12-1-2021

Detecting risk level in individuals misusing fentanyl utilizing posts from an online community on Reddit

Sanjana Garg

Jordan Taylor

Mai El Sherief

Erin Kasson

Talayeh Aledavood

See next page for additional authors

Authors

Sanjana Garg, Jordan Taylor, Mai El Sherief, Erin Kasson, Talayeh Aledavood, Raven Riordan, Nina Kaiser, Patricia Cavazos-Rehg, and Munmun De Choudhury

Internet Interventions 26 (2021) 100467

Contents lists available at ScienceDirect

Internet Interventions

journal homepage: www.elsevier.com/locate/invent

Detecting risk level in individuals misusing fentanyl utilizing posts from an online community on Reddit

Sanjana Garg [a,1], Jordan Taylor [a,1], Mai El Sherief [a], Erin Kasson [b], Talayeh Aledavood [c], Raven Riordan [b], Nina Kaiser [b], Patricia Cavazos-Rehg [b], Munmun De Choudhury [a,*]

[a] College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, United States of America
[b] Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63130, United States of America
[c] Department of Computer Science, Aalto University, Espoo, Finland

ARTICLE INFO

Keywords:
Fentanyl
Opioids
Overdose
Machine learning
Detection
Social media

ABSTRACT

Introduction: Opioid misuse is a public health crisis in the US, and misuse of synthetic opioids such as fentanyl have driven the most recent waves of opioid-related deaths. Because those who misuse fentanyl are often a hidden and high-risk group, innovative methods for identifying individuals at risk for fentanyl misuse are needed. Machine learning has been used in the past to investigate discussions surrounding substance use on Reddit, and this study leverages similar techniques to identify risky content from discussions of fentanyl on this platform.

Methods: A codebook was developed by clinical domain experts with 12 categories indicative of fentanyl misuse risk, and this was used to manually label 391 Reddit posts and comments. Using this data, we built machine learning classification models to identify fentanyl risk.

Results: Our machine learning risk model was able to detect posts or comments labeled as risky by our clinical experts with 76% accuracy and 76% sensitivity. Furthermore, we provide a vocabulary of community-specific, colloquial words for fentanyl and its analogues.

Discussion: This study uses an interdisciplinary approach leveraging machine learning techniques and clinical domain expertise to automatically detect risky discourse, which may elicit and benefit from timely intervention. Moreover, our vocabulary of online terms for fentanyl and its analogues expands our understanding of online "street" nomenclature for opiates. Through an improved understanding of substance misuse risk factors, these findings allow for identification of risk concepts among those misusing fentanyl to inform outreach and intervention strategies tailored to this at-risk group.

1. Introduction

Amidst the opioid epidemic (Gostin et al., 2017), synthetic opioid misuse in particular has become an urgent public health crisis since 2013 when these illicitly manufactured synthetics started to become more readily available (DEA, 2015; DEA, 2018), contributing to nearly 12 times the number of overdose related deaths in 2019 than in 2013 (CDC, 2019). Fentanyl is a synthetic opioid that in particular is considered a serious threat (Springer et al., 2019), as it has driven the most recent wave of synthetic opioid deaths (Spencer et al., 2019; CDC, 2018). In 2016, fentanyl became the drug most frequently mentioned in relation to overdose deaths in the United States, surpassing heroin (Hedegaard et al., 2018). It is a highly potent drug, making it incredibly easy for users to become addicted as well as users of other drugs to unintentionally overdose on it, and it is often laced into substances without user's knowledge, making risk for overdose much higher (Jones et al., 2018; NIDA, 2019). In fact, a study found that 73% of the participants who tested positive for fentanyl did not report fentanyl misuse, suggesting that they unknowingly injected or consumed the drug (LaRue et al., 2019; Amlani et al., 2015).

Many instances of overdose and harm related to fentanyl use in the United States are unintentional, often related to use of heroin, cocaine, and other drugs which are laced with fentanyl to increase their euphoric effects (CDC, 2021; NIDA, 2019). With regard to individuals who engage

* Corresponding author at: College of Computing, Georgia Institute of Technology, 756 W Peachtree St NW, Atlanta, GA 30308, United States of America.
E-mail address: mchoudhu@cc.gatech.edu (M. De Choudhury).
[1] Co-first authors.

https://doi.org/10.1016/j.invent.2021.100467
Received 26 July 2021; Received in revised form 25 September 2021; Accepted 1 October 2021
Available online 20 October 2021

in intentional fentanyl misuse, however, theories about motivations for misuse include a high tolerance for other drugs including opioids that requires a more potent drug such as fentanyl to maintain a high, motivations related to addiction/dependence and misuse to reduce or mitigate withdrawal symptoms, or addiction resulting from accidental exposure to fentanyl (Buresh et al., 2019). Understanding motivations for and risks related to patterns of intentional fentanyl misuse via firsthand experiences is crucial to better adapt prevention and intervention strategies effectively. Yet, due to the illegal nature and stigma surrounding drug misuse, populations of illicit drug users are often difficult to reach. Research into motivations for illicit drug misuse also reflects a dynamic relationship between feelings of separation or isolation from others (i.e., social pain) as well as physical pain, and such pain may encourage cyclical, continued misuse of substances to reduce these feelings (Eisenberger, 2012; Sullivan and Ballantyne, 2021). Fentanyl misuse in particular is highly stigmatized; sufferers are often known to feel reticent to share their experiences with clinicians, researchers, and even family and friends (Nelson and Perrone, 2012). Consequently, for those labeled an "addict," this stigma causes many negative outcomes, including shame, embarrassment and unwillingness to enter treatment (Livingston et al., 2012). This makes it difficult to gather data to understand practices and risk factors surrounding fentanyl misuse through interviews or patient disclosures.

Pseudonymous social media sites can empower those with stigmatized identities to disclose experiences and seek support with diminished fear of offline harm (Andalibi et al., 2016). As such, pseudonymous social media has been used to study stigmatized experiences in the domains of LGBTQ+ minority stress (Saha et al., 2019), sexual abuse (Andalibi et al., 2016), parenting (Ammari et al., 2019), and mental health (De Choudhury and Sushovan, 2014; Pavalanathan and De Choudhury, 2015; De Choudhury et al., 2016; Naslund et al., 2016; De Choudhury and Kiciman, 2017; Andalibi et al., 2017; Cavazos-Rehg et al., 2017; Guntuku et al., 2017; Paul and Dredze, 2017; Coppersmith et al., 2018). One widely used pseudonymous social media site is Reddit, which offers topic-specific forums, known as subreddits, where users can vote and comment on each other's posts anonymously, empowering users to discuss stigmatized topics (Singer et al., 2014; Betton et al., 2015; Andalibi et al., 2016; De Choudhury et al., 2016; Robinson et al., 2019). According to a Pew survey, in America, Reddit is used by about 15% of adult men, 8% of adult women, 22% of those ages 18 to 29, and 14% of those ages 30 to 49 (Perrin and Anderson, 2019). In the same survey, 14%, 12%, and 4% of Hispanic, White, and Black Americans respectively and 9%, 10%, and 15% of those with annual incomes less than $30,000, $30,000 to $74,999, and greater than $75,000 respectively reported using Reddit. Previous studies have used machine learning and natural language processing methods to analyze Reddit posts regarding casual drug discussions, opioid addiction, and alternative treatments for opioid use recovery (Park and Conway, 2018; Chancellor et al., 2019; Lu et al., 2019; Alambo et al., 2021). However, no known studies focus on fentanyl discussions in particular. Further, to our knowledge, none of the existing data and language analytic studies have focused on identifying or understanding specific risky behaviors associated with fentanyl misuse.

