Abstract—Opioid addiction is a severe public health threat in the U.S., causing massive deaths and many social problems. Accurate relapse prediction is of practical importance for recovering patients since relapse prediction promotes timely relapse preventions that help patients stay clean. In this paper, we introduce a Generative Adversarial Networks (GAN) model to predict the addiction relapses based on sentiment images and social influences. Experimental results on real social media data from Reddit.com demonstrate that the GAN model delivers a better performance than comparable alternative techniques. The sentiment images generated by the model show that relapse is closely connected with two emotions ‘joy’ and ‘negative’. This work is one of the first attempts to predict relapses using massive social media data and generative adversarial nets. The proposed method, combined with knowledge of social media mining, has the potential to revolutionize the practice of opioid addiction prevention and treatment.

Keywords—Opioid Addicts Detection, Opioid Relapse Prediction, Generative Adversarial Nets

I. INTRODUCTION

Opioid addiction is a severe public health threat in the U.S., causing massive deaths and many social problems [1]. According to the latest statistics of National Institute on Drug Abuse (NIDA, 2017), more than 115 Americans die after overdosing on opioids on a daily basis, and nearly 64,000 people died of drug overdoses in the US in 2016, the most lethal year of the drug overdose epidemic (NIDA, 2017). Moreover, millions of Americans have been influenced by opioid-related problems. It is estimated that 2.1 million people suffer from substance use disorders related to prescription opioid pain relievers in the United States alone [1], where there has been a significant increase in opioid-related deaths from 2000 to 2017, and the death toll is still rising [2]. The opioid crisis has social impacts beyond the increased death toll. The status of the opioid crisis in the US is shown in Figure 1, other consequences include a rise in the number of infants born dependent on opioids [3], [4] as well as the spread of infectious diseases such as HIV and hepatitis C [5]. Opioid epidemic has now deteriorated into a full-scale national pandemic, leading to national concern because of its negative impacts on health, social security and economics.

Effective ways to mitigate the opioid crisis are thus urgently needed. Most of the previous work in this area has tended to focus on discovering drug-related adverse events [6], for example by proposing platforms to identify prescription drug abuse [7], monitor medication abuse via social media [8], or detect opioid addicts [9], [10]. As yet, however, there have been few attempts at opioid addiction relapse prediction. The ability to predict relapses is critically important; as the data shown in Figure 2 indicate, since this could potentially provide substantial support for better addiction prevention and treatment.

The addiction relapse process involves three main phases, namely emotional, mental, and physical relapses [11], [12]. Relapse is a gradual process that could start days to weeks before the actual consumption of drugs. Typically, addiction relapse begins on an emotional level, which then goes on to become a physical relapse if no action is taken to support the recovering addict. These emotional states, once confirmed, can thus be regarded as an indicator for relapse prediction. An accurate relapse prediction is essential to address the opioid epidemic since a precise relapse prediction makes it possible to implement a well-timed relapse prevention plan. Moreover, if combined with medical care and mental support, relapse prediction can prevent relapse or extend the period that an addict stays clean. Even though relapse prediction could potentially be helpful for relapse prevention, however, it remains a very challenging task for two main reasons: (1) addicts tend to be very emotional as a result of the way drugs affect their brains, making their responses and behavior patterns inherently unpredictable [13]; and (2) traditional supervised methods for relapse prediction demand a significant amount of training data.
In this study, we gathered appropriate social media data, especially data from groups such as “opiates”, “opiatesRecovery” and “Drugs” on Reddit, and built a classifier to extract data about the addicts. We then quantified the emotions represented in the comments made by the addicts and transformed these sentiment scores into sentiment images that could be fed into the discriminator. Using the GAN framework, the generator was able to learn the emotion patterns using a convolution neural network while the discriminator inferred the relapse probabilities based on a neural network with a softmax layer. Our key contributions are:

- We designed an innovative Generative Adversarial Network to predict opioid relapse that includes a discriminator that is capable of learning the data distribution patterns of both relapsing and non-relapsing addicts.
- This work provides an entirely new approach to opioid relapse prediction. Classifiers are trained to identify two types of addicts: those who are struggling to get off drugs and those who enjoy the drugs and have no plans to rehabilitate. Using social media data, we seek to predict whether a recovering opioid addict will relapse or not.
- We propose the concept of sentiment images that adequately capture the dynamic emotion fluctuations of addicts, as well as considering the social influences between users and the rich relationships between them.
- We performed a comprehensive set of experiments using a real-world dataset that produced results that consistently outperformed those obtained using other approaches, which could potentially revolutionize the practice of opioid addiction prevention and treatment.

