Identifying and Characterizing Opioid Addiction States Using Social Media Posts

Deeptanshu Jha[1] Samantha R. La Marca[1], and Rahul Singh[1, 2,†]

[1]Department of Computer Science, San Francisco State University, and [2]Center for Discovery and Innovation in Parasitic Diseases, University of California, San Diego

***Abstract*—Opioid addiction constitutes a significant contemporary health crisis that is multifarious in its complexity. Modeling the epidemiology of any addiction is challenging in its own right. For opioid addiction, the challenge is exacerbated due to the difficulties in collecting real-time data and the circumscribed nature of information opioid users may disclose owing to stigma associated with prescription misuse. Given this context, identifying the progression of individuals through the stages of (opioid) addiction is one of the more acute problems in epidemiological modeling whose solution is crucial for designing specific interventions at both personal and population levels. We describe a computational approach for determining and characterizing addiction stages of opioid users from their social media posts. The proposed approach combines recurrent neural network learning with information-theoretic analysis of word-associations and context-based word embedding to determine addiction stage-specific language usage. Users who have a high likelihood for relapsing back to drug-use are identified and characterized using propensity score matching and logistic regression. Experimental evaluations indicate that the proposed approach can distinguish between various addiction stages and identify users prone to relapse with high accuracy as evidenced by F1 scores of 0.88 and 0.79 respectively.**

Keywords—Computational epidemiology, opioid addiction, addiction state analysis, social media

I. INTRODUCTION

Opioids are powerful pain killers that interact with the opioid pain receptors on the nerve cells in the brain and possess a high potential for inducing addiction. Misuse of opioid pain relievers is one of the gravest public health crises of our time, with oxycodone, hydrocodone, morphine, and codeine being the most commonly misused opioids [1. The complexity of dealing with opioid addiction is exacerbated by a number of factors among the most significant of which are its insidious link with pain relief in palliative contexts and social stigma - which impairs disclosure and data collection.

Development of opioid addiction is multi-stage [2] and influenced by the context of each individual user. Typically, process begins with prescriptive use or due to experimentation with opioid drugs. Early misuse does not necessarily lead to addiction as some users may stop. Regular use however, leads to compulsive and uncontrollable opiate craving as well as increased tolerance to the drug. When a user ceases intake, they typically pass through the following stages (Murthy 2017): (1) *withdrawal***:** characterized by a combination of physical and emotional symptoms. Early symptoms start within 6-12 (30) hours for short (long)-acting opioids. These symptoms include muscle aches, agitation, anxiety, hypertension, increased heart rate, and trouble falling and staying asleep. Late symptoms include nausea, vomiting, sweating, diarrhea, and cramps [3]. (2) *Recovery*: recovery from opioid addiction is a long-term process requiring continuous effort and diligence. The recovery process is typically associated with treatment regimens and accompanied by indications that the substance user is attempting to achieve better health, lifestyle, and purpose. (3) *Relapse*: this stage is characterized by the return of an individual to using drugs after a period of sobriety.

Recognizing the aforementioned addiction states is critical to intervention design and treatment of individuals. Furthermore, determining the distribution of a population in terms of these stages can be helpful in population-level modeling as well as for planning large-scale interventions, and allocation of resources. Currently, such determinations depend on data collected using surveys or obtained from clinical records. However, a number of factors may limit the accuracy of such determinations; most significantly perhaps, the stigma associated with substance misuse impacts the accuracy of surveys due to underreporting [4] and prevents people from seeking medical assistance. Since physicians are the primary source of prescription medications, opioid users have been found to be inhibited in discussing substance abuse problems with health practitioners [5]. Finally, data collected via surveys as well as patient records tends to lag the true temporal-state of the real-world epidemic.

Social media constitutes a novel source of addiction related information, the leveraging of which may ameliorate many of the aforementioned challenges. In particular, the anonymity offered by social media platforms engenders candid conversations and disclosures. Furthermore, social media posts occur in real-time and can be analyzed rapidly to remove the time-lag inherent when information is sought from traditional sources such as surveys and patient records. Consequently, the design of methods for using of social media-based information in epidemiology has garnered recent interest.

