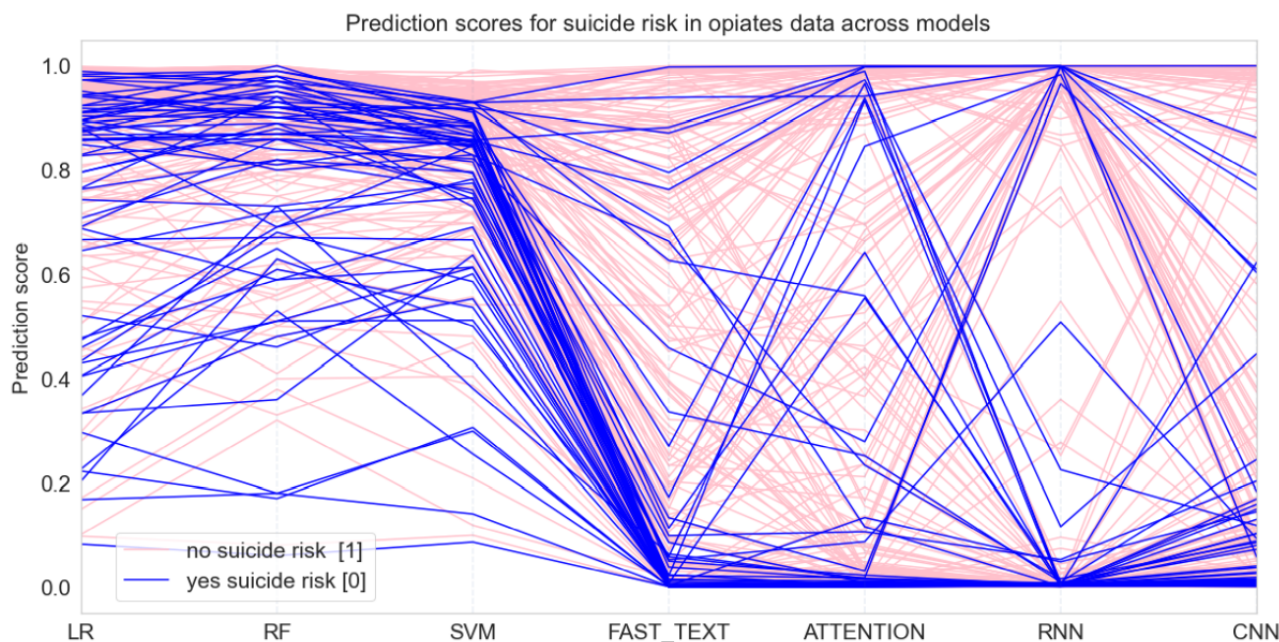

(a) Prediction Accuracy with Varying Percentage of Suicide-Risk Posts

■ Logistic Regression ■ Support Vector Machine ■ Recurrent Neural Network ■ Convolutional Neural Network — AdaBoost

(b) Feature Importance of Weak Learner in AdaBoost with Varying Percentage of Suicide-Risk Posts

Figure 4 shows the parallel prediction performance for C1 at a more fine-grain level; 50% of each label was sampled for the visual. Each line represents a post. The categorical *x*-axis lists the models, and the intersection of a line with the *y*-axis of a model is the prediction score of the model. The darker lines are MTURK-labeled as indicative of suicide risk (0 on the y-axis), whereas the lighter lines are nonindicative of suicide risk (1 on the y-axis).

Figure 4. Model predictions for individual posts represented as a line. A prediction closer to 1 indicates no predicted suicidal language and a prediction closer to 0 indicates yes predicted suicidal language. Individual posts are color coded by their MTURK label. ATTENTION: attention-based bidirectional recurrent neural network; CNN: convolutional neural network; FAST_TEXT: FastText; LR: logistic regression: RF: random forest; RNN: recurrent neural network; SVM: support vector machine.