II. RELATED WORK

We review three categories of related work as follows.

a) Biomedical Knowledge Mining on Social Network:

As social media are profoundly and widely intertwined with people’s everyday lives, they contain more and more biomedical information about users. As a result, social media represent an increasingly important data source for discovering biomedical knowledge in applications. Bian et al. proposed the use of a Support Vector Machine to identify drug users and
adverse events based on Twitter messages [6], while Fan et al. introduced an AutoDOA to automatically detect opioid addicts by building a Heterogeneous Information Network (HIN) [10]. Other researchers have also used supervised classification techniques to detect posts that involve medication abuse [8], or developed web platforms to track prescription drug abuse on social media [7]. All of these studies apply data analytics to social media data to detect drug addicts or detect drug abuse events, laying the foundation for further studies on topics such as opioid addict relapse prediction.

b) Emotional States and Addiction Relapse: Negative emotional states, such as stress, depression, anxiety, and other emotional states are known to be related to relapse. Shiffman evaluated reports of relapse in ex-smokers, finding that most of the subjects (71%) had negative emotions preceding their relapse, with the most common mood state being anxiety, followed by anger, frustration and depression [20]. Cummings et al. found that negative emotional states accounted for 30% of all relapse [21]. Those studies and their results indicate that emotions can indeed be useful indicators of people’s status.

c) Adversarial Neural Networks for Classification: A number of methods for learning a classifier from unlabeled data or only partially labeled data have been proposed. Traditional solutions to this problem lie either in generative clustering methods, which try to model the data distribution directly or in discriminative clustering methods, which separate the data without modeling its distribution [22]. There are also various methods for unsupervised and semi-supervised learning utilized by the neural network community. Among all the neural networks, Generative Adversarial Nets (GANs) stand out for their performance over a wide range of applications. Since their introduction by Goodfellow et al. [2], they have been used to generate images or videos [24], transform text to images [25], transform image styles [26], generate sequences of text [27], protect medical information privacy [28], and classify images [17], among others. Springenberg proposed CatGAN that are based on an objective function that trades off the variety of new architectural features and training procedures that can be applied to GANs classifier; their proposed model was verified on MNIST, CIFAR-10 and SVHN [16]. These earlier studies on GAN laid the foundation for applying GAN to some interesting fields, including classification and prediction.

III. Problem Formulation

In this section, after introducing the formal definition of the problem, we present the solution. Our mainly focus here is building a practical and reasonable model using social media data to predict who is likely to relapse in the next week. The reason why we are limiting our goal to predict relapse in the next week is based on several facts. First, relapse predictions for a shorter or a longer periods makes no sense, as 99% of the recovering addicts will continue to abstain for a day since their last relapse or since beginning of the detoxification process, while more than 85% will relapse within a year [29]. Moreover, the majority of current relapse prediction research is measured in terms of weeks [30], [31], hence our focus on predicting relapse in the week ahead, utilizing it as the default setting in our experiments.

Definition 3.1 (Drug Relapse): Drug relapse is defined as the resumption of drug use after a period of abstinence.

A relapse is generally associated with younger age, heavy use before treatment, a history of injecting, and not following up with aftercare [32], [33]. However, it is also frequently associated with extreme emotions and stressors [34]–[36]. Specifically, people in rehabilitation suffer from extreme emotional swings. Because of this, this paper predicts the probability of relapsing based on emotional behaviors.

Definition 3.2 (Social Community): Each social community of an addict consists of several social groups, which is called subreddit, denoted as \( S_j \). A subreddit consists of a collection of posts, which is denoted by \( s^j_i \). We can conclude that the influence an addict receives is

\[
P = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} p^i_k.
\]

For an addict \( a_j \), we quantify it with \( < P_j, F_j, Y_j > \), where \( P_j \) is the social influence received, \( Y_j \) is the relapsing indicator, and \( F_j \) represents the features of the addicts. In short, we combine \( < P_j, F_j > \) and rewrite it as

\[
X_j = < P_j, F_j >.
\]

Definition 3.3 (Task): Given a set of addicts \( A = \{a_1, a_2, a_3, \ldots \} \), and their corresponding feature variable \( X_j = < P_j, F_j > \) derived from their social community \( S_j = \{s^1_j, s^2_j, s^3_j, \ldots \} \), a task is to learn a function \( f : X_j \rightarrow Y_j \), such that \( Y_j \in [0, 1] \), where \( X_j \) and \( Y_j \) are feature variable and target variable respectively.