A. Problem formulation and data collection

Given the social media posts of a person engaged in opioid misuse, we seek to identify and subsequently characterize the addiction stages: *use*, *withdrawal*, *recovery*, and *relapse* as defined above (for clarity these stages are italicized and underlined hereafter) from the social media posts of users. We formulate this problem as that of sequence labeling: given the temporally ordered sequence of social media posts $(p_1(t),$

†Corresponding author email: rahul@sfsu.edu

978-1-6654-0126-5/21/$31.00 ©2021 IEEE

$p_2(t),..., p_n(t)$) from a opioid user, our goal is to determine a corresponding sequence of labels $l = (l_1(t), l_2(t), ..., l_n(t))$, where $l_i \in \{use, withdrawal, and, recovery\}$. The state relapse can then be identified as the transition from: (1) recovery (or withdrawal) to use or (2) the transition from recovery to withdrawal. In addition to solving the labeling problem, we present results that show how through predictive modeling it is possible to identify users who have a high likelihood of relapsing into drug use.

Our research is based on publicly available data from the social media forum Reddit. We used data from two sub-forums: r/opiates (a subreddit for opiates users) and r/OpiatesRecovery (a subreddit where members help each other to give-up opioids). To ensure privacy, all data used by us are anonymized and the contents of the posts paraphrased when reported. Two sets of data were collected by us; the first (hereafter, the DRUG-DATASET) consisted of 6,656,671 posts made by 1,278,603 unique redditors in 117 drug use subreddits. The second dataset contained 35,868 unique redditors who had posted in either of the r/opiates or r/OpiatesRecovery subreddits. We queried their entire post history of these redditors using Reddit's open ReST API PRAW [6]. Subsequently, we selected 155 redditors from this data set who had together made 2,587 posts in either r/opiates or r/OpiatesRecovery. These posts were independently analyzed by two researchers and annotated with the labels use, withdrawal, and recovery defined analogously to [7]. The annotation was done both in a context-sensitive manner (*i.e.* by taking into account the sequence of posts made by a user) as well as in a context-independent manner. The former annotation is henceforth called context sensitive annotation (CSA) and the latter is called context independent annotation (CIA).

B. Overview of the approach

Our approach has the following steps: (1) *Annotation of Reddit posts to identify stages of drug use.* In this first step, the drug use and recovery process of 155 users described via 2,587 posts in the subreddits: 'opiates' and 'OpiatesRecovery' was mapped to the stages use, withdrawal, and recovery. The annotated data generated in this supervised labeling step was subsequently used for model construction and evaluation. (2) *Inferring the stage-specific vocabulary.* We use a measure of association called point-wise mutual information (PMI) to identify stage-specific *n*-grams (n=1, 2) to characterize and represent the posts in terms of their linguistic content. (3) *Creation of document vectors to represent the posts and comments in the drug user's post history.* A vector representation of the *n*-grams comprising each post is computed using word embedding. Subsequently, each post is numerically summarized by a vector representing the average of its constituent *n*-grams. (4) *Incorporation of recovery-specific information.* The summary representation of a post is augmented using two statistics (when present): the number of days a user had been clean (*i.e.* not used opiates) and the count of symptoms and therapies mentioned in a post. These features are identified using regular expressions and string matching. (5) *Predictive modeling and addiction stage identification*: A recurrent neural network is used in conjunction with conditional random fields to identify the stages of drug use. This learning architecture allows us to capture the temporal context(s) of each post by incorporating information from both the preceding and subsequent posts. (6) *Identifying users who might relapse.* We use a statistical matching technique called nearest neighbor propensity score matching (PSM) to identify *n*-grams (n = 1,2) that characterized relapse. These *n*-grams are then used to train a logistic regression-based classifier to identify users who were in recovery but exhibited language use that heightened their possibility to relapse.