Observations Regarding C2: Detecting Opioid Abuse in the Context of Suicide Ideation

Our hypothesis for C2 was that after being trained on data from r/depression versus the opiates group, the models would be capable of detecting phrases used by individuals who struggle with drug addiction. Although the models all achieved good results in classifying r/depression versus opiate data, they failed to extract opioid-containing posts with respect to the MTURK labels. One possible explanation is that users express themselves differently in different contexts (subreddits). Another possibility is that the language semantics in r/suicidewatch forgo characteristics found in opiates, such as details about withdrawal or relapse that take up the majority of space in any opiate context post. For example, a post in r/suicidewatch may briefly mention a few keywords such as "hydros" or "benzos" in a sentence about overdosing, while the remainder of the post focuses on personal struggles. On the other hand, if a user from r/opiates posts about wanting to commit suicide in r/opiatesrecovery, they will express sentiments regarding the drug because they

know their audience will understand. It is possible that the user is subconsciously assuming that the audience in r/suicidewatch knows little about opioid addiction and therefore omits the consequences of opioid addiction on their life. As a result, an MTURK worker who reads the entire post may observe that the user has easy access to opioids and is experienced with drug mixing; however, the model may fail to pick up such meanings because these exact expressions did not appear in the data it was trained on.

From our new hypothesis regarding language usage in C2, we re-evaluated the C2 predictions. Instead of using MTURK-determined labels, we constructed a set of heuristic labels such that a post is implicative of opioid addiction if the post contains any word from the C2 keywords in Table 2 and is not implicative of opioid addiction otherwise. The findings are shown in Table 6. Although the overall performance was not ideal, all scores increased, supporting our hypothesis that the presence of specific words is important and that MTURK workers' directed attention likely differs from that of the models.

Table 6. Model prediction accuracies with heuristic labels determined by the keyword presence.

Model	LR[a]	RF[b]	SVM[c]	FAST[d]	RNN[e]	ATTENTION[f]	CNN[g]
All data	0.606	0.588	0.648	0.666	0.774	0.69	0.712
Opioid addiction	0.264	0.26	0.347	0.338	0.56	0.403	0.463
Nonopioid addiction	0.9	0.87	0.907	0.948	0.926	0.937	0.926

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]FAST: FastText.

[e]RNN: recurrent neural network.

[f]ATTENTION: attention-based bidirectional recurrent neural network.

[g]CNN: convolutional neural network.

Content Overview for Predicted Suicidal Posts in Opiate Data

We constructed a bigram word2vec from posts predicted as *suicidal* in the entire opiate data set (opiate data: n=23,740, as shown in Table 3) to holistically assess the quality of these posts. We then queried for the top 15 most similar words to "suicidal" for each model, as it is more likely to appear with first-person pronouns as opposed to other derivations of the root "suicid." Table 7 summarizes the results from the most similar to least similar words.

Table 7. Top words most similar to suicidal from the subset of opiates data that was predicted to belong to the category suicidal.

Model	Top words
LR[a]	hour, dead, lack, asleep, wake up, sobriety, self, they are, cause, at least, however, hell, later, group, state
RF[b]	easy, living, yet, hit, waiting, probably, cold, the same, such, by, tomorrow, body, constantly, saying, working
SVM[c]	making, everyone, once, pills, without, soon, lol, nothing, around, sorry, thing, withdrawal, start, mental, tolerance
FAST[d]	suicidal thoughts, depressive, extreme, emotional, severe anxiety, existing, depressed, irritable, insomnia, severe, nauseous, mood swings, paranoid, fatigued, having trouble
RNN[e]	severely, depressed, diagnosed with, isolated, unbearable, bipolar, suicidal thoughts, anxious, ptsd, an alcoholic, ocd, overwhelmed, irritable, lethargic, extreme
ATTENTION[f]	diagnosed with, depressive, suicidal thoughts, social anxiety, bipolar, extreme, crippling, irritable, tremors, emotional, borderline, severe depression, brain fog, existing, gad
CNN[g]	paranoid, unhappy, depressed, isolated, apathetic, irritable, an alcoholic, suicidal thoughts, trauma, severely, diagnosed with, brain fog, anxious, manic, emotionless

[a]LR: logistic regression.