Definition 3.4 (Relapsing Prediction Task): Given a set of addicts \( A = \{a_1, a_2, a_3, \ldots \} \), and their state features \( X_j = \{P_j, F_j \} \in R^m \), find a mapping \( \Psi : R^m \rightarrow Y \), where \( Y \in [0, 1] \), such that the task error is minimized. Formally, we have

\[
\arg \min_{\Psi} \sum_{j=1}^{m} ||f(X_j) - Y_j||^2_2 \quad (1)
\]

In the following part, we will discuss how to apply Generative Adversarial Networks to the relapsing prediction problem. Given real data with distribution \( P_r(x) \) over a feature space \( X \), and expect to learn a distribution \( P_g(x) \) that is as close as possible to \( P_r(x) \). The closeness between two probability distributions \( P_1 \) and \( P_2 \) is defined by the divergence function \( D(P_1, P_2) \). So the optimality is defined by

\[
P_g^* = \arg \inf_{P_g} D(P_g, P_r)
\]

where \( P_g \) is the data distribution generated by generator, \( P_r \) is the data distribution of addicts and \( \inf_{P_g} D(P_g, P_r) \) is the infimum of \( D(P_g, P_r) \).

A. Generative Adversarial Networks

Goodfellow et al. (2014), who introduced the Generative Adversarial Networks (GANs) framework, proposed it as a
sort of game between a generator and a discriminator. A

generator is trained to create fake samples that could fool the
discriminator, while the discriminator is trained to distinguish
between the real samples and the fakes. Thus, after each
iteration, the generator gets better at creating realistic fakes,
while the discriminator improves its ability to tell which are
real and which fakes. The model reaches equilibrium when the
generator begins to create samples that fool the discriminator.

The GANs framework consists of two neural networks that
are trained jointly. The first neural work is implemented as a
generator, and it is initialized with some random noise \( x_0 \). The
discriminator is trained using traditional supervised learning
techniques. The input of the discriminator then consists of
two types, real data (labeled) and fake data created by the
generator. The output is scalar, simply indicating if the data
is real or fake. In theory, these two opponents playing in the
game are local differential Nash equilibrium, which is

\[ J = J(G) + (1 - J(D)) \]

between the real samples and the fakes. Thus, after each
iteration, the generator gets better at creating realistic fakes,
while the discriminator improves its ability to tell which are
real and which fakes. The model reaches equilibrium when the
generator begins to create samples that fool the discriminator.

The GANs framework consists of two neural networks that
are trained jointly. The first neural work is implemented as a
generator, and it is initialized with some random noise \( x_0 \). The
discriminator is trained using traditional supervised learning
techniques. The input of the discriminator then consists of
two types, real data (labeled) and fake data created by the
generator. The output is scalar, simply indicating if the data
is real or fake. In theory, these two opponents playing in the
game are local differential Nash equilibrium, which is

\[ J = J(G) + (1 - J(D)) \]

between the real samples and the fakes. Thus, after each
iteration, the generator gets better at creating realistic fakes,
while the discriminator improves its ability to tell which are
real and which fakes. The model reaches equilibrium when the
generator begins to create samples that fool the discriminator.

Algorithm 1: Generative Adversarial Nets \([23]\)

Initialize \( \theta_d \) for D and \( \theta_G \) for G;

for each training iteration do

for K steps do

Sample \( m \) examples \( x^1, x^2, ..., x^m \) from data
distribution \( P_{data}(x) \);

Sample \( m \) noise samples \( z^1, z^2, ..., z^m \) from the
prior \( P_{prior}(z) \);

Obtaining generated data \( \tilde{x}^1, \tilde{x}^2, ..., \tilde{x}^m \) by using
\( \tilde{x}^i = G(z^i) \);

Update discriminator parameters \( \theta_d \) to maximize
\[ V = \frac{1}{m} \sum_{i=1}^{m} \log D(x^i) + \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(\tilde{x}^i)) \]
and \( \theta_d \leftarrow \theta_d + \eta \nabla V(\theta_d) \);

end

Sample another \( m \) noise samples \( z^1, z^2, ..., z^m \) from the
prior \( P_{prior}(z) \);

Update generator parameters \( \theta_G \) to minimize
\[ V = \frac{1}{m} \sum_{i=1}^{m} \log D(x^i) + \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(z^i)) \]
and \( \theta_G \leftarrow \theta_G - \eta \nabla V(\theta_G) \);

end


B. Cost Functions

Since the discriminators of GANs are similar to a supervised
binary classifier, they have the same cost function. Suppose the
discriminator is approximated by a function \( D : \mathbb{R}^n \rightarrow [0,1] \),
and let \( (x_i, y_i) \in (\mathbb{R}^n, 0, 1) \) be the input-label pairing for training
data points, the cross-entropy of the binary discriminator
is defined as

\[ H(G) = -\sum_{i=1}^{N} (y_i \log D(x_i) + (1 - y_i) \log(1 - D(x_i))) \] (3)