C. Prior work and contributions of the proposed research

Among early works, in [8-10] manual analysis was employed to qualitatively assess themes and attributes underlying opioid misuse based on data from the social media forums, Reddit, Twitter and Instagram. In one of the early algorithmic approaches [11] classification algorithms were used to automatically identify prescription drug abuse tweets. Rather than using the content of posts in [12] activity in subreddits that were reflective of users interests in issues such as as mental health, relationships, and parenting were used to identify drug users open to addiction-recovery interventions. More recently, supervised learning and term embeddings were used in [13] to identify clinically unverified treatments adopted by the drug users to stave off withdrawal effects. Finally, the SMARTS software [14], can identify individuals open to recovery interventions based on their social media posts. None of these works however, have investigated the problem of identifying addiction states. The most significant work in this context till date has been [7]. In it, the process of opioid use and recovery was divided into three stages: *using*, *withdrawing*, and *recovery* based on data from the MedHelp's addiction recovery community, Forum 77. Subsequently, activity features (such as the number of posts authored, number of comments received) and linguistic features such as the frequency of phase-specific terms and LIWC categories were used to train a conditional random field model to classify stages of opioid use.

Our paper makes the following advances to the state-of-the-art. *First*, we use dense word-embeddings in conjunction with a bidirectional long short term memory recurrent neural network coupled with a CRF to model addiction states. This learning method leads to higher accuracy in identifying addiction states when compared to earlier works such as [7]. *Second*, our use of posts from communities involved in *both* recreational opioid use as well as opioid addiction recovery leads to improved linguistic characterization of addiction states. For example, in [7] use is characterized by terms such as "withdrawals", "rehab", "counseling", "treatment", and "stop". These terms arguably, better characterize cessation of drug use. By contrast, our approach identifies terms such as: "shot", "rush", "iv", "plugging", "tolerance", "snort", and "nodding" (a popular slang used to describe being high on opiates), which are all manifestly more indicative of the recreational drug use. *Third*, our method can not only identify relapse, but also, due to its linguistic characterization, is able to predictively identify users who may relapse.

II. METHOD

A. Characterizing state-specific vocabulary

We used information-theoretic measures to estimate term specific associations by analyzing the frequency of a term in a specific context to its frequency in a generic background distribution. To capture the linguistic stage-specificity, we identified *n*-grams (*n*=1, 2) in the posts and utilized point-wise mutual information (PMI) to determine the state specificity of these *n*-grams. Given a state *s* of drug use and an *n*-gram *g*, the PMI is defined as the ratio of the probability of observing *s* and *g* together (*i.e.* their joint probability) to the probabilities of observing *s* and *g* independently. We normalized the PMI measure in the range [-1, +1] as depicted in Eq. (1). Consequently, the value of -1 denoted that *s* and *g* do not co-occur, zero denoted their independence, and 1 denoted complete co-occurrence.

$$PMI = \ln\left(\frac{p(s,n)}{p(s)\times p(n)}\right)/-\ln(p(s,n)) \quad (1)$$

We computed the PMI values for posts labelled via CIA. Of the 1,870 posts that were thus labelled, *use*, *withdrawal*, and *recovery* stages accounted for 762, 778, and 330 posts respectively leading to *p(use)*, *p(withdrawal)*, and *p(recovery)* values of 0.40, .41, and 0.17 respectively. Next, the PMI value for every *n*-gram belonging to posts in each of the *use*, *withdrawal*, and *recovery* stages were calculated and the 3000 highest (by PMI value) *n*-grams for each stage were used to represent their respective vocabulary. That is, posts in a stage were expunged of *n*-grams that did not occur in the vocabulary of that stage.