[b]RF: random forest.

[c]SVM: support vector machine.

[d]FAST: FastText.

[e]RNN: recurrent neural network.

[f]ATTENTION: attention-based bidirectional recurrent neural network.

[g]CNN: convolutional neural network.

Discussion

Overview

The goal of this study is to make use of the structure and data from Reddit to build models that could ultimately predict (C1) suicidal language among opioid users and (C2) opioid usage among individuals with suicidal thinking. We evaluated several models with combinations of semantic inputs and well-known NN text classifiers in their ability to classify and then predict out-of-sample data to extract posts containing both suicide ideation and opioid abuse. MTURK was then employed to provide heuristic ground truths to the out-of-sample data.

Limitations

Owing to the large quantity of data, we assumed that all posts under a subreddit were reflective of the official subreddit purpose (eg, recovering opioid users post in r/opiatesrecovery) and that deviations were neutralized because of the sheer volume of the majority content. A manual review of predicted suicidality among opioid subreddits (C1) found that many cases were wrongly predicted as suicidal because the users were

experiencing extreme withdrawal, anger due to relapse, and even anger due to sobriety because of the inability to use opioids as a coping mechanism. Inclusion of r/offmychest in the control subreddits was an attempt to offset such strong negative emotions that likely do not imply suicidality. A suicidal post may also be predicted as nonsuicidal because of calm resignation language or because the user is writing in self-reflection or to share past experiences.

A manual review of predicted opioid usage in r/suicidewatch (C2) found that poor predictions may have resulted from the sentiment mixture present in the opioid reddits. As the opioid group comprises active users and users in recovery/withdrawal (ie, r/opiates vs r/opiatesrecovery), the language variation could be extreme and vary from anger (ie, *My dealer is ****ing late*) to negative (ie, *I can't find happiness outside of heroin*) to joy (ie, *I've been clean for 70 days!*). All of these posts were given the same label of *opiates* as they were grouped under opioid language. Hence, the presence of drug-specific words is needed to predict opioid usage.

MTURK Limitations

We attempted to overcome MTURK limitations by providing brief context and instructions. Regardless, limitations included lack of domain knowledge, experience with suicide ideation or drug misuse, and subjectivity of interpretation. There is also possible carelessness in annotation, as some posts were relatively long and not all posts may have been read thoroughly. Furthermore, we provided a single post out of context from the user's life; it is difficult to determine how alarming the post is in reality because we cannot know what happened to the user later in time. There was also no way to control how many and which users participated in the assessment—the participation of many distinct workers may have introduced random variations in annotation consistency. Manual review of several posts that were disagreed upon showed that lack of domain expertise can make annotation difficult; for example, a post that was disagreed upon mentioned being *tired of waking up*, in *withdrawal*, and that it is *easier to end it all*, but also mentioned *wanting to get sober*. One MTURK worker may have considered this post as hopeful, whereas another worker may have considered the post to be on the edge of giving up. Despite these limitations, these annotations provide insight into the public perception regarding what is considered suicidal when opiate addiction is involved.

Insights

Studies conducted on opiates and other drug-addicted individuals have shown that opiate addiction is often comorbid with other psychological disturbances, such as personality disorders or trauma [64,65]. We observed this in the simple overview of words most related to *suicidal*. Among the NNs, several personality disorder terms and mental illnesses were extracted, such as *ptsd* (posttraumatic stress disorder), *borderline, bipolar*, and *gad* (general anxiety disorder). Characteristics of withdrawal were also captured: *irritable, insomnia, nauseous*. We can speculate that individuals with personality disorders going into withdrawal phrases are at higher risk of attempting suicide than other subgroups of the population.

Future Directions

There are many possible directions for utilizing Reddit metadata. For subreddits with mixed emotional sentiments such as r/opiates, coarse grain classification can first be applied to discard either too positive or negative posts depending on context. Given the interactivity of Reddit, analyses of comments may show what sort of support is provided by peers who struggle. The peer aspect of Reddit is attractive, and integration of nonjudgmental peer support may help recover OUD users to be more willing to rely on real-life aid. This is one step toward decreasing the frequent social isolation brought upon by OUD.