In the case of GANs, since \( x_i \) only comes from two sources:
either \( x_i \sim p_{data} \), the true data distribution, or \( x_i = G(z) \)
where \( z \sim p_{generator} \) (the generator’s distribution) based on
some input code \( z \). The \( z \) samples might be from any distribution,
but for simplicity we will specify it as \( z \sim \text{unif}[0, 1] \).
Also, the best model needs precisely half of the data to come
from each of these two sources. To apply this to the cross-
entropy function above, we need to transform this probabilistically,
after replacing the sums with expectations, the \( \log(1 - D(z)) \)
term with \( \log(1 - D(z)) \) under some sampled code \( z \) for the
generator. We get

\[ J(G) = -\frac{1}{2} E_{x \sim p_{data}} [\log D(x)] - \frac{1}{2} E_z [\log(1 - D(G(z)))] \] (4)

The cost function of the generator varies according to differ-
ent specifications. First of all, in the zero-sum game, the sum
of all players’ costs is always zero. That is \( J(G) = -J(D) \),
and hence the cost function is

\[ J(G) = \frac{1}{2} E_{x \sim p_{data}} [\log D(x)] + \frac{1}{2} E_z [\log(1 - D(G(z)))] \] (5)

Because \( J(G) \) is closely tied to \( J(D) \), we can summarize
the entire game with a value function specifying the discrimi-
ator’s payoff:

\[ V(\theta_D, \theta_G) = -J(D)(\theta_G, \theta_D) \] (6)

The solution to a minimax game involves minimization in
an outer loop and maximization in an inner loop:

\[ \theta^{(G)} = \arg \min_{\theta_G} \max_{\theta_D} V(\theta_D, \theta_G) \] (7)

The primary purpose of the generative adversarial net is for
generating reliable samples according to some constraints, but
it can be modified into a classifier with some machine learning
techniques. Next, we will explain how to adjust it to fit for
relapse prediction.
C. Relapsing Prediction

The relapse prediction model is a combination of GAN and a standard neural network classifier. For a standard classifier that classifies a data point \( x \) into one of \( K \) possible classes, it takes in \( x \) as input and outputs a \( K \)-dimensional vector of logits \( \{l_1, ..., l_K\} \), that can be turned into class probabilities by applying the softmax: 

\[
p_{\text{model}}(y = i| x) = \frac{\exp(l_i)}{\sum_{k=1}^{K} \exp(l_k)}.
\]

This model is then trained by minimizing the cross-entropy between the observed labels and the model predictive distribution \( p_{\text{model}}(y|x) \). Similarly, a GAN classifier is built with standard neural classifier by simply adding samples from the GAN generator to the data set. The samples are then labeled with a new ‘generated’ class \( y = K + 1 \), and correspondingly increasing the dimension of the output of the classifier from \( K \) to \( K + 1 \). In this case, \( p_{\text{model}}(y = K + 1|x) \) indicates the probability that \( x \) is fake, corresponding to \( 1 - D(x) \) in the original GAN framework. This process is given in Figure 3. This model can also learn from the unlabeled data, that corresponds to one of the \( K \) classes of real data by maximizing \( \log p_{\text{model}}(y \in \{1, ..., K\}|x) \) \[10\]. Assuming half of the data set consists of real data and half of it is generated (this is arbitrary), the loss function for training the classifier then becomes

\[
L = -E_{x,y \sim \text{data}-}(x,y) [\log p_{\text{model}}(y|x)] - E_{x \sim G} [\log p_{\text{model}}(y = K + 1|x)]
\]

\[(8)\]

\[
L_{\text{supervised}} = -E_{x,y \sim \text{data}(x,y)} [\log p_{\text{model}}(y|x, y < K + 1)]
\]

\[(9)\]

\[
L_{\text{unsupervised}} = -E_{x \sim \text{data}(x)} [\log (1 - p_{\text{model}}(y = K + 1|x))] - E_{x \sim G} [\log p_{\text{model}}(y = K + 1|x)]
\]

\[(10)\]

Once the generative adversarial net model has been trained, we obtain a model that can be used either to predict relapse or to generate sample data. On the one hand, the discriminator has now learned the data distribution of the relapsing population and can thus be used to predict relapses. On the other hand, since the generator has learned what real relapsing data looks like, it can produce realistic fake data by imitating the real data. This process is shown in Figure 3.