The skip-gram word2vec technique [15], was next used to generate a 50-dimensional embeddings of all the *n*-grams from 6,656,671 posts in the DRUG-DATASET. Given a text corpus for training, word2vec generates embeddings where terms sharing common contexts in the corpus are located close to each other. Once the embedding was computed, each post was represented by a vector (in this 50-dimensional space) which was computed by taking the dimension-wise average of the vectors of its constituent *n*-grams. In the *withdrawal* and *recovery* stages, we observed that redditors sometimes mentioned the number of days they had refrained from using opioids. This indicator variable is hereafter called *days-clean*. It was identified using regular expressions applied to each post including title and self-comments and used as an additional feature to characterize addiction behavior. Of the 125,479 posts in our Opioid_Dataset, we identified 3,607 posts to have this information. If, for a post, there was no data on the number of *days-clean*, a value of zero was assigned to this feature. The final elements of the feature vector for a post consisted of the number of withdrawal symptoms and chemical- or natural product-therapies (used for mitigating such symptoms) mentioned in the post.

B. Addiction state modeling

We used a combination of bidirectional Long-Short-Term-Memory networks (Bi-LSTMN) and conditional random fields (CRF) to model the drug addiction stage(s) for a given post sequence of a drug user; a diagrammatic representation of the learning architecture can be found in Figure 1). The Bi-LSTMN was used to generate a context sensitive intermediate representation for each user post. This intermediate representation was then fed to the CRF to predict the addiction stage-label for the post.

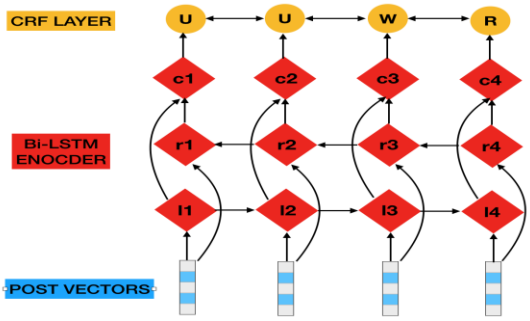

Figure 1. The Bi-LSTM-CRF architecture for addiction-stage modeling. Each post vector is fed to the bidirectional LSTM layer where l_i (r_i) represents the post in terms of its left (right) context. Concatenation of these vectors yields a representation of the post in its context c_i which is used as input for the CRF layer.

To describe the model construction, we note that the sequence of posts made by a user is assumed to describe their addiction journey. Recurrent neural networks (RNNs) constitute one of the artificial neural network frameworks for modeling sequential information; RNNs take a sequence *P* of (posts) vectors ($p_1, p_2, \ldots, p_n$) and return another sequence ($h_1, h_2, \ldots, h_n$) that encodes information for every time step in *P*. However, RNNs are known to perform poorly for long sequences [16]. Long-Short-Term-Memory networks (LSTMNs) represent an RNN architecture which is used for alleviating this issue. Each LSTMN unit consists of a memory cell along with input, output, and forget gates. The memory cell stores information over arbitrary time periods and the gates control the information flow between cells [17]. Given the input sequence *P* of *n* posts, the computation of the LSTMN is guided by Eqs. (2)-(7).

$$i_n = \sigma(W_i h_{n-1} + U_i p_n + b_i) \quad (2)$$
$$f_n = \sigma(W_f h_{n-1} + U_f p_n + b_f \quad (3)$$
$$\tilde{c}_n = \tanh(W_c h_{n-1} + U_c p_n + b_c) \quad (4)$$
$$c_n = f_n \bullet c_{n-1} + i_n \bullet \tilde{c}_n \quad (5)$$
$$o_n = \sigma(W_o h_{n-1} + U_o p_n + b_o) \quad (6)$$
$$h_n = o_n \bullet \tanh(c_n) \quad (7)$$

In the above equations, for the n^{th} post p_n, the input gate, forget gate, and output gate activation vectors are denoted by i_n, f_n, an o_n, respectively while c_n denotes the corresponding cell activation vector and h_n denotes the hidden vector. Further, σ is the logistic sigmoid function and • denotes the Hadamard product. The matrices U_q and W_q denote the weights of the input and recurrent connections with the subscript denoting an input gate (*i*), output gate (*o*), forget gate (f), or memory cell (*c*). Similarly, b_q denote the respective bias

vectors. For a post p_n the LSTMN generates a representation by capturing the context of p_n in terms of posts preceding it. Hereafter, we term this the left-context and denote it by $\overrightarrow{h_n}$. In the Bi-LSTMNs used by us, a separate LSTMN is used to read the post sequence in reverse and generate a right context representation $\overleftarrow{h_n}$ for p_n [18]. The final representation for a post was obtained by concatenating the left and right context representations. As we shall demonstrate, such a bidirectional model helped us generate a highly accurate context specific representation of every post in the sequence.