In response, the present study examines the content in the social media platform Reddit, specifically within the subreddit r/fentanyl. We use posts and comments from the r/fentanyl subreddit to assess fentanyl misuse risk factors using a mixed methods approach — first by developing a codebook using qualitative content analysis on the forum, and then building and validating supervised machine learning classifiers to detect risk. By identifying risk associated with intentional fentanyl misuse from social media, this work improves understanding of this substance for better clinical research, treatment and interventions, and outreach to populations which may be difficult to reach via conventional means.

2. Methods

2.1. Social media data

We collected public data from the subreddit r/fentanyl. The r/fentanyl forum describes itself as dedicated to harm reduction and the exchange of information about fentanyl and its analogues that offers firsthand user experiences and advice with the goal "to dispel some common myths about these substances." We collected all posts and comments (content) from the beginning of the subreddit's history in May of 2015 to January 2020 using PRAW[2] library (Boe, 2016) and Google Big Query (BigQuery, 2019).[3] Our dataset from r/fentanyl is summarized in Table 1; it includes 6459 posts and comments from 1124 unique users. However, as can be seen in Table 2 120 posts and 361 comments were written by users who later deleted their account, so the number of authors in our dataset is likely larger than 1124.

Since these data were collected from the publicly available subreddit r/fentanyl, this study does not constitute human subjects research and was not subject to institutional review. To protect user identities, we have not included usernames, nor direct quotes; also, example quotes are paraphrased to reduce traceability.

To help with the automated detection of fentanyl misuse on this Reddit forum, we first identified posts and comments that were first-person reports of fentanyl use; this is important because these communities harbor a variety of content ranging from attitudes about fentanyl use, personal experiences, news and misinformation, side-effects of use, as well as experiences about relapse and abstinence. For this, we employed a machine learning classifier developed on the annotated medication intake Twitter dataset by Klein et al. (2017). This classifier classifies posts into two categories: intake (self-reported) and no-intake. Using this classifier in a transfer learning setting (Howard et al., 2020), a total of 1628 posts and comments in our initial dataset of 6459 were identified to be about first-hand self-reports of fentanyl intake (Table 1). We used this classifier to select a 60% intake 40% no-intake content split on r/fentanyl for 391 annotated posts or comments. We chose to include data classified as no-intake because our codebook contained categories not associated with drug intake, such as discussing withdrawal, tolerance, or color. As can be seen in Table 2, the distribution of posts and comments from deleted users in our annotated dataset is comparable to the proportion of posts and comments with text from deleted users in our entire dataset. We can also see in Table 3 that while most users only have one post or comment in our annotated dataset, these users tend to be more frequent posters or commenters than the median user. This is to be expected because the number of posts and comments per user follows a power law distribution.

Table 1
Data description from r/fentanyl.

	#	# users	# avg. words	# intake	# no-intake
Posts	804	422	88	207	597
Comments	5655	980	54	1421	4234
Total	6459	1124	59	1628	4831

[2] PRAW ("Python Reddit API Wrapper") is a Python package that provides an interface to scrape data from Reddit using Reddit's API.

[3] Google BigQuery is a platform that enables analysis over big data of the order of petabytes. Along With that, it also contains Reddit posts and comments from different subreddits stored as datasets. We used Google BigQuery only for retrieving data and not for analysis.

Table 2
Post and comment distribution from users who deleted their account.

		Num from deleted author	Total	Percent from deleted author
All data	All	481	6459	7.45%
	Posts	120	804	14.93%
	Comments	361	5655	6.38%
All data with text*	All	83	5744	1.44%
	Posts	13	387	3.36%
	Comments	70	5357	1.31%
Annotated data	All	5	391	1.28%
	Posts	1	42	2.38%
	Comments	4	349	1.15%

* Some types of posts don't have a text body, and some comment and post texts in our dataset simply say they were removed or deleted.

Table 3
Description of the number of posts and comments per author in our entire dataset and in our annotated dataset.

		Mean per Author in Entire Dataset	Median per Author in Entire Dataset
All data authors	All	5.32 (±15.14)	2.00
	Posts	0.61 (±1.39)	0.00
	Comments	4.71 (±14.47)	1.00
Annotated data authors	All	16.41 (±31.49)	7.00
	Posts	1.26 (±1.96)	1.00
	Comments	15.15 (±30.19)	7.00

		Mean per author in annotated dataset	Mean per author in annotated dataset
Annotated data authors	All	1.78 (±1.91)	1.00
	Posts	0.19 (±0.41)	0.00
	Comments	1.59 (±1.83)	1.00

2.2. Qualitative data annotation

On this filtered dataset, we now describe a qualitative approach to code risk levels. Given a lack of existing frameworks to support coding of risk levels in social media posts, inductive and deductive methods (Braun and Clarke, 2006) were used to develop a codebook to delineate factors related to risk in posts and comments from the sample. First, using an inductive approach, a subset of roughly 100 posts from the filtered intake sample were reviewed by human coders to determine the types of risk behaviors commonly discussed within this subreddit. These general themes (e.g., tolerance/withdrawal, access to substance, route of administration, identification) were then compared to empirically supported factors related to risk of substance misuse as outlined in previous literature [i.e., injection drug use (Kenney et al., 2018), higher physical and mental morbidity burdens (Smolina et al., 2020)]. More specifically, final annotation codes created to specifically identify imminent substance misuse risk with consideration of unique fentanyl misuse risk factors included: mentions (1) he/she is a regular drug user (Degenhardt et al., 2010), (2) a high substance tolerance (Darke and Hall, 2003) or withdrawal (Bluthenthal et al., 2020), (3) a previous overdose or knowing others who have overdosed (Britton et al., 2010), (4) polysubstance use (Betts et al., 2015; Coffin et al., 2003), (5) current access to or actively seeking the substance (Paulozzi, 2012), (6) functional (Barash et al., 2017) and quality of life impairments (Zibbell et al., 2019), (7) intravenous method of use (Britton et al., 2010), and (8) drugs being cut with another substance (LaRue et al., 2019). Factors further extended to seeking advice on dosage or use methods, as well as supportive commenting for risky drug use (Webster, 2017).

Once the codebook was established and refined, two clinical annotators reviewed batches of around 200 posts/comments at a time to assign codes and risk level sums for each. Inter-rater reliability ranged from 0.71 to 0.99 for specific risk codes and was 0.64 for risk level assigned, all within or exceeding substantial agreement (Landis and Koch, 1975; McHugh, 2012). A third consensus coder further reviewed and coded those upon which there was disagreement, which occurred in 36% of cases (Syed and Nelson, 2015). These annotations were then used to inform the machine learning models. Coders read each post or comment and coded a "0" if none of the risk factors were in the post/comment or "1" if a risk concept was present in the post/comment. Coders then summed the number of codes present into a total score to identify level of risk for that post/comment. If 1 or more codes were present in the post/comment, this post was categorized as a post with elevated risk (coded with "1"). If no code was present or the post contained too little information to code, the post was categorized as a post with low risk (coded with "0"). We acknowledge that many members of this community are at some level of risk, which is why we refer to the "0" class as "low risk" rather than "no risk." Also, we note that our codebook addresses the risk factors disclosed within the text of posts or comments, not account-level risk. The annotated data is described in Table 4.

2.3. Machine learning based risk detection

Our fentanyl use risk codebook is extensive to capture the multiple facets of risk surrounding the use of this substance, but this extensiveness makes it intractable and expensive for experts to label every post/comment. Therefore, to understand discursive risk on r/fentanyl, we built multiple machine learning classifiers using the content annotated by domain experts. Generally speaking, classification is the process of predicting the class of given data points. We built four classifiers, established in the literature, using the annotated data: logistic regression, Support Vector Machine (Noble, 2006), random forest (Breiman, 2001), and long short-term neural network (LSTM) classifiers (Hochreiter and Schmidhuber, 1997). These classifiers used features that captured the frequency, co-occurrence of words and rarity of words in posts specific to a risk code. We used 80% of our annotated data to train our models (that is, the models learned patterns embedded in the data) and tested on the remaining 20% (that is, based on the patterns learned during training, for an unseen data point, the models guessed which category it was the most likely to belong to). Since we had significant class imbalance, we employed SMOTE, or (Chawla et al., 2002). Further details of our models are expanded upon in the "Classification Models" section of the supplementary document. In summary, our classifiers used the language within our expert-annotated posts and comments to predict whether the risk level assigned by domain experts indicated low risk or elevated risk (Fig. 1).