IV. MODEL DEVELOPMENT, APPLICATION AND RESULTS

A. Experimental Design

In this section, we utilize an experimental study using real data from Reddit.com to comprehensively evaluate the performance of our proposed new GANs model. A series of complex intermediate experiments were involved including, for example, natural language processing and image processing, and these are described in turn below. Finally, the proposed method is evaluated by comparing its performance with those of alternative methods. The evaluation metrics are listed in Table I

\[\text{Table I} \quad \text{PERFORMANCE INDICES OF CLASSIFICATION}\]

| Indices | Description |
|---------|-------------|
| TP      | # of correct classification as positive |
| TN      | # of correct classification as negative |
| FP      | # mistakenly classified as positive |
| FN      | # mistakenly classified as negative |
| Precision | TP/(TP+FP) |
| Recall  | TP/(TP+FN) |
| ACC     | (TP+TN)/(TP+TN+FP+FN) |
| F1      | 2*Precision*Recall/(Precision+Recall) |

The GANs model described here consists of a generator and a discriminator that are trained together. In this case, the generator generates data about relapsing addicts, while the discriminator evaluates whether the input data it receives is from a real relapsing addict or not. The input data of the discriminator therefore consists of a combination of real data from addicts and fake data generated by the generator. Meanwhile, the generator continues to create new data that imitates real data from relapsing addicts in an attempt to fool the discriminator. The data it generates is fed into the discriminator, along with a stream of real data from actual relapsing addicts. The goal of the discriminator is to identify the data produced by the generator as fake. The discriminator receives three categories of data: data about addicts who have relapsed, data about addicts who are struggling to stay clean and data generated by the generator. The output of the discriminator is thus a vector of probabilities with three elements that indicate the probability that the input belongs to each category. The model that emerges at the end of this process can be used both to predict relapse probabilities and generate sample data.

\[\text{Table II} \quad \text{DATA DESCRIPTION}\]

| Dataset | # of users | # of posts | # of comments |
|---------|------------|------------|--------------|
| Group_1 (Opiates) | 1137 | 474 | - |
| Group_2 (Opi. Rec.) | 1621 | 456 | - |
| Group_3 (Drugs) | 1271 | 331 | - |
| personal_1 (Opiates) | 1137 | - | 40258 |
| personal_2 (Opi. Rec.) | 1621 | - | 30247 |
| personal_3 (Drugs) | 1271 | - | 32403 |

B. Data Collection

A web crawling tool was developed using PRAW (Python Reddit API Wrapper) to collect data from Reddit.com, a social media platform open to specific groups and sessions. The dataset for this paper came primarily from three subreddits (see Definition \[32\] on Reddit.com, namely, "Opiates", "OpiatesRecovery" and "Drugs". Within each subreddit, there are a series of discussion posts (see Definition \[32\]). We collected the data using keyword-based method, with the keywords being a list
of opioid-related drug names and its abbreviations, slangs or jargons that are related to opioids, such as ‘oxy (oxytocin), ’’Xanax’, ‘dope’, etc. The dataset consists of two parts. The first consists of the subreddit, which include comments and interactions within specific posts. After collecting the subreddit data, the user IDs of the users who either create a post or make at least one comment are extracted to produce a list user IDs from the corresponding subreddit. The second part of the dataset consists of personal data, which was collected based on the user list extracted from the first part. Based on each user ID, comments from the previous 37 days was collected using keyword-based methods; and comments/posts from the first 30 days were then used to build sentiment images, which are explained in more detail in Section IV-F, and posts/comments from the last 7 days used to label individual users as relapsed or not lapsed based on self-reporting. This involved manually reading the posts/comments from the last 7 days if a user posts/comments that he/she has relapsed, then they are labeled as a relapsed addict. In summary, we collected comments from 1,261 Reddit posts. By extracting user id from subreddits, we also collected comments containing personal history data from 4,029 Reddit users, including about 102,908 comments. A summary of the dataset is given in Table II.

![Image](image.png)

**Fig. 4.** Word Correlation for words from group discussion of ‘OpiatesRecovery’ with \( \phi \geq 0.15. \)

### C. Text Co-occurrence and Correlation

To better understand the social interactions between different users, we began by generating word correlation graphs to obtain a full picture of their conversations. Here, the textual analysis was based on the relationship between words, with the objective being to examine words that tend to immediately follow others, or that tend to co-occur within the same posts. We first tokenized comments into consecutive sequences of words. By checking how often word X was followed by word Y, we were then able to model the relationships between them using the Phi efficient metric, which is defined as

\[
\phi = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_1n_0n_0n_1}}
\]

where \( n_{11} \) represents the number of posts where both \( X \) and \( Y \) appear, \( n_{00} \) is the number where neither appears, and \( n_{10} \) is the cases where one appears without the other.

By gradually increasing the Phi coefficient, we manually filtered out the words that were most likely to appear with each other. After analyzing these posts, we visualized them as shown in Figure 4 and Figure 5. By analyzing and annotating these keywords, we summarized the most frequent topics as follows:

- Talking about the detoxification processes and medicines, such as ‘naloxone’, ‘bupe’, ‘doctor’, and ‘detox’.
- Describing pains and prescriptions from the doctor, as described by keyword such as ‘pain’, ‘doctor’, ‘painkiller’.
- Discussing withdrawal symptoms.
- There are two types of addicts: the first type includes the ones who are in recovery or struggling to recover, while the other type includes the ones who are not in the addiction recovery or not seeking recovery.