The final component of our learning architecture employs a linear-chain CRF to capture the relations between adjacent stage labels. Its incorporation was motivated by the observation that capturing the correlations between adjacent stages led to improved labeling accuracy in sequence labeling tasks [19]. Given the output from the Bi-LSTMN, the CRF jointly models the probability of the entire sequence of stage labels $l = (l_1, l_2, \ldots l_n)$. If T denotes the set of all possible stage label transitions, then, using a linear-chain CRF model, the conditional probability of the stage labels given the Bi-LSTMN output is given by Eq. (8).

$$P(l|h; W, b) = \frac{\prod_{i=1}^{n} \exp(W'_{l_{i-1},l_i} h + b_{l_{i-1},l_i})}{\sum_{l \in T} \prod_{i=1}^{n} \exp(W'_{l_{i-1},l_i} h + b_{l_{i-1},l_i})} \quad (8)$$

In Eq.(8), W and b are weight matrices and their subscripts indicate the weight vector for the label (l_{i-1}, l_i). We used maximum likelihood estimation to train the CRF layer and for a training dataset $\{(h_i, l_i)\}$ the final log-likelihood was:

$$L(W,b) = \sum_{(h_i,l_i)} \log P(l_i \mid h_i; W, b) \quad (9)$$

Finally, the Viterbi algorithm was used to generate the optimal labeled-sequence l^*:

$$l^* = \underset{l \in T}{\operatorname{argmax}} P(l \mid h; W, b) \quad (10)$$

To predict if a recovering drug user may relapse, the labels from the Bi-LSTM-CRF model were used to create a data set of redditors who had at least one post from *withdrawal* or *recovery*. This set (hereafter called the Recovery_Dataset) consisted of 3,082 redditors. A user was determined to have relapsed if the following state transitions occurred: (1) *withdrawal* to *use*, (2) *recovery* to *use*, or (3) *recovery* to *withdrawal*. In the aforementioned dataset, 2,179 users were found to have relapsed while 903 users did not. Finally, the terms identified through PSM were used to train a Logistic Regression model to predict the likelihood of relapse. To create a model for predicting the propensity for relapse, we considered the redditors in the Recovery_Dataset as subjects and every *n*-gram present in their posts before they relapsed as treatments. We then used propensity score matching (PSM) [20] to determine the *n*-grams (n=1, 2) that impacted the likelihood for a drug user to relapse.

Addiction State	State-specific Vocabulary
use	vein (0.32), shot(0.31), tar(0.31), shoot (0.30), veins(0.30), rush (0.29), tolerance (0.29), amc(0.29), femoral (0.28), speedball (0.28), dope (0.28), nick (0.28), blood (0.28), cotton (0.28), rig (0.27), powder(0.27), speedballs(0.27), denmark (0.27), snort(0.26)
withdrawal	day (0.26), days (0.25), sleep (0.25), kratom (0.25), taper (0.24),subs (0.24), lope(0.23), symptoms (0.23), rls(0.22), tapering (0.22), withdrawal (0.22), detox (0.22), wd (0.22),quit (0.22), feel (0.21), suboxone (0.21),mg (0.21), hours_sleep (0.21), hours oct (0.21)
recovery	months_clean (0.46), months (0.46), clean (0.46) ,life (0.43), years_clean (0.39),defects (0.38),recovery (0.37), years(0.37), struggling (0.37), clean_today (0. 37), ugly (0.36), wedding (0.36), fiancée (0.36), steps(0.51),year_clean (0.36), meetings (0.36), year (0.36), sober (0.36), son (0. 36), custody (0. 36)

Table 1: Top 20 *n*-grams for *use*, *withdrawal*, and *recovery* with corresponding PMI in brackets.