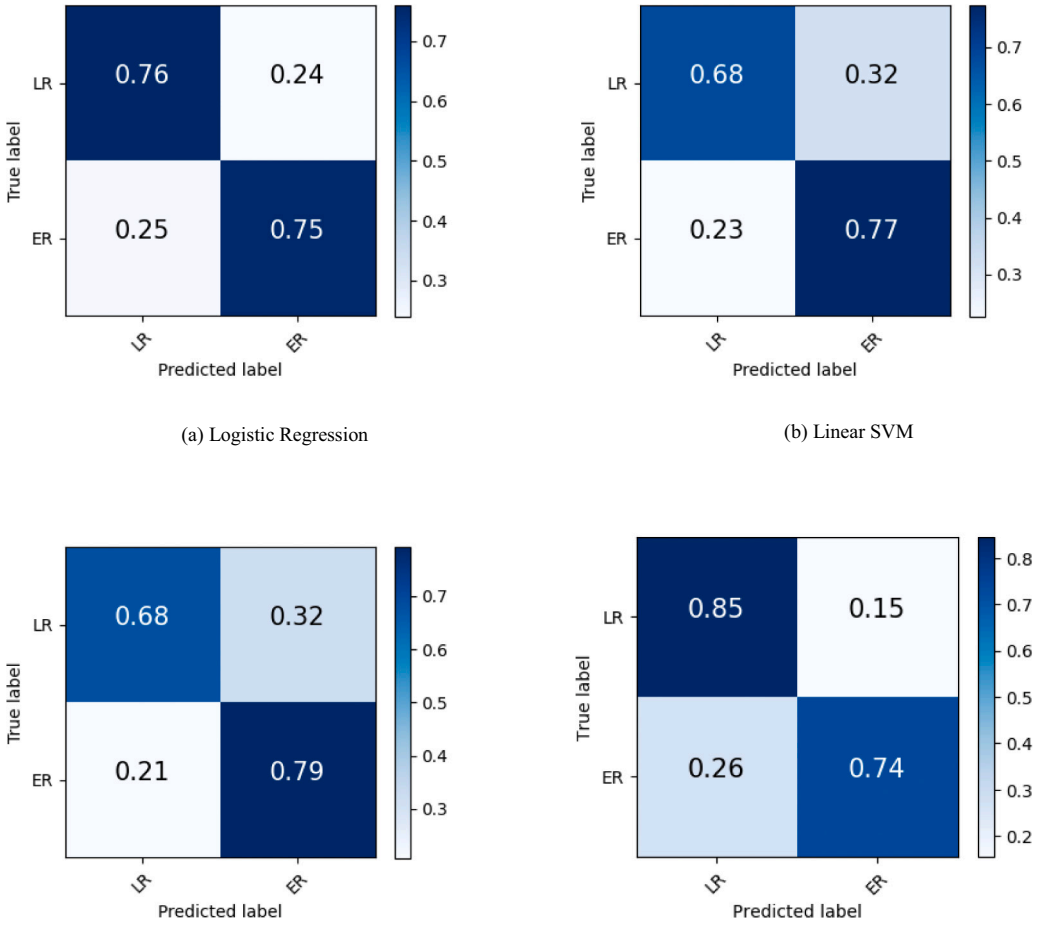
3. Results

The performance of our risk models is summarized in Table 5 and includes the mean precision, recall, F1, accuracy and AUC across each fold corresponding to the hyperparameter setting with highest mean AUC reported for each classifier. We also include the performance of a classifier with the same aforementioned hyperparameters trained on our entire training set and tested on our test set. On our test set each model's accuracy ranged from 0.74 to 0.76, precision from 0.71 to 0.76, and recall from 0.73 to 0.79. However, the differences between the models were not statistically significant when one considers the standard deviation across each metric in Table 5 during cross validation. In summary, no model outperformed another on risk level classification. Lastly, while our models seem to have performed similarly on posts and comments, such as our LSTM model correctly classifying 8 of the 10 posts in our test set, there are too few posts to draw conclusions about the performance comparing posts and comments.

Additionally, to ensure our models were learning signals directly relevant to the risk of fentanyl and thus establish the construct validity of our models (O'Leary-Kelly and Vokurka, 1998), we created a baseline model using length of the text and the number of drug occurrences as features to compare with our more complex language models (or the

Table 4
Annotated data statistics.

	Low risk					Elevated risk				
	#	# users	# avg. words	Intake	No-intake	#	# users	# avg. words	Intake	No-intake
Posts	2	2	177	0	2	40	37	255	31	9
Comments	144	99	26	55	89	205	135	103	156	49
Total	146	101	28	55	91	245	162	128	187	58



(a) Logistic Regression

(b) Linear SVM

(c) Random Forest

(d) LSTM

Fig. 1. Confusion matrix for each classifier. (LR — low risk, ER — elevated risk).

classifiers described in Methods section). Based on Table 5, all our models outperformed this baseline of 0.62 F1 score. We note that each of our classifiers performed similarly with respect to F1 score on the held-out test set: the F1 scores for each classifier range from 0.72 to 0.75. We can also see in the ROC curves in Fig. 2 that the models perform similarly with respect to false positives and false negatives as the decision boundary threshold is varied. Relatedly, each model has a similar AUC. To elucidate these results further, Fig. 2 presents the ROC curves for each classifier.

For simpler classification models like random forest, logistic regression and support vector machine, one can quantify the importance of features by ranking the coefficient of each feature in the trained model, higher the coefficient implies higher correlation of the feature with the positive class (in our case high-risk class). Table 6 shows the top 15 important features (words or phrases) in the risk classification task. As expected, the word "drugxyz", which is used to represent community specific drug names like "fent" and "carfent," features prominently in the top two salient words or phrases for every classifier. We further observe that the first person pronoun "I" appears in the top 15 most important words or phrases for our random forest and logistic regression classifier, and the phrase "I know" is in the top 15 for our support vector machine classifier. Upon inspection, "I know" is likely correlated with elevated risk because the phrase is used to bolster one's credibility when providing advice, such as "I know because I'm a regular fentanyl user", to provide social proof (Cialdini, 1987) when giving advice, "I know people who...", and to hedge (Lakoff, 1975) risky personal narratives, such as "I know it's stupid but...". Moreover, words associated with procurement, such as "get" and "buy", and dosing, such as "mg" and "one," appear in the top 15 most important words and phrases for our random forest and logistic regression classifiers. Meanwhile, the word "terrible," which is associated with withdrawal and drug use personal narratives, appears in the 15 most important words and phrases for both our logistic regression and support vector machine classifiers. It is important to note that our neural network based LSTM model cannot be

Table 5
Macro-average model performances on 5-fold cross validation on 80% of our annotated data and performances of models trained on our training set (80% of our annotated data) and evaluated on our test set (20% of our annotated data).