In general, exploration of Reddit data limits provides many possible study areas, given the amount of active subreddits. Among all posts in r/opiates, we found a handful in which a user who has successfully quit opioid use for many years and returns to relay hope for a full recovery. Such individuals can be seen as *success stories.* It is of particular interest to examine the difference between these individuals and those who end in tragedy, as OUD recovery is difficult and relapse prone [66]. One could also consider the interaction among suicide ideation, OUD, and chronic pain, especially because many OUD cases begin with an individual's or close person's prescription medication [10]. Another focus is the possibility of influencing the OUD individual's behavior; a study in the context of behaviorally targeted advertising for health care based on user searches implies that behavior, as determined by input search terms, is predictive of the individual being influenced by the target advertisement to act, such as visiting a health provider [67]. Therefore, internet behavior may be more closely linked to real-life behavior than expected, and targeted advertising might encourage OUD individuals who are midwithdrawal and fighting relapse to seek help.

The tendency of NNs to achieve high accuracy among purely suicide-labeled posts but perform poorly when the data set is mixed prompts the question of which semantic dimensions are actually being captured. Although we use NNs as a black box in this experiment, future studies may want to focus directly on latent semantic dimensions that may aid in parameter tuning and generalizability. In particular, it would be very interesting to extract the exact features that might generalize across textual data domains by examining the word embeddings that the NNs had trained from scratch to aid in predictions on out-of-sample data [68].

We focused on extracting posts that are indicative of suicide ideation in the context of OUD and did not perform much analysis on extracted posts. Many studies have investigated the content of suicidal individuals, but much fewer studies have been conducted in specific contexts, possibly due to the lack of available data. Gathered predictions are capable of revealing thoughts of suicidal opioid users, despite the limitations of this study. In aiding those recovering from OUD, it may be important to consider beginning at the personal level [69]. Communication between the professional and the patient is important, and personal stories are capable of decreasing stigma by deconstructing stigmatic barriers [70]. As these posts are likely to be raw and honest, they can allow clinical professionals to become more familiar with the mindsets of these individuals.

A less stigmatic perception of patients can lead to patient empowerment, which is necessary in the long term [66,71]. Consider a post that was agreed upon as showing the risk of suicide, summarized quote:

> *as soon as something goes missing, or something goes wrong, you're the first fucking person everyone suspects is at fault…it MUST have been the fucking drug addict, right? … I may be a dopehead, but I have never been a thief.*

It is easy to advocate understanding and help for those struggling with opioid addiction; however, it is harder to recognize how their own actions may affect those with opioid addiction. We believe that concrete case examples are essential as a first step and that suicide risk assessment should be done on an individual level. Our study seeks to aid in suicide prevention by helping those with OUD gain understanding—our goal is not to substitute for risk assessment using prediction models [72]. Application of gathered insights from the extracted posts can help decrease stigma and clarify wrong assumptions made about drug misusage and suicidal ideation.

Conclusions

The goal of this paper is to make use of big data from Reddit to detect suicidality among opioid users. The structure of Reddit offers categories for data sets, and the social media setting can provide case specifics that may not be obtainable elsewhere. This study can serve as a proof of concept for the use of social media site attributes to aid in machine learning methods and for research directions on the feasibility of NNs to abstract textual data and perform in independent areas. A comparison with a high-performing baseline model implies that the absence of hard-coded features may allow more flexibility and accuracy for models when running on out-of-sample data.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Classification metrics using combinations of different model inputs.
[XLSX File (Microsoft Excel File), 131 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Illustration of workflow for C1.
[PNG File , 95 KB-Multimedia Appendix 2]