### D. Addict Classifier

Two classifiers were designed and implemented to filter out the research targets: an addict classifier and a recovering classifier. The first of these differentiates between addicts and non-addicts, while the second “within addicts” classifier separates those who are in recovery and who are not seeking recovery. Since an opioid rehabilitation must go through the detoxification process, it is meaningless to predict whether the second type of addicts will relapse. The following research mainly focuses on the first type of addicts, who are in recovery but struggling with relapse temptations.

The addict classifier is used to classify addicts from non-addicts. Based on users’ historical posts, we employed a Support Vector Machine (SVM) classifier to determine whether a user was addicted. In most of the cases, users admitted their drug dependencies. Where we were unable to identify a user as an addict, they were labeled as a non-addict. The dataset for this classifier, which consisted of 1,000 users (419 addicts and 581 non-addicts), took four people four days to compile as it required the researchers to read through all the comments made by each user and label them accordingly. Of the entire labeled dataset, 70% (namely, 70% of the users) was used as the training dataset, and the rest was used for testing. Once the classifier had been trained, we applied it to the unlabeled dataset. Identified addicts were then fed into the next classifier to determine whether they were still enjoying using drugs or recovering.

The second classifier was designed to identify the addicts who were recovering. As mentioned earlier, there are two types of addicts. The first type of addict enjoys drugs, and hence will almost certainly relapse, so the relapse prediction for this group makes no sense. A similar method was then applied...
to identify these addicts who were willing to become drug-free but struggling with the process. This took another four days to label 1,000 users (with 375 recovering users and 625 users who still enjoyed drugs). Among this labeled data, 70% was used as the training dataset and the rest for testing. The performance of the two classifiers is shown in Table III.

**TABLE III**

| Addict Classification     | Recovering Classification |
|---------------------------|---------------------------|
|                           | Accuracy | Recall | Precision | F1     | Accuracy | Recall | Precision | F1     |
| Log. Regr.                | 0.8180   | 0.8277 | 0.8180     | 0.8228 | 0.7433   | 0.7834 | 0.7334     | 0.7575 |
| KNN                       | 0.7027   | 0.8144 | 0.7592     | 0.7858 | 0.7521   | 0.8121 | 0.7327     | 0.7703 |
| SVM                       | 0.9176   | 0.9429 | 0.9346     | 0.9388 | 0.8762   | 0.9290 | 0.8447     | 0.8849 |

E. Emotion Measure

A word-emotion association lexicon provided by Mohammad et al. [38] was utilized in this study to quantify the emotional content of each comment from a user. Ten categories of emotion are identified in this lexicon: anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive. Some sample words from each category are given in Figure 8. Each of these categories is associated with an emotion count variable. If a term used in a comment is listed in one lexicon category, the count variable for that category is increased by one. The process is repeated for the entire comment database.
Once the process was completed, the counts were normalized by dividing each count variable over the maximum count within the same comment to provide a relative comparison between emotions. The emotion category with the maximum value then represents the most influential or dominant emotion of this user. The smaller the value of the normalized emotional count, the less influential that emotion is.

In particular, suppose $L_i$ is the subset of a lexicon which contains words in emotion $i$. The emotion count $\text{emotion\_count}_i$ and normalized count $\text{n\_emotion\_count}_i$ for the emotion $i$ are:

$$\text{emotion\_count}_i = \sum \text{word}_j$$

$$\text{n\_emotion\_count}_i = \frac{\text{emotion\_count}_i}{\max(\text{emotion\_counts})}$$

Where:

$$\text{word}_j = \begin{cases} 1 & \text{if word}_j \text{ in } L_i \\ 0 & \text{otherwise} \end{cases}$$

F. Sentiment Image

A sentiment image for an addict is defined using a sentiment matrix. After quantifying the sentiment scores for every comment, we obtained a list of sentiment scores for each addict: for each comment, the sentiment score is arranged as [Anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive]. We calculated the average sentiment score for every three comments because we collected 30 comments from each user and the emotion vector has a length of 10. Next, we stacked up the average sentiment score for the thirty posts to form the sentiment image, a 10*10 matrix with each row representing the average sentiment score of three comments. The sentiment images of the addicts are then fed into the discriminator as real data from relapsing addicts.