Features	Cross validation				
	Precision	Recall	Macro-F1	Accuracy	AUC
N-Gram	0.82	0.81	0.81	0.81	0.91
L + D	(±0.11)	(±0.09)	(±0.09)	(±0.09)	(±0.12)
N-Gram	0.82	0.81	0.81	0.81	0.89
L + D	(±0.10)	(±0.09)	(±0.08)	(±0.09)	(±0.12)
TFIDF	0.84	0.83	0.82	0.83	0.92
L + D	(±0.04)	(±0.04)	(±0.05)	(±0.04)	(±0.06)
BERT	0.82	0.78	0.78	0.81	0.87
	(±0.05)	(±0.06)	(±0.06)	(±0.05)	(±0.04)
Baseline	0.75	0.75	0.71	0.72	0.86
	(±0.08)	(±0.09)	(±0.10)	(±0.10)	(±0.09)

Features	Test set				
	Precision	Recall	Macro-F1	Accuracy	AUC
N-Gram L + D	0.73	0.76	0.74	0.76	0.81
N-Gram L + D	0.71	0.73	0.72	0.74	0.78
TFIDF L + D	0.72	0.74	0.73	0.76	0.79
BERT	0.76	0.79	0.76	0.77	0.83
Baseline	0.64	0.66	0.61	0.62	0.66

LR — logistic regression, SVM — linear support vector machine, RF — random forest, LSTM NN — long short-term neural network.
L + D — lemmatized and debiased (see Supplement section "Debiasing" for more information about debiasing).

similarly interpreted due to its complex internal structure (Castelvecchi, 2016).

We then conducted qualitative analysis on what was being detected as elevated risk by our classifier. Table 7 shows some examples of instances correctly classified by our logistic regression classifier as elevated risk along with the risk factors associated with each instance. In this table darker color represents higher importance of that word during classification. For example, "off", "get", "tolerance" are important features for people with high tolerance or withdrawal. Similarly, "my", "family", "pay" are important words for functional and quality of life impairments risk factor. The example for use of additional substances is also notable because it shows the benefit of debiasing drug names. Since the classifier is able to map words "carfent" "logue", "butyr", "xanax" to "drugxyz", it can learn the underlying notion that multiple substances are being mentioned which helps with the task of risk detection, though they may be rare drugs to occur.

4. Discussion

In an effort to alleviate the severe public health threat fentanyl misuse poses (CDC, 2018), the present study utilizes machine learning supported by manual domain expert annotation on data from a public, anonymous forum, r/fentanyl, to identify content with elevated risk factors around fentanyl misuse. This way our findings provide novel data and language analytic methods on the study of specific risky behaviors associated with substance misuse. Notable strengths of this paper include the use of a popular social media platform that protects users' privacy while facilitating authentic conversation around stigmatizing or incriminating topics. This allows us to evaluate firsthand experiences, fentanyl misuse and personal risk factors among a high risk and masked population to adapt prevention and rehabilitation programs effectively. Summarily, our work may help support the development and provision of timely treatment and interventions to those in need, while also expanding outreach methods to populations that are difficult to reach via traditional means.

4.1. Practical implications

This research shows that the anonymity afforded by social media sites like Reddit allows individuals to discuss stigmatized topics such as illegal substances (Birnholtz et al., 2015) and fentanyl misuse. Moreover, some of these individuals may be at elevated risk that could be detected via computational methods; consequently, this work can pave the way to provide preventative support or clinical intervention to especially vulnerable individuals on these forums. Furthermore, the r/fentanyl forum facilitates conversation surrounding harm reduction and information about fentanyl and its analogues, providing both a place for advice seeking and social support, as well as an exchange of information related to the use of these substances. Reddit's popularity and ability to facilitate discussion on specific stigmatizing topics, coupled with the computational methods developed in this work can, therefore, aid in both the timely and targeted outreach to fentanyl misusers, an especially hard-to reach population (Miller and Sønderlund, 2010; Wejnert and Heckathorn, 2012). This may, in turn, satisfy the need to identify individuals for harm reduction interventions, while maintaining their privacy and encouraging real conversations among other individuals using fentanyl, which may be therapeutic or harm reducing within themselves (Latkin et al., 2003).

Especially in light of increased treatment barriers due to COVID-19, utilizing accessible methods to learn about, target and engage high-risk members of difficult to reach communities in treatment is critical. Accordingly, given the above noted potential for practical use, we discuss two implications for online communities discussing substance use and misuse.

First, prior research suggests that Reddit moderators play an active role in managing the content on their subreddit, defining community-specific rules, establishing norms, and providing support to people who post acutely concerning content (De Choudhury and Sushovan, 2014; De Choudhury et al., 2016; Chandrasekharan et al., 2019). Moreover, auto-moderation tools play an important role in subreddit moderation, empowering moderators with technology-mediated approaches for triaging concerning content, especially in contexts where fully manual triage might be demanding on moderators' time and effort (Jhaver et al., 2019). In fact, research by Matias et al. that conducted an online experiment on auto-moderation strategies has found the approach to be helpful to content moderation, as well as to enforce and uphold community norms against harassment (Matias, 2019). In light of this research, our work could inform the design of tools to support the work of substance use related online community moderators. For instance, our results could inform the design of tools to help moderators target users posting risky content for interventional outreach, as discussed in recent studies on marginalized populations appropriating social media for health needs (Andalibi et al., 2016; Saha et al., 2020; Wadden et al., 2021). Facebook similarly uses artificial intelligence (AI) to provide resources to those identified as being at risk for suicide based on a partnership with the National Suicide Prevention Lifeline (Constine, 2017), and in March 2020 Reddit announced a partnership with Crisis Text Line to allow users to flag other users who may be in crisis (Perez, 2020). We envision that by pairing our computational approach and using content contributed by non-profit organizations for targeted outreach, moderators will be better equipped to deal with risky messages, in turn, not only helping improve the overall quality of discussions in these forums, but also generate for and share information with public health entities about strategies to invest in prevention and intervention campaigns.

Notwithstanding these opportunities and implications for content moderation, we strongly discourage using our methods to remove content or ban users from posting on these forums. As shown by Chancellor et al. in the context of online pro-eating disorder communities, content removal can be both ineffective and harmful for addressing deviant behavior on social media (Chancellor et al., 2016). When individuals experience vulnerability, they tend to reach out to others to "buffer"

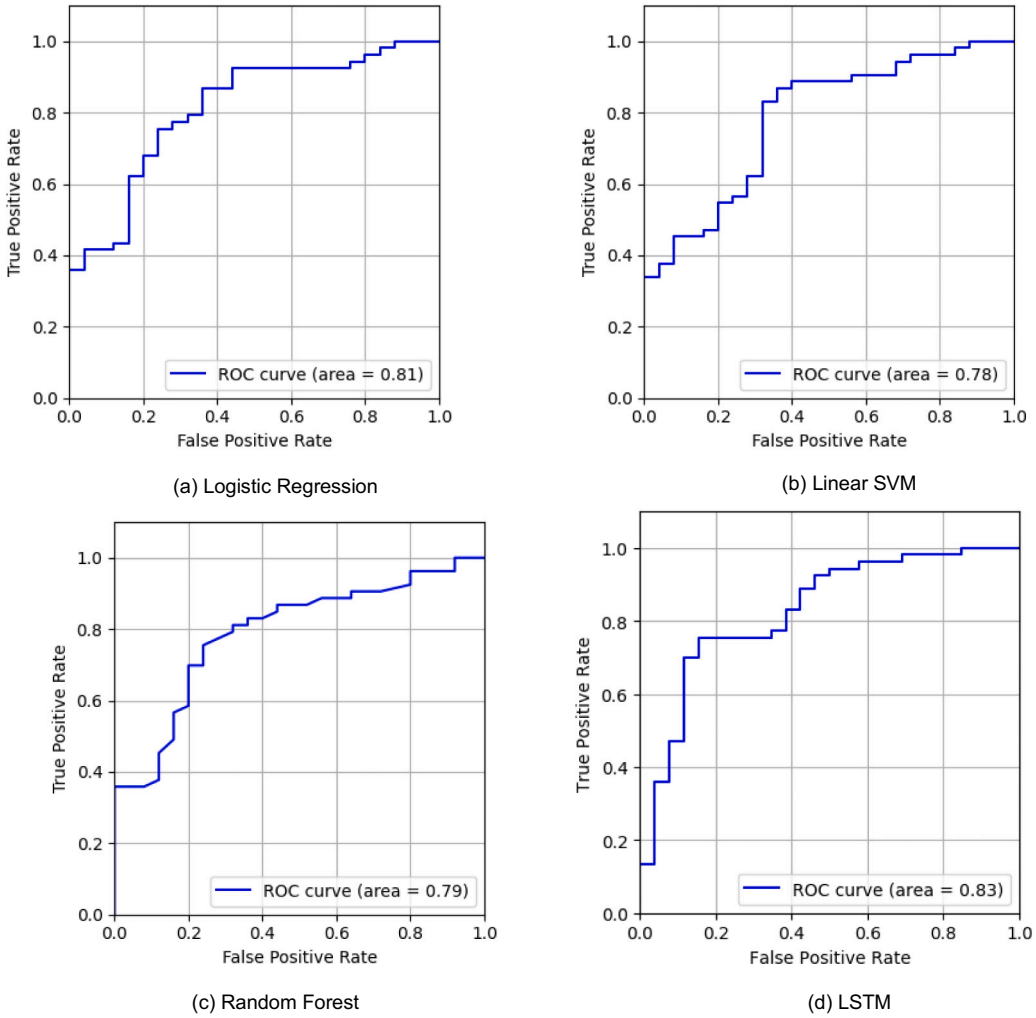


Fig. 2. ROC (receiver operating characteristic) curves for each classifier.