References

1. What is the US Opioid Epidemic? The US Department of Health and Human Services (HHS). URL: https://www.hhs.gov/opioids/about-the-epidemic/index.html [accessed 2020-11-10]
2. Oquendo MA, Volkow ND. Suicide: a silent contributor to opioid-overdose deaths. N Engl J Med 2018 Apr 26;378(17):1567-1569. [doi: 10.1056/NEJMp1801417] [Medline: 29694805]
3. Hedegaard H, Curtin S, Warner M. Suicide Mortality in the United States, 1999-2017. NCHS Data Brief. 2018. URL: https://www.cdc.gov/nchs/data/databriefs/db330-h.pdf [accessed 2020-11-10]
4. Suicide Statistics. American Foundation for Suicide Prevention (AFSP). 2018. URL: https://afsp.org/suicide-statistics/ [accessed 2020-11-10]
5. Norton A. Many Opioid Overdoses May Be Suicides. WebMD. 2018 Apr 25. URL: https://www.webmd.com/mental-health/addiction/news/20180425/many-opioid-overdoses-may-be-suicides [accessed 2020-11-10]
6. Madadi P, Persaud N. Suicide by means of opioid overdose in patients with chronic pain. Curr Pain Headache Rep 2014 Nov;18(11):460. [doi: 10.1007/s11916-014-0460-1] [Medline: 25249422]
7. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Arlington, VA: American Psychiatric Association; 2013.
8. Effective Treatments for Opioid Addiction. National Institute on Drug Abuse (NIDA). 2016. URL: https://www.drugabuse.gov/publications/effective-treatments-opioid-addiction/effective-treatments-opioid-addiction [accessed 2020-11-10]
9. van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev 2010 Apr;117(2):575-600 [FREE Full text] [doi: 10.1037/a0018697] [Medline: 20438238]
10. Han B, Compton WM, Blanco C, Crane E, Lee J, Jones CM. Prescription opioid use, misuse, and use disorders in US adults: 2015 national survey on drug use and health. Ann Intern Med 2017 Sep 05;167(5):293-301. [doi: 10.7326/M17-0865] [Medline: 28761945]

11. Madadi P, Hildebrandt D, Lauwers AE, Koren G. Characteristics of opioid-users whose death was related to opioid-toxicity: a population-based study in Ontario, Canada. PLoS One 2013;8(4):e60600 [FREE Full text] [doi: 10.1371/journal.pone.0060600] [Medline: 23577131]

12. Braden JB, Sullivan MD, Ray GT, Saunders K, Merrill J, Silverberg MJ, et al. Trends in long-term opioid therapy for noncancer pain among persons with a history of depression. Gen Hosp Psychiatry 2009;31(6):564-570 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.07.003] [Medline: 19892215]

13. Hooley JM, Franklin JC, Nock MK. Chronic pain and suicide: understanding the association. Curr Pain Headache Rep 2014;18(8):435. [doi: 10.1007/s11916-014-0435-2] [Medline: 24916035]

14. Hendy HM, Black P, Can SH, Fleischut A, Aksen D. Opioid abuse as maladaptive coping to life stressors in US adults. J Drug Issues 2018 Jul 5;48(4):560-571. [doi: 10.1177/0022042618783454]

15. Sinha R. Chronic stress, drug use, and vulnerability to addiction. Ann N Y Acad Sci 2008 Oct;1141:105-130 [FREE Full text] [doi: 10.1196/annals.1441.030] [Medline: 18991954]

16. Santos DS, Andrade DT, Wan H, Ronzani T, Corrigan P. The Stigmatization of Drug Use as Mechanism of Legitimation of Exclusion. Switzerland: Springer; 2018:15-25.

17. Townsend L, Gearing RE, Polyanskaya O. Influence of health beliefs and stigma on choosing internet support groups over formal mental health services. Psychiatr Serv 2012 Apr;63(4):370-376. [doi: 10.1176/appi.ps.201100196] [Medline: 22337006]

18. Mackey TK, Kalyanam J, Katsuki T, Lanckriet G. Twitter-based detection of illegal online sale of prescription opioid. Am J Public Health 2017 Dec;107(12):1910-1915. [doi: 10.2105/AJPH.2017.303994] [Medline: 29048960]

19. Kalyanam J, Katsuki T, Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav 2017 Feb;65:289-295. [doi: 10.1016/j.addbeh.2016.08.019] [Medline: 27568339]