III. Experiments, Evaluations and Results

A. Identifying the addiction state-specific vocabulary

We begin by presenting the addiction state-specific vocabulary as determined by using PMI (Table 1). From this table, we can observe that the *n*-grams associated with *use* were found to be dominated by terms related to drug administration and post consumption responses. For *withdrawal*, we observed terms describing the symptoms experienced by users during the withdrawal process, the drugs consumed to avoid the withdrawal symptoms, method(s) of withdrawals, and various descriptions associated with the notion underlying the variable *days-clean*. Finally, *recovery* was characterized arguably by the most diverse set of terms among all the three stages. It includes terms that describe the duration users have refrained from drug use. Interestingly, compared to *withdrawal*, the time-frame corresponding to this information is much longer (months versus days). Users were also found to be contemplative and philosophical about life as well as motivated to move on with their lives, exemplified by terms such as "wedding", "fiancée", "sober". Users were also found by us to discussed programs such as Alcoholics Anonymous (AA)/ Narcotics Anonymous (NA) and the 12 steps ideology endorsed by the AA/NA. Finally, some drug users struggled with their recovery and their language-use contained terms like "defects", "struggling", "clean today", and "ugly".

B. Common withdrawal symptoms and alternate therapies

We used the posts predicted as *withdrawal* posts and our list of symptoms and alternative therapies to identify the most common withdrawal symptom and the alternative therapies being mentioned by the drug users on Reddit. Our analysis

showed that the top five symptoms experienced by the drug users in the *withdrawal* state to be: (1) insomnia, (2) diarrhea, (3) restless legs, (4) nausea and (5) body ache. Furthermore, the top five alternative therapies being employed by the drug users in the *withdrawal* state were: (1) kratom, (2) gabapentin, (3) xanax, (4) loperamide, and (5) clonidine.

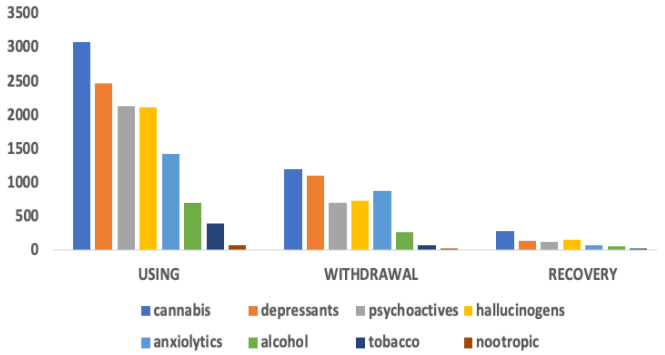

Figure 2. Drug co-ingestion patterns present in different states of substance misuse.

C. Drug co-ingestion patterns

We analyzed the posts from the 5,374 users to study drug co-ingestion patterns present at different stages of the opioid use and recovery. To do so, we created a list of drug terms belonging to the following classes: (1) alcohol, (2) tobacco, (3) depressants, (4) anxiolytics, (5) hallucinogens, (6) psychoactives, and (7) nootropics. We then used a public drug term glossary made available by the DEA to populate these classes. Finally, string matching was employed to identify if a drug from an aforementioned class was present in a post. Figure 2 presents the number of users who were found to co-ingest a particular class of drug with opioids grouped in terms of *use*, *withdrawal*, and *recovery*. As it can be observed from this figure, the number of users who engaged in co-ingestion was highest at *use* stage and progressively decreased at *withdrawal* and *recovery*. The most common co-ingested class of drug for each stage was found to be 'cannabis' followed by 'depressants'.