Table 6
Comparison of top features across three top performing classifiers. Weights denote the feature importance.

RF	β	LR	β	SVM	β
drugxyz	0.038	one	1.064	drugxyz still	0.93
i	0.032	drugxyz	0.797	better	0.796
have	0.028	it if	0.77	one	0.603
my	0.02	drugxyz still	0.699	it if	0.509
it	0.017	all	0.645	all	0.401
do	0.016	better	0.641	terrible	0.395
me	0.015	i	0.635	connection	0.385
shit	0.013	some	0.631	if it	0.381
one	0.012	me	0.627	take it	0.374
get	0.012	i know	0.619	who have	0.372
way	0.011	off	0.598	i know	0.37
like	0.01	terrible	0.595	maybe	0.363
with	0.01	maybe	0.584	make	0.331
mg	0.01	buy	0.571	lab	0.33
some	0.01	make	0.558	methyl	0.329

themselves against negative emotions and actions, and online communities provide one such powerful mechanism (Coyne and Downey, 1991). Therefore, any intervention, based on this work, to empower the community moderators would need to ensure that communities like the one studied in this paper, continue to provide an outlet to seek out these kinds of "safety valves" to individuals to regulate their emotions when they need (Acton, 1973).

Second, we note that past studies among Reddit users misusing opioids (Cavazos-Rehg et al., 2019; Cavazos-Rehg et al., 2021) identified numerous barriers to treatment (e.g., stigma/shame, attitudes towards treatment, treatment readiness) and such barriers may also be reflected in limited openness to direct, proactive outreach strategies among users within these online communities. As a potential approach towards addressing this challenge, based on prior research on peer support in online health communities (Yang et al., 2019; Yang et al., 2019a), our work could be used to empower community moderators, such that they can make appropriate support provisions involving veteran members or other supportive members in the community, who are willing to do so and have been screened to be equipped to help. In addition, our work could be used to connect with and directly query members of communities for specific input to inform feasible, acceptable, and actionable methods of outreach to reduce harm and provide support to better tailor the use of such risk information when detected. These above approaches will certainly have to be tempered with appropriate privacy protections and ethical considerations, so that individuals continue to feel safe in the attempts to discuss substance misuse challenges on online forums, and so that risk detection does not increase harm.

4.2. *Methodological implications*

Next, there are some methodological implications worth discussing related to implementation. As mentioned earlier, no model performed

Table 7
Correctly classified examples with associated risk factors and top features highlighted.

Risk Factor	Example with top contributing features
High tolerance or withdrawal	I used to get a good nod off those sort of doses , but my tolerance has increased a bit, and now those doses are the minimums to stay well for me . When I want to get high , I just take a few extra sprays . Or I get some real dope - which is much better for getting high , but way too expensive to maintain a proper habit .
Functional and quality of life impairments	Great to hear that. It's a long road. I just heard that my employer is getting rid of my position and my next check will be a whole week short which puts my family in a bad position and we might get evicted. I'm seriously pondering killing myself so my family can collect insurance and get out of this hole. I'm so lost . I get clean and now this happens . I don't know what to do. I did this to us and I should pay .
Use of additional substances + Supportive comment for risky drug use	Dude, I' ve been wanting to give it a shot for a long time now , hearing of people mixing up a mg per L blows my mind. Depending on the logue' I have , it 's anywhere from 50/150 mg per 30 mL. Careful you titrate slowly with that carfent , I know many some batches circulating on the DN are under 90%, possibly substantially under, be cautious and report back how it went ! I just dropped from 75 mg of vaporized butyr a day and tried switching to subs for the first time, 19 hours in I was able to take a 3 mg dose and some xanax to knock me out for a hour or two, woke up feeling 8/10 which was a nice suprise. Edit: make sure you glove up , mask up , and robe up when you get ready to undergo the titration procedure with the car .

statistically significantly better than another across accuracy, precision, recall, and f1, but Fig. 2 shows the sensitivity and specificity tradeoffs associated with each model's output decision boundary thresholds for binary classification. The preferred threshold would depend on the application of this classifier in a real-world scenario. For example, if a risk classifier with low specificity is used to support moderators of a drug related online community, then the implications of a community member being falsely detected to post risky content could alienate users and suppress the potential use of these online communities as a "safety valve", a concern echoed by Chancellor et al. (2016) and Jhaver et al. (2019). On the other hand, clinical use of the insights of the risk detection model for screening purposes may prefer greater false positives over greater false negatives, because it would minimize the likelihood that individuals showing exacerbated levels of risky fentanyl use are missed and therefore precluded from getting an intervention. Future research must carefully consider these tradeoffs when deciding on decision boundary thresholds.

Although each model performed similarly on classification, there are noticeable differences in how easy the models' features and decisions are to interpret and audit. For instance, our LSTM neural network model used BERT embeddings (Devlin et al., 2018), representing each post or comment as a 512×768 matrix, features that are difficult to interpret directly. On the other hand, the *n*-gram feature vectors, or vectors made by counting words and phrases in the text, used by our logistic regression model and SVM model, are easier to interpret directly. Additionally, it is easier to understand the output of a logistic regression classification model than an LSTM because the former classifies by applying a simple, linear function to input vectors. Meanwhile, LSTM model outputs are complex functions of their input vectors. This means it would be easier for a domain expert to audit the output of our logistic regression model with n-gram features than our BERT LSTM model. As we saw no significant classification improvements between our logistic regression and LSTM models, those using this work should prefer the simpler logistic regression model to the latter to empower domain experts to more easily audit model outputs.

Additionally, the list of community specific drug names we found to improve our models in Table 8 can inform future research on online communities where opioids, fentanyl or fentanyl analogues are discussed. For example, Sarker et al. (2019) found a statistically significant correlation between Pennsylvania county-level overdose death rates and the misuse-indicating social media posts labeled using a machine learning classifier built on a set of "prescription and illicit opioid names, including street names and misspellings" to collect opioid-mentioning Twitter posts. Balsamo et al. (2019) used similar methods to construct a vocabulary of over 700,000 terms associated with opiate related subreddits, but fentanyl analogues appear infrequently because the list was constructed from multiple opiate related subreddits and terms occurring in less than 100 posts or comments were removed. Furthermore, Balsamo et al.'s (2019) vocabulary is not annotated for drug names. Our approach that focused on automatic identification of fentanyl analogues extends this research, opening opportunities for future work that examines harm reduction strategies associated with various genres of opiates.

Moreover, while our aforementioned vocabulary is applicable to the narrow domain of fentanyl analogues, our method of using word embeddings to find online community specific drug names, discussed in the "Data Filtering for Annotation" section of our supplement, can be used to help clinicians understand other colloquial drug names. Notably, Lee and Antin (2012) found a misalignment between the drug names used by substance use researchers and those used by adult drug users. As such, our method may help clinicians better design surveys using colloquial drug names.