20. Chan B, Lopez A, Sarkar U. The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. PLoS One 2015;10(8):e0135072 [FREE Full text] [doi: 10.1371/journal.pone.0135072] [Medline: 26252774]

21. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]

22. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related Twitter posts. Crisis 2017 Sep;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]

23. Reddit by the Numbers. URL: https://www.redditinc.com/press [accessed 2020-11-10]

24. Johnson GJ, Ambrose PJ. Neo-tribes. Commun ACM 2006 Jan 1;49(1):107-113. [doi: 10.1145/1107458.1107463]

25. Choudhury MD, De S. Mental Health Discourse on Reddit: Self-disclosure, Social Support, and Anonymity. In: The 8th International Conference on Weblogs and Social Media. 2014 Presented at: WSM'14; August 5-8, 2014; Palo Alto, California URL: https://www.semanticscholar.org/paper/Mental-Health-Discourse-on-reddit%3A-Self-Disclosure%2C-Choudhury-De/e58943b8d656fbb07a14f36b28ae211da74992a9

26. DeAndrea DC, Anthony JC. Online peer support for mental health problems in the United States: 2004-2010. Psychol Med 2013 Nov;43(11):2277-2288 [FREE Full text] [doi: 10.1017/S0033291713000172] [Medline: 23410539]

27. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Escobar Lee K, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: a study of community interactions. Drug Alcohol Depend 2017 Dec 1;181:5-10 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.09.010] [Medline: 29024875]

28. Ribeiro JD, Joiner TE. The interpersonal-psychological theory of suicidal behavior: current status and future directions. J Clin Psychol 2009 Dec;65(12):1291-1299. [doi: 10.1002/jclp.20621] [Medline: 19827114]

29. Park A, Conway M. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]

30. Ji S, Yu CP, Fung S, Pan S, Long G. Supervised learning for suicidal ideation detection in online user content. Complexity 2018 Sep 9;2018:1-10. [doi: 10.1155/2018/6157249]

31. Sharma E, Choudhury M. Mental Health Supportits Relationship to Linguistic Accommodation in Online Communities. In: The CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI'18; April 14-16, 2018; Montreal QC, Canada. [doi: 10.1145/3173574.3174215]

32. Kumar M, Dredze M, Coppersmith G, De CM. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. In: The ACM Conference on Hypertext and Social Media. 2015 Presented at: ACM'15; September 20-22, 2015; Guzelyurt, Northern Cyprus URL: http://europepmc.org/abstract/MED/28713876 [doi: 10.1145/2700171.2791026]

33. Grant RN, Kucher D, León AM, Gemmell JF, Raicu DS, Fodeh SJ. Automatic extraction of informal topics from online suicidal ideation. BMC Bioinformatics 2018 Jun 13;19(Suppl 8):211 [FREE Full text] [doi: 10.1186/s12859-018-2197-z] [Medline: 29897319]

34. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz JH, et al. Social media based analysis of opioid epidemic using reddit. AMIA Annu Symp Proc 2018;2018:867-876 [FREE Full text] [Medline: 30815129]

35. Stirman SW, Pennebaker JW. Word use in the poetry of suicidal and nonsuicidal poets. Psychosom Med 2001;63(4):517-522. [doi: 10.1097/00006842-200107000-00001] [Medline: 11485104]

36. Handelman LD, Lester D. The content of suicide notes from attempters and completers. Crisis 2007;28(2):102-104. [doi: 10.1027/0227-5910.28.2.102] [Medline: 17722692]

37. Litvinova TA, Seredin PV, Litvinova OA, Romanchenko OV. Identification of suicidal tendencies of individuals based on the quantitative analysis of their internet texts. CyS 2017 Jun 30;21(2):243-252. [doi: 10.13053/cys-21-2-2721]

38. Zhang X, LeCun Y. Text understanding from scratch. arXiv Preprint posted online April 4, 2016 [FREE Full text]

39. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural language processing (almost) from scratch. J Mach Learn Res 2011;12:2493-2537 [FREE Full text]

40. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. 'When 'bad' is 'good'': identifying personal communication and sentiment in drug-related Tweets. JMIR Public Health Surveill 2016 Oct 24;2(2):e162 [FREE Full text] [doi: 10.2196/publichealth.6327] [Medline: 27777215]

41. Husseini OA, Buddhitha P, Husseini OM, Inkpen D. Deep Learning for Depression Detection of Twitter Users. In: The Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: CLCP'18; June 5-7, 2018; New Orleans, LA p. 88-97. [doi: 10.18653/v1/w18-0609]

42. Kim Y. Convolutional Neural Networks for Sentence Classification. In: The 2014 Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: NLP'14; October 23-26, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1181]

43. Husseini OA, Buddhitha P, Husseini OM, Inkpen D. Deep Learning for Depression Detection of Twitter Users. In: The Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: CLCP'18; June 2-7, 2018; New Orleans, LA p. 88-97. [doi: 10.18653/v1/w18-0609]

44. Johnson R, Zhang T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. In: The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015 Presented at: HLT'15; August 5-8, 2015; Denver, Colorado p. 103-112. [doi: 10.3115/v1/n15-1011]

45. Singh D, Wang A. Detecting Depression Through Tweets. Stanford University. 2016. URL: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/reports/6879557.pdf [accessed 2020-11-10]

46. Suicidewatch: Peer Support for Anyone Struggling With Suicidal Thoughts. Reddit. URL: https://www.reddit.com/r/SuicideWatch/ [accessed 2020-11-10]

47. Baumgartner J. Pushshift Reddit API. GitHub. 2017. URL: https://github.com/pushshift/api [accessed 2020-11-10]

48. Mellor T, Boe B. PRAW: The Python Reddit API Wrapper. GitHub. URL: https://github.com/praw-dev [accessed 2020-11-10]

49. Aladağ AE, Muderrisoglu S, Akbas NB, Zahmacioglu O, Bingol HO. Detecting Suicidal Ideation on Forums: Proof-of-Concept Study. J Med Internet Res 2018 Jun 21;20(6):e215 [FREE Full text] [doi: 10.2196/jmir.9840] [Medline: 29929945]

50. Shen J, Rudzicz F. Detecting Anxiety through Reddit. In: The Fourth Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality. 2017 Presented at: CLCP'17; August 3-8, 2017; Vancouver, BC. [doi: 10.18653/v1/w17-3107]

51. Shen J, Rudzicz F. Detecting Anxiety through Reddit. In: The Fourth Workshop on Computational Linguistics and Clinical Psychology - From Linguistic Signal to Clinical Reality. 2017 Presented at: CLCP'17; August 5-8, 2017; Vancouver, BC. [doi: 10.18653/v1/w17-3107]

52. Nerbonne J. The secret life of pronouns. What our words say about us. Literary Linguistic Comput 2013 Feb 6;29(1):139-142. [doi: 10.1093/llc/fqt006]

53. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830. [doi: 10.1016/j.patcog.2011.04.006]

54. Rehurek R, Sojka P. Software Framework for Topic Modelling With Large Corpora. In: The LREC Workshop on New Challenges for NLP Frameworks. 2010 Presented at: NLP'10; April 2-9, 2010; Valletta, Malta.

55. Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation. In: The Conference on Empirical Methods in Natural Language Processing. 2014 Presented at: EMNLP'14; October 25-29, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]

56. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of Tricks for Efficient Text Classification. In: The 15th Conference of the European Chapter of the Association for Computational Linguistics. 2017 Presented at: ACL'17; April 3-7, 2017; Valencia, Spain. [doi: 10.18653/v1/e17-2068]

57. Liu P, Qiu X, Huang X. Recurrent Neural Network for Text Classification With Multi-task Learning. In: The Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016 Presented at: CAI'16; April 7-9, 2016; New York, USA. [doi: 10.5555/3060832.3061023]

58. Raffel C, Ellis D. Feed-forward networks with attention can solve some long-term memory problems. arXiv Preprint posted online December 29, 2015. [FREE Full text]