Social influences or interactions are also considered when building the sentiment images. Just like a color image that has 3 RGB (Red, Green and Blue) channels, the sentiment image we utilized here has two channels. The first of these channels is a 10*10 matrix, which is built by stacking up ten sentiment vectors. The second channel is constructed by taking the social network influences into consideration. Specifically, if an addict starts a discussion by asking a question or describing his or her current status, other users may make comments and then the discussion creator may reply. By exchanging opinions in this way, people who exchange opinions can influence each other. For the purposes of this research, we defined social network influence as follows:

$$\text{influence} = \begin{cases} \frac{1}{n} \sum_{i=1}^{n} S_i & \text{if n comments is replied} \\ 0 & \text{no reply} \end{cases}$$

If no one replies, the discussion creator receives a 0 influence score. If dozens of users reply and the creator replies, then the creator receives an average of all the comments. For instance, in Figure 4. we show the social interaction with user ‘**luntold12**’, where ‘*’ represents a character that is hidden to protect the users’ privacy. Figure 7 shows the social interaction for users from ‘Opiates’, ‘OpidatesRecovery’ and ‘Drugs’. By analyzing the group interaction, we find that in most of the posts, the total number of comments are smaller than 10. Those participating in ‘Drugs’ discuss illegal drug usage more, while people in ‘OpidatesRecovery’ and ‘Opiates’ tend to discuss how to get rid of drugs.

The 2-channel sentiment images were then fed into the generative adversarial network to train the model.

G. Relapse Prediction

The relapse prediction model was modified from the discriminator of a standard generative adversarial net by applying a softmax layer to the output of a discriminator [16]. We trained the model for 7,000 epochs with 1,000 labeled sentiment images (consisting of 330 addicts who have managed to stay clean for a while, and 670 addicts who have relapsed). Since no labeled dataset is available online, we labeled the data of 1,000 users and transformed the data into images with two layers (sentiment layer and social influence layer, Channel=2). For each mini-batch, we randomly selected 128 images. For optimization we used Adam [39] with a learning rate=0.0001. Our implementation of the GANs model is based on Tensorflow [40]. The data flows inside the model can be visualized as shown in Figure 9 and Figure 10. The dimensions of the tensors are calculated as the tensor dimension calculation in convolution neural networks [41]. Figure 9 shows the input of a discriminator as a mixture of sentiment images and data from the generator, with the output being...
Fig. 11. These sentiment images are from four addicts produced by GAN when model converged. It can be observed that the emotion ‘joy’ and emotion ‘negative’ are closely connected to relapse.

Fig. 12. Generator loss changes in training.

Fig. 13. Discriminator loss changes in training.

V. Conclusion

In this paper, we present a new application where generative adversarial nets are utilized to predict whether a recovering opioid addict will relapse or not within the next week. This paper is one of the first attempts to predict opioid addict relapses using massive social media data from Reddit.com. By combining social media data with generative adversarial networks, our model makes predictions based on addicts’ previous records and current sentiment status. Our experimental results demonstrate that the new GANs model consistently outperforms traditional prediction techniques and could thus be incorporated into a strategy for supporting practitioners working on opioid relapse prevention and addiction treatment.

Unfortunately, the limitations of this study are clear: the size of manually labeled training data used in the experimental is not large enough to draw strong and consolidated conclusions. We are aware of the existence of this limitation but leave it as it was because processing and manually labeling a large amount of training data is very costly. This limitation can be fixed by investing more time to collect, process and manually label more training data. We leave the limitation for future research.

REFERENCES

[1] M. W. Holly Hedegaard and A. M. Minino, “Drug overdose deaths in the united states,” https://www.cdc.gov/nchs/products/databriefs/db294.htm, December-2017, online, accessed 25-March-2018.

[2] NIDA, “Overdose death rates,” https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates, September-2017, online, accessed 25-March-2018.
TABLE IV

| Relapse prediction with different sizes of training data |
|--------------------------------------------------------|
|   | 90% | 80% | 70% | 60% | 50% |
| Log.Reg ACC | 0.7150 | 0.6920 | 0.6520 | 0.6320 | 0.6020 |
| F1 | 0.7775 | 0.7582 | 0.7268 | 0.7065 | 0.6801 |
| SVM ACC | 0.8210 | 0.7980 | 0.7680 | 0.7380 | 0.7180 |
| F1 | 0.8500 | 0.8412 | 0.8162 | 0.7907 | 0.7749 |
| KNN ACC | 0.7120 | 0.7980 | 0.6590 | 0.6290 | 0.5990 |
| F1 | 0.7736 | 0.8412 | 0.7283 | 0.7020 | 0.6753 |
| GAN ACC | 0.8390 | 0.8190 | 0.7840 | 0.7490 | 0.7140 |
| F1 | 0.8547 | 0.8491 | 0.8305 | 0.8013 | 0.7716 |

[87] J. R. Geddes, S. M. Carney, C. Davies, T. A. Furukawa, D. J. Kupfer, E. Frank, and G. M. Goodwin, “Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review,” The Lancet, vol. 361, no. 9358, pp. 653–661, 2003.