Table 8
Drug related words among the 200 word embedding tokens most similar to the seed word or words. The numbers in parentheses represent the cosine similarity between the drug related word on the right and the seed word(s) on the left.

Seed word(s)	Drug-related words from embedding
Fent & fentanyl	heroin (0.95), hcl (0.95), analog (0.9), analogue (0.9), opioid (0.89), fhcl (0.86), analogous (0.85), acetyl (0.85), carfentanil (0.85), tetrahydrofuranfentanyl (0.85), maf (0.84), herion (0.84), morphine (0.84), mr-2096 (0.83), 4phenylfent (0.83), carfent (0.83), butyr (0.83), h (0.82), opiate (0.82), furanyl (0.82), actyl (0.81), butry (0.81), acroyl (0.81), acryl (0.81), citrate (0.81), hcls (0.81), ope (0.8), carf (0.8), r-30490 (0.8), fenta (0.8), carfentinil (0.8), spice (0.8), lofentanil (0.8), fentas (0.8), 4fib (0.79), junk (0.79), tranq (0.79), butyryl (0.79), drug (0.78), dirt (0.78), act (0.78), r30 (0.78)
Butyr	furanyl (0.99), acryl (0.99), citrate (0.99), cyclopropyl (0.99), iso (0.99), fuf (0.98), acetyl (0.98), butyryl (0.98), morphine (0.98), actyl (0.98), 4-furanylbutyrfentanyl (0.98), u47700 (0.98), isobutyrfentanyl (0.98), butyrfentanyl (0.98), 3-methyl (0.98), butry (0.98), 4-anpp (0.98), benzyl (0.97), 4-meo (0.97), thff (0.97), fenta (0.97), acrylfentanyl (0.97), para (0.97), 3-mf (0.97), acetylfentanyl (0.97), snow (0.97), bf (0.97), dxm (0.97), 3-mff (0.97), 2memaf (0.97), fentanil (0.97), isotonitazene (0.97), analogous (0.97), cocaine (0.97), acetamide (0.97), methoxyacetylfentanyl (0.97), flouro (0.97), sedative (0.97), noid (0.97), fluoro (0.97), thiofentanyl (0.97), 3fuf (0.97), 3-me (0.97), 3-meo (0.97), methylfentanyl (0.96), diamorphine (0.96), benzylfentanyl (0.96), logue (0.96), salvia (0.96), benzoyl (0.96), alprazolam (0.96)

An analysis of the misclassified posts in Table 9 also points to some interesting insights. We observed that drug color attributes could not be detected by our classifier. This can be attributed to the rarity of color attribute in our annotated dataset and also gives an insight into how people on the r/fentanyl subreddit talk about drug attributes. We also observe that Dilaudid (brand name for opioid analgesic) is missing in our corpus of drug names to debias, which points to the fact that external sources of drug names or brand names for drugs could be used to supplement this corpus to enhance classification performance of risk detection. We also include examples which were low risk but were classified as elevated risk. These two instances point to our intentionally restrictive codebook which emphasizes on factors like regular user or high tolerance and does not annotate instances with only a mention of usage as risk.

4.3. Limitations, conclusions, and future directions

We acknowledge some limitations towards generalizability posed by the focus on a single online community, r/fentanyl. While there are other subreddits on substance misuse that could have been considered in this research, this particular forum allowed us to scope a dataset comprising postings on an opioid frequently misused. Our work also does not consider lurkers on the Reddit platform — individuals who browse and consume content but do not post; in fact, it is noted that for most online platforms, a small minority of users generate a majority of the content (Van Mierlo, 2014). In the light of these issues, we caution against drawing generalizable conclusions about population-level trends on fentanyl misuse behaviors and risk factors beyond the one studied.

A second limitation of our study is that our codebook and risk classifiers label the risk of individual posts or comments based solely on their text versus exploring self-reported user level risk. In addition, our codebook and classifiers can struggle to classify the risk of comments which are ambiguous out of the context of their parent post and surrounding comments. Also, our list of drug names for debiasing does not include every possible drug name, so the comment "Dilaudid only seems to give me a rush if it's my first shot of the day. Weird. Fu-f is water soluble though?" was likely misclassified because the brand name opiate "Dilaudid" was not in our drug name corpus. Furthermore, our codebook was developed using both inductive (e.g., reviewing r/fentanyl approaches) and deductive approaches (e.g., referencing past literature on opioid/fentanyl risks) to be used with posts and comments on r/fentanyl, so it may not transfer to more general online communities related to opioids. Accordingly, future work could explore user-level opioid use risk based on a user's entire posting and commenting history. Furthermore, future studies can explore trends in the occurrences of community specific drug names over time to understand the rise and fall in popularity of specific fentanyl analogues. On a related note, although we did not apply our risk classifier to automatically label the unlabeled posts

Table 9
Misclassified examples.

Type	Example
Elevated risk misclassified as low risk	Orange is not legitimate , you 've been scammed.
	The color is interesting, looks oxidized. I would love to see this analyzed, as I've only seen white or waxy yellow CF.
	How can I get the nausea and vomiting under control?
	Dilaudid only seems to give me a rush if it's my first shot of the day. Weird. Fu-f is water soluble though ?
Low risk misclassified as elevated risk	true pure MAF is stronger than plain ole fent. (in my experience it is at least 2x potency when pure , as my dose was effectively halved going from fent to maf) however since maf is no longer being synthesized what remains is degrading and probably cut to hell limited stocks .
	I know that 3 to 7 hours is good legs trust me that fluroisobytylfentanyl I had to dose every hour !

and comments in the r/fentanyl subreddit, future research could do so to examine the prevalence of risk in different discussions pertaining to fentanyl and its analogues, as well as study how they evolve over time. It would also be worthwhile to harness similar accessible and secure technology over a larger sample to gather additional rich qualitative data and expand upon both the population and its specific members' unique needs. This will ensure tailored, efficient measures are targeted and delivered to those individuals most in need.

In using an interdisciplinary approach including machine learning techniques and clinical human coding posts/comments regarding the misuse of fentanyl and its analogues, our team was able to automatically detect risk and identify users who may benefit from substance use support and intervention. This work improves upon our understanding of substance misuse risk factors and furthers our ability to identify such risk concepts among underrepresented populations to inform outreach and intervention strategies tailored to this at-risk group. The findings in this study will not only help to create novel, efficient methods to successfully identify those at high risk for fentanyl misuse, but may also inform future studies aiming to develop and adapt similar models to facilitate timely detection of other substance use and mental health risk factors.

Declaration of competing interest

The authors have no conflicts of interest to report.

Acknowledgements

Funding for this work was provided by the National Institutes of Health (NIH) [Grant No: K02 DA043657 (Dr. Cavazos-Rehg) and Grant No: R01MH117172 (Dr. De Choudhury)], and through a postdoctoral fellowship to Dr. Aledavood from the James S. McDonnell Foundation. We would also like to acknowledge Vivian Agbonavbare and Nnenna Anako for their work to manually code posts and comments for this study.