D. Evaluation of addiction state classification accuracy

In this study we investigated the classification accuracy corresponding to different feature sets and learning methods. We used three different feature sets that are summarized in Table 2. The performance of the Bi-LSTM-CRF model for these three features is shown in Tables 3-5. Our experiments were carried out on two independent test sets randomly sampled at different times. This was done to test the generalizability and applicability of our methods on differing datasets. Test set 1 consists of 995 posts of 75 users and Test Set 2 consists of 744 posts belonging to 54 drug users. Posts for the users in test sets were mapped to *use*, *withdrawal* and *recovery* by the authors and were used as ground truth label to evaluate the model performance. In these experiments, the Bi-LSTM-CRF model with JS1 as features set was found to have the best *F1* score across both the two datasets. In terms of individual classes, we achieved the highest *F1* score for *use*, followed by *withdrawal*, and *recovery*. We postulate that this differentiation was due to the distinct linguistic characteristics of posts in *use* as opposed to *withdrawal* and *recovery*. In particular, *withdrawal* and *recovery* have linguistic characteristics that were less distinct as compared to *use*. Finally, the importance of the variables capturing the duration of *days-clean* can be noted: their incorporation improved the *F1* scores for both data sets.

Feature name	**Feature length**	**Feature composition**
JS1	60	Post vector (50), no of stage specific term present in posts and comments (6), number of alternative therapies and symptoms mentioned (1), days clean (1), a binary variable to indicate if the number of days clean is in days (1), binary variables that indicates indicate if the number of days clean is in days or months
JS2	58	Post vector (50), no of stage specific term present in posts and comments (6), number of alternative therapies and symptoms mentioned (1), and days clean (1). *The reader may note the absence of variables indicating the duration of the variable days-clean*. (c) F77 (feature length 54) – the features used in [7].
F77	54	Features used in [7].

Table 2. Different feature sets utilized to compare the results of the classification algorithms. Number in parentheses represents the size of the features.

Feature Set	**Precision**	**Recall**	**F1**
JS1	0.89, 0.94	0.90,0.91	0.90,0.92
JS2	0.88,0.95	0.91,0.90	0.90,0.92
F77	0.89,0.91	0.81,0.88	0.85,0.89

Table 3. Performance of the proposed Bi-LSTM-CRF method for *use* on two independently collected data sets.

Feature Set	**Precision**	**Recall**	**F1**
JS1	0.88,0.88	0.87,0.88	0.88,0.88
JS2	0.86,0.87	0.85, 0.88	0.85,0.87
F77	0.75,0.85	0.90,0.90	0.82,0.87

Table 4. Performance of the proposed Bi-LSTM-CRF method for *withdrawal* on two independently collected data sets.

Feature Set	**Precision**	**Recall**	**F1**
JS1	0.81,0.80	0.80,0.86	0.81,0.83
JS2	0.78,0.77	0.73,0.86	0.76,0.81
F77	0.83,0.82	0.70,0.76	0.76,0.79

Table 5. Performance of the proposed Bi-LSTM-CRF method for *recovery* on two independently collected data sets.

E. Assessing the predictions for propensity to relapse

In this section we examine the PSM results and the classification accuracy in predicting users who relapse. In Table 6 we report the top 5 *n*-grams and paraphrased example posts with the highest ATE scores that were found to decrease the likelihood of continued abstinence from drugs. We postulate that these tokens reflect the presence of drug yearning in the posts of users who eventually relapse. Terms ,such as, "enjoy high", "just smoke", "hate love" display that users in withdrawal crave drugs and mention of such terms signal towards a future relapse.

In the final experiment, we investigated the effectiveness of a Logistic Regression model (LRM) to classify redditors in terms of their tendency to relapse. We set aside 10% of our user set (3,082) as our held-out validation set (377 redditors). The classifier parameter were tuned using *k*-fold cross-validation ($k = 10$) on the remaining 90% users. The LRM was tested on a dataset of 377 users (containing 290 users who relapsed and 87 users who did not relapse). Table 7 displays the results of the classification; we found that redditors who relapsed were classified with greater accuracy (F1 = 0.80) than redditors who did not relapse (F1 = 0.77). The area under the curve (AUC) for the LRM was 0.79.

Token	ATE	Paraphrased post
enjoy high	-0.87	I used to enjoy being high.
just smoke	-0.8	I just smoke weed now.
blaming	-0.8	Am I blaming my friends for my craving
dose pm	-0.8	My sub dose is at 9 pm.
hate love	-0.8	I hate myself rn but I love dope.