[88] S. Shiffman, “Relapse following smoking cessation: a situational analysis,” Journal of consulting and clinical psychology, vol. 50, no. 1, p. 71, 1982.

[89] G. J. R. Cummings, C. and G. A. Marllatt, “he addictive disorders: Treatment of alcoholism, drug abuse, smoking, and obesity,” no. 3, pp. 291–322, 1980.

[90] J. T. Springenberg, “Unsupervised and semi-supervised learning with categorical generative adversarial networks,” arXiv preprint arXiv:1511.06390, 2015.

[91] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in Neural Information Processing Systems 27, 2014, pp. 2672–2680.

[92] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.

[93] A. M. Elgammal, B. Liu, M. Elosheney, and M. Mazzone, “CAN: creative adversarial networks, generating art by learning about styles and deviating from style norms,” CoRR, 2017. [Online]. Available: http://arxiv.org/abs/1706.07068

[94] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” arXiv preprint, 2017.

[95] L. Yu, W. Zhang, J. Wang, and Y. Yu, “Seqgan: Sequence generative adversarial nets with policy gradient,” in AAAI, 2017, pp. 2852–2858.

[96] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and G. Greenspan, “Synthetic data augmentation using gan for improved liver lesion classification,” arXiv preprint arXiv:1801.02385, 2018.

[97] R. Sinha, “New findings on biological factors predicting addiction relapse vulnerability,” Current psychiatry reports, vol. 13, no. 5, p. 398, 2011.

[98] L. O. Bauer, “Predicting relapse to alcohol and drug abuse via quantitative electroencephalography,” Neuropsychopharmacology, vol. 25, no. 3, pp. 332–340, 2001.

[99] R. L. Ashare, E. P. Wileyto, K. A. Perkins, and R. A. Schnoll, “The first seven days of a quit attempt predicts relapse: Validation of a measure for screening medications for nicotine dependence,” Journal of addiction medicine, vol. 7, no. 4, p. 249, 2013.

[100] B. P. Smyth, J. Barry, E. Keenan, and K. Ducray, “Lapse and relapse following inpatient treatment of opiate dependence.” Irish medical journal, 2010.

[101] H. Chalana, T. Kundal, V. Gupta, and A. S. Malhari, “Predictors of relapse after inpatient opioid detoxification during 1-year follow-up,” Journal of addiction medicine, vol. 2016, 2016.

[102] J. D. Ford and E. Russo, “Trauma-focused, present-centered, emotional self-regulation approach to integrated treatment for posttraumatic stress and addiction: trauma adaptive recovery group education and therapy (target).” American journal of psychotherapy, vol. 60, no. 4, 2006.

[103] J. Elster, Strong feelings: Emotion, addiction, and human behavior. mit Press, 2009.

[104] R. Sinha, H. C. Fox, K. A. Hong, K. Bergquist, Z. Bhagwagar, and K. M. Siedlarz, “Enhanced negative emotion and alcohol craving, and altered physiological responses following stress and cue exposure in alcohol dependent individuals,” Neuropsychopharmacology, vol. 34, no. 5, p. 1198, 2009.

[105] L. J. Ratliff, S. A. Burden, and S. S. Sastry, “Characterization and computation of local nash equilibria in continuous games,” in Proceedings of the 2013 ACM CIKM. ACM, 2017, pp. 1259–1267.

[106] theraleighhouse, “3 stages of drug addiction relapse: Emotional, mental & physical,” https://www.theraleighhouse.com/addiction-blog/resources-myself/3-stages-addiction-relapse-emotional-mental-physical, 7-September-2017, online, accessed 25-March-2018.

[107] K. D. Brownell, G. A. Marllatt, E. Lichtenstein, and G. T. Wilson, “Understanding and preventing relapse,” 1986.

[108] G. F. Koob, “Dynamics of neuronal circuits in addiction: reward, antireward, and emotional memory,” Pharmacopsychiatry, vol. 42, no. Suppl 1, p. S32, 2009.

[109] theraleighhouse, “Seeking drug abuse treatment: Know what to ask,” https://www.drugabuse.gov/publications/seeking-drug-abuse-treatment-know-what-to-ask/introduction, 7-Sept-2017, online, accessed 25-March-2018.

[110] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in neural information processing systems, 2014, pp. 2672–2680.

[111] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” in Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.

[112] I. Goodfellow, “Nips 2016 tutorial: Generative adversarial networks,” arXiv preprint arXiv:1701.00160, 2016.

[113] G. A. Marllatt and D. M. Donovan, Relapse prevention: Maintenance strategies in the treatment of addictive behaviors. Guilford press, 2005.