References

Acton, J.P., 1973. Evaluating Public Programs to Save Lives.
Alambo, A., Padhee, S., Banerjee, T., Thirunarayan, K., 2021, January. Covid-19 and mental health/substance use disorders on reddit: a longitudinal study. In: International Conference on Pattern Recognition. Springer, Cham, pp. 20–27.
Amlani, A., McKee, G., Khamis, N., Raghukumar, G., Tsang, E., Buxton, J.A., 2015. Why the FUSS (Fentanyl Urine Screen Study)? A cross-sectional survey to characterize an emerging threat to people who use drugs in British Columbia, Canada. Harm Reduct. J. 12 (1), 54.
Ammari, Tawfiq, Schoenebeck, Sarita, Romero, Daniel, 2019. Self-declared throwaway accounts on Reddit: how platform affordances and shared norms enable parenting disclosure and support. In: Proceedings of the ACM on Human-Computer Interaction 3, No. CSCW, pp. 1–30.
Andalibi, N., Haimson, O.L., De Choudhury, M., Forte, A., 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In: Paper Presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.
Andalibi, N., Ozturk, P., Forte, A., 2017. Sensitive self-disclosures, responses, and social support on Instagram: the case of #depression. In: Paper Presented at the Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.
Balsamo, D., Bajardi, P., Panisson, A., 2019. Firsthand opiates abuse on social media: monitoring geospatial patterns of interest through a digital cohort. In: The World Wide Web Conference, pp. 2572–2579.
Barash, J.A., Somerville, N., DeMaria Jr., A., 2017. Cluster of an unusual amnestic syndrome—Massachusetts, 2012–2016. MMWR. Morbid. Mortal. Wkl. Rep. 66 (3), 76.
Betton, V., Borschmann, R., Docherty, M., Coleman, S., Brown, M., Henderson, C., 2015. The role of social media in reducing stigma and discrimination. Br. J. Psychiatry 206 (6), 443–444.
Betts, K.S., McIlwraith, F., Dietze, P., Whittaker, E., Burns, L., Cogger, S., Alati, R., 2015. Can differences in the type, nature or amount of polysubstance use explain the increased risk of non-fatal overdose among psychologically distressed people who inject drugs? Drug Alcohol Depend. 154, 76–84.
BigQuery, 2019. Google BigQuery. Retrieved from. https://bigquery.cloud.google.com/dataset/fh-bigquery:reddit_comments.
Birnholtz, J., Merola, N.A.R., Paul, A., 2015. "Is it weird to still be a virgin" anonymous, locally targeted questions on Facebook confession boards. In: Paper Presented at the Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.
Bluthenthal, R.N., Simpson, K., Ceasar, R.C., Zhao, J., Wenger, L., Kral, A.H., 2020. Opioid withdrawal symptoms, frequency, and pain characteristics as correlates of health risk among people who inject drugs. Drug Alcohol Depend. 107932.
Boe, B., 2016. Python reddit api wrapper (PRAW). https://praw.readthedocs.io/en/latest/.
Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Britton, P.C., Wines Jr., J.D., Conner, K.R., 2010. Non-fatal overdose in the 12 months following treatment for substance use disorders. Drug Alcohol Depend. 107 (1), 51–55.
Buresh, M., Genberg, B.L., Astemborski, J., Kirk, G.D., Mehta, S.H., 2019. Recent fentanyl use among people who inject drugs: results from a rapid assessment in Baltimore, Maryland. Int. J. Drug Policy 74, 41–46.
Castelvecchi, D., 2016. Can we open the black box of AI? Nat. News 538 (7623), 20.
Cavazos-Rehg, P., Grucza, R., Krauss, M.J., Smarsh, A., Anako, N., Kasson, E., Bierut, L.J., 2019. Utilizing social media to explore overdose and HIV/HCV risk behaviors among current opioid misusers. Drug Alcohol Depend. 205, 107690.
Cavazos-Rehg, P., Xu, C., Krauss, M.J., Min, C., Winograd, R., Grucza, R., Bierut, L.J., 2021. Understanding barriers to treatment among individuals not engaged in treatment who misuse opioids: a structural equation modeling approach. In: Substance Abuse, pp. 1–20.
Cavazos-Rehg, P.A., Krauss, M.J., Sowles, S.J., Connolly, S., Rosas, C., Bharadwaj, M., Bierut, L.J., 2017. An analysis of depression, self-harm, and suicidal ideation content on Tumblr. Crisis 38 (1), 44.
Centers for Disease Control and Prevention (CDC), 2018. Drug overdose deaths in the United States, 1999–2018. Retrieved from. https://www.cdc.gov/nchs/products/databriefs/db356.htm.
Centers for Disease Control and Prevention (CDC), 2019. Opioid overdose: fentanyl. Retrieved from. https://www.cdc.gov/drugoverdose/opioids/fentanyl.html.
Chancellor, S., Pater, J.A., Clear, T., Gilbert, E., De Choudhury, M., 2016. #thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pp. 1201–1213.
Chancellor, S., Nitzburg, G., Hu, A., Zampieri, F., De Choudhury, M., 2019. Discovering alternative treatments for opioid use recovery using social media. In: Paper Presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.
Chandrasekharan, E., Gandhi, C., Mustelier, M.W., Gilbert, E., 2019. Crossmod: a cross-community learning-based system to assist reddit moderators. In: Proceedings of the ACM on Human–Computer Interaction, 3(CSCW), pp. 1–30.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.
Cialdini, R.B., 1987. Influence, Vol. 3. A. Michel, Port Harcourt.
Coffin, P.O., Galea, S., Ahern, J., Leon, A.C., Vlahov, D., Tardiff, K., 2003. Opiates, cocaine and alcohol combinations in accidental drug overdose deaths in New York City, 1990–98. Addiction 98 (6), 739–747.
Constine, J., 2017. Facebook rolls out AI to detect suicidal posts before they're reported. Zugang. https://techcrunch.com/2017/11/27/facebook-ai-suicide-prevention/(15.12.2019).
Coppersmith, G., Leary, R., Crutchley, P., Fine, A., 2018. Natural language processing of social media as screening for suicide risk. Biomed. Inform. Insights 10, 1178222618792860.
Coyne, J., Downey, G., 1991. Social factors and psychopathology: stress, social support, and coping processes. Ann. Rev. Psychol. 42 (1), 401–425.
Darke, S., Hall, W., 2003. Heroin overdose: research and evidence-based intervention. J. Urban Health 80 (2), 189–200.
De Choudhury, M., Kiciman, E., 2017. The language of social support in social media and its effect on suicidal ideation risk. In: Paper Presented at the Proceedings of the International AAAI Conference on Weblogs and Social Media.
De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., Kumar, M., 2016. Discovering shifts to suicidal ideation from mental health content in social media. In: Paper Presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.
De Choudhury, Munmun, Sushovan, De, 2014. Mental health discourse on Reddit: self-disclosure, social support, and anonymity. Proceedings of the International AAAI Conference on Web and Social Media, vol. 8 no. 1.
Degenhardt, L., Mathers, B., Vickerman, P., Rhodes, T., Latkin, C., Hickman, M., 2010. Prevention of HIV infection for people who inject drugs: why individual, structural, and combination approaches are needed. Lancet 376 (9737), 285–301.
Devlin, J., Chang, M., Lee, K., Toutanova, K., 2018. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.
Drug Enforcement Administration, 2015. National Forensic Laboratory Information System Special Report: Opiates and Related Drugs Reported in NFLIS, 2009–2014.
Drug Enforcement Administration, 2018. Drug Enforcement Administration Emerging Threat Report: Annual 2017.
Eisenberger, N.I., 2012. The pain of social disconnection: examining the shared neural underpinnings of physical and social pain. Nat. Rev. Neurosci. 13 (6), 421–434.
Gostin, L.O., Hodge, J.G., Noe, S.A., 2017. Reframing the opioid epidemic as a national emergency. JAMA 318 (16), 1539–1540.

Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H., Eichstaedt, J.C., 2017. Detecting depression and mental illness on social media: an integrative review. Curr. Opin. Behav. Sci. 18, 43–49.
Hedegaard, H., Bastian, B.A., Trinidad, J.P., Spencer, M., Warner, M., 2018. Drugs most Frequently Involved in Drug Overdose Deaths: United States, 2011–2016.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Howard, D., Maslej, M.M., Lee, J., Ritchie, J., Woollard, G., French, L., 2020, May. Transfer learning for risk classification of social media posts: model evaluation study. J. Med. Internet Res. 22 (5), e15371.
Jhaver, S., Birman, I., Gilbert, E., Bruckman, A., 2019. Human-machine collaboration for content regulation: the case of Reddit automoderator. ACM Trans. Comput.-Human Interact. (TOCHI) 26 (5), 1–35.
Jones, C.M., Einstein, E.B., Compton, W.M., 2018. Changes in synthetic opioid involvement in drug overdose deaths in the United States, 2010-2016. Jama 319 (17), 1819–1821.
Kenney, S.R., Anderson, B.J., Conti, M.T., Bailey, G.L., Stein, M.D., 2018. Expected and actual fentanyl exposure among persons seeking opioid withdrawal management. J. Subst. Abus. Treat. 86, 65–69.
Klein, A., Sarker, A., Rouhizadeh, M., O'Connor, K., Gonzalez, G., 2017. Detecting personal medication intake in twitter: an annotated corpus and baseline classification system. In: BioNLP 2017, Vancouver, Canada. Association for Computational Linguistics, pp. 136–142 (August).
Lakoff, G., 1975. Hedges: a study in meaning criteria and the logic of fuzzy concepts. In: Contemporary Research in Philosophical Logic and Linguistic Semantics. Springer, Dordrecht, pp. 221–271.
Landis, J.R., Koch, G.G., 1975. A review of statistical methods in the analysis of data arising from observer reliability studies (Part I). Stat. Neerland. 29 (3), 101–123.
LaRue, L., Twillman, R.K., Dawson, E., Whitley, P., Frasco, M.A., Huskey, A., Guevara, M. G., 2019. Rate of fentanyl positivity among urine drug test results positive for cocaine or methamphetamine. JAMA Netw. Open 2 (4), e192851.
Latkin, C.A., Sherman, S., Knowlton, A., 2003. HIV prevention among drug users: outcome of a network-oriented peer outreach intervention. Health Psychol. 22 (4), 332.
Lee, J.P., Antin, T.M., 2012. How do researchers categorize drugs, and how do drug users categorize them? Contemp. Drug Probl. 38 (3), 387–427.
Livingston, J.D., Milne, T., Fang, M.L., Amari, E., 2012. The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. Addiction 107 (1), 39–50.
Lu, J., Sridhar, S., Pandey, R., Hasan, M.A., Mohler, G., 2019. Redditors in recovery: text mining reddit to investigate transitions into drug addiction. arXiv preprint arXiv:1903.04081.
Matias, J.N., 2019. Preventing harassment and increasing group participation through social norms in 2,190 online science discussions. Proc. Natl. Acad. Sci. 116 (20), 9785–9789.
McHugh, M.L., 2012. Interrater reliability: the kappa statistic. Biochem. Med. 22 (3), 276–282.
Miller, P.G., Sønderlund, A.L., 2010. Using the internet to research hidden populations of illicit drug users: a review. Addiction 105 (9), 1557–1567.
Naslund, J.A., Aschbrenner, K.A., Marsch, L.A., Bartels, S.J., 2016. The future of mental health care: peer-to-peer support and social media. Epidemiol. Psychiat. Sci. 25 (2), 113–122.
National Institute on Drug Abuse (NIDA), 2019. Fentanyl Drug Facts. National Institutes of Health. Retrieved from. https://www.drugabuse.gov/publications/drugfacts/fentanyl.
Nelson, L.S., Perrone, J., 2012. Curbing the opioid epidemic in the United States: the risk evaluation and mitigation strategy (REMS). JAMA 308 (5), 457–458.
Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24 (12), 1565–1567.
O'Leary-Kelly, S.W., Vokurka, R.J., 1998. The empirical assessment of construct validity. J. Oper. Manag. 16 (4), 387–405.
Park, A., Conway, M., 2018. Opioid surveillance using social media: how urls are shared among reddit members. Onl. J. Publ. Health Inform. 10 (1).
Paul, M.J., Dredze, M., 2017. Social monitoring for public health. Synth. Lect. Inform. Concepts Retriev. Serv. 9 (5), 1–183.
Paulozzi, L.J., 2012. Prescription drug overdoses: a review. J. Saf. Res. 43 (4), 283–289.
Pavalanathan, U., De Choudhury, M., 2015. Identity management and mental health discourse in social media. In: Proceedings of the 24th International Conference on World Wide Web, pp. 315–321 (May).
Perez, S., 2020. Reddit partners and integrates with mental health service crisis text line. https://rb.gy/hvjw2u. Mar.
Perrin, A., Anderson, M., 2019. Share of US Adults Using Social Media, Including Facebook, Is Mostly Unchanged Since 2018. Pew Research Center, p. 10.
Robinson, P., Turk, D., Jilka, S., Cella, M., 2019. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc. Psychiatry Psychiatr. Epidemiol. 54 (1), 51–58.
Saha, K., Kim, S.C., Reddy, M.D., Carter, A.J., Sharma, E., Haimson, O.L., De Choudhury, M., 2019. The language of lgbtq+ minority stress experiences on social media. In: Proceedings of the ACM on Human-Computer Interaction 3 (CSCW), pp. 1–22.
Saha, K., Ernala, S.K., Dutta, S., Sharma, E., De Choudhury, M., 2020. Understanding moderation in online mental health communities. In: International Conference on Human-Computer Interaction, July. Springer, Cham, pp. 87–107.
Sarker, A., Gonzalez-Hernandez, G., Ruan, Y., Perrone, J., 2019. Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter. JAMA Netw. Open 2 (11), e1914672.
Singer, P., Flöck, F., Meinhart, C., Zeitfogel, E., Strohmaier, M., 2014. Evolution of Reddit: from the front page of the internet to a self-referential community? Paper Presented at the Proceedings of the 23rd International Conference on World Wide Web.
Smolina, K., Crabtree, A., Chong, M., Park, M., Mill, C., Zhao, B., Schütz, C.G., 2020. Prescription-related risk factors for opioid-related overdoses in the era of fentanyl contamination of illicit drug supply: a retrospective case-control study. In: Substance Abuse, pp. 1–7.
Spencer, M., Warner, M., Bastian, B.A., Trinidad, J.P., Hedegaard, H., 2019. Drug Overdose Deaths Involving Fentanyl, 2011–2016.
Springer, Y., Gladden, R., O'Donnell, J., Seth, P., 2019. Notes from the field: fentanyl drug submissions-United States, 2010–2017. MMWR Morb. Mortal. Wkly Rep. 68 (2), 41–43.
Sullivan, M.D., Ballantyne, J.C., 2021. When physical and social pain coexist: insights into opioid therapy. Ann. Fam. Med. 19 (1), 79–82.
Syed, M., Nelson, S.C., 2015. Guidelines for establishing reliability when coding narrative data. Emerg. Adult. 3 (6), 375–387.
Van Mierlo, T., 2014. The 1% rule in four digital health social networks: an observational study. J. Med. Internet Res. 16 (2), e33.
Wadden, D., August, T., Li, Q., Althoff, T., 2021. The effect of moderation on online mental health conversations. In: Proceedings of the International AAAI Conference on Web and Social Media, May, Vol. 15, pp. 751–763.
Webster, L.R., 2017. Risk factors for opioid-use disorder and overdose. Anesth. Analg. 125 (5), 1741–1748.
Wejnert, C., Heckathorn, D., 2012. Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research. In: Sage Library of Research Methods: SAGE Internet Research Methods, v2-311.
Yang, D., Yao, Z., Seering, J., Kraut, R., 2019. The channel matters: self-disclosure, reciprocity and social support in online cancer support groups. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, May, pp. 1–15.
Yang, D., Kraut, R.E., Smith, T., Mayfield, E., Jurafsky, D., 2019a. Seekers, providers, welcomers, and storytellers: modeling social roles in online health communities. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, May, pp. 1–14.
Zibbell, J., Howard, J., Clarke, S., Ferrell, A., Karon, S., 2019. Non-Fatal Opioid Overdose and Associated Health Outcomes: Final Summary Report. Office of the Assistant Secretary for Planning and Evaluation.