Table 6. Statistically significant ($p < 0.05$) treatment tokens that decreases the likelihood of continued recovery.

	Precision	Recall	F1
Relapse	0.75	0.85	0.80
No Relapse	0.83	0.71	0.77

Table 7. Classification accuracy for the Logistic regression model.

IV. Discussions and Conclusions

In this paper we have described an algorithmic approach to address the fundamental challenges of detecting the different states of drug use and recovery using social media data. The proposed method analyzes the state-specific language use from social media discussions to predict the addiction state(s) for a user. Our approach constitutes a powerful tool for helping focus interventions for specific substance users. If applied at-scale, it can also be used to generate, at population levels, retrospective and predictive state predictions and analyses.

V. Acknowledgements

This research was funded in part by the National Science Foundation through grant IIS 1817239.

REFERENCES

[1] National Institute on Drug Abuse. Commonly abused prescription drugs. 2012

[2] Lyvers, M. Drug addiction as a physical disease: The role of physical dependence and other chronic drug-induced neurophysiological changes in compulsive drug self-administration. *Experimental and clinical psychopharmacology*. 6(1):107., 1998

[3] US Department of Health and Human Services, HHS guide for clinicians on the appropriate dosage reduction or discontinuation of long-term opioid analgesics, 2019

[4] Richter, L *et al,* Stigma and Addiction Treatment. In The Stigma of Addiction, pp. 93-130,. Springer, Cham., 2019

[5] Centers For Disease Control and Prevention. Physicians are a leading source of prescription opioids for the highest-risk users. 2014

[6] The Python Reddit API Wrapper, 2019

[7] MacLean, D. *et al,* Forum77: An analysis of an online health forum dedicated to addiction recovery. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 1511-1526, 2015

[8] Chan, B., Lopez, A. and Sarkar, U., The canary in the coal mine tweets: social media reveals public perceptions of non-medical use of opioids. *PloS one*, *10*(8), 2015

[9] D'Agostino, A.R., Optican, A.R., Sowles, S.J., Krauss, M.J., Lee, K.E. and Cavazos-Rehg, P.A., Social networking online to recover from opioid use disorder: A study of community interactions. *Drug and alcohol dependence*, *181*, pp.5-10, 2017

[10] Cherian, R., Westbrook, M., Ramo, D. and Sarkar, U.,. Representations of codeine misuse on instagram: content analysis. *JMIR public health and surveillance*, *4*(1), p.e22., 2018

[11] Sarker, A. *et al,* Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. Drug safety. 39(3):231-40, 2016

[12] Eshleman, R., Jha D. and Singh, R Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media. In2017 IEEE International Conference on Bioinformatics and Biomedicine, pp. 849-854, 2017.

[13] Chancellor, S. *et al,* Discovering alternative treatments for opioid use recovery using social media. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* pp. 124, 2019

[14] Jha D. and Singh R, SMARTS: the social media-based addiction recovery and intervention targeting server, *Bioinformatics*, Vol. 36(6), pp. 1970-1972, 2020

[15] Mikolov, T. *et al,* Distributed representations of words and phrases and their compositionality. InAdvances in neural information processing systems 2013, pp. 3111-3119, 2013

[16] Bengio, Y. *et al,* Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*. 1994 Mar 2;5(2):157-66, 1994

[17] Hochreiter, S and Schmidhuber, J. Long short-term memory, *Neural Computation*. 9 (8): 1735–1780, 1997

[18] Graves A, and Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks. *Proceedings. IEEE International Joint Conference on Neural Networks*, Vol. 4, pp. 2047-2052, 2005

[19] Zhang, Y., Wang, X., Hou, Z. and Li, J., Clinical named entity recognition from Chinese electronic health records via machine learning methods. *JMIR medical informatics*, *6*(4), p.e50, 2018

[20] Rosenbaum PR, and Rubin DB The central role of the propensity score in observational studies for causal effects. *Biometrika*. 70(1):41-55, 1983