59. Rakhlin A. Convolutional Neural Networks for Sentence Classification. GitHub. 2016. URL: https://github.com/alexander-rakhlin/CNN-for-Sentence-Classification-in-Keras [accessed 2020-11-10]

60. Zhang Y, Wallace B. A Sensitivity Analysis of (and Practitioners Guide to) Convolutional Neural Networks for Sentence Classification. In: The Eighth International Joint Conference on Natural Language Processing. 2017 Presented at: NLP'17; November 4-9, 2017; Taipei, Taiwan URL: https://arxiv.org/abs/1510.03820

61. Amazon Mechanical Turk. URL: https://www.mturk.com/ [accessed 2020-11-10]

62. Tong L, Cheng Q, Homan C, Silenzio V. Learning From Various Labeling Strategies for Suicide-related Messages on Social Media: an Experimental Study. arXiv Preprint posted online January 30, 2017 [FREE Full text]

63. Shneidman ES. Suicide as psychache. J Nerv Ment Dis 1993 Mar;181(3):145-147. [doi: 10.1097/00005053-199303000-00001] [Medline: 8445372]

64. DeJong CA, van den Brink W, Harteveld FM, van der Wielen EG. Personality disorders in alcoholics and drug addicts. Compr Psychiatry 1993;34(2):87-94. [doi: 10.1016/0010-440x(93)90052-6] [Medline: 8387417]

65. Khoury L, Tang YL, Bradley B, Cubells JF, Ressler KJ. Substance use, childhood traumatic experience, and posttraumatic stress disorder in an urban civilian population. Depress Anxiety 2010 Dec;27(12):1077-1086 [FREE Full text] [doi: 10.1002/da.20751] [Medline: 21049532]

66. Livingston JD, Milne T, Fang ML, Amari E. The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. Addiction 2012 Jan;107(1):39-50 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03601.x] [Medline: 21815959]

67. Agarwal V, Zhang L, Zhu J, Fang S, Cheng T, Hong C, et al. Impact of predicting health care utilization via web search behavior: a data-driven analysis. J Med Internet Res 2016 Sep 21;18(9):e251 [FREE Full text] [doi: 10.2196/jmir.6240] [Medline: 27655225]

68. Shin J, Madotto A, Fung P. Interpreting Word Embeddings With Eigenvector Analysis. In: NIPS 2018 Interpretability and Robustness for Audio, Speech and Language Workshop. 2018 Presented at: NIPS'18; March 6-10, 2018; Montreal, Canada URL: https://openreview.net/pdf?id=rJfJiR5ooX

69. Ling W. Prescription opioid addiction and chronic pain: more than a feeling. Drug Alcohol Depend 2017 Apr 1;173(Suppl 1):S73-S74 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.08.643] [Medline: 28363323]

70. McGinty E, Pescosolido B, Kennedy-Hendricks A, Barry CL. Communication strategies to counter stigma and improve mental illness and substance use disorder policy. Psychiatr Serv 2018 Feb 1;69(2):136-146 [FREE Full text] [doi: 10.1176/appi.ps.201700076] [Medline: 28967320]

71. Kelly JF, Westerhoff CM. Does it matter how we refer to individuals with substance-related conditions? A randomized study of two commonly used terms. Int J Drug Policy 2010 May;21(3):202-207. [doi: 10.1016/j.drugpo.2009.10.010] [Medline: 20005692]

72. Large MM. The role of prediction in suicide prevention. Dialogues Clin Neurosci 2018 Sep;20(3):197-205 [FREE Full text] [Medline: 30581289]

Abbreviations

API: Application Programming Interface
ATTENTION: attention-based bidirectional recurrent neural network
CNN: convolutional neural network
FAST: FastText
GloVe: Global Vectors for Word Representations
LR: logistic regression
MTURK: Amazon Mechanical Turk
NN: neural network
OUD: opioid use disorder
PRAW: Python Reddit API Wrapper
RF: random forest
RNN: recurrent neural network
SVM: support vector machine
TF-IDF: term frequency-inverse document frequency

XSL•FO
RenderX