Original Paper

A Machine Learning Approach for the Detection and Characterization of Illicit Drug Dealers on Instagram: Model Evaluation Study

Jiawei Li[1,2], MS; Qing Xu[1,2], MAS; Neal Shah[2], BS; Tim K Mackey[1,2,3,4], MAS, PhD

[1]Department of Healthcare Research and Policy, University of California - San Diego, Extension, La Jolla, CA, United States

[2]Global Health Policy Institute, La Jolla, CA, United States

[3]Department of Anesthesiology, University of California - San Diego, School of Medicine, La Jolla, CA, United States

[4]Division of Infectious Disease and Global Public Health, University of California - San Diego, School of Medicine, La Jolla, CA, United States

Corresponding Author:
Tim K Mackey, MAS, PhD
Department of Anesthesiology
University of California - San Diego, School of Medicine
8950 Villa La Jolla Drive
A124
La Jolla, CA, 92037
United States
Phone: 1 9514914161
Email: tmackey@ucsd.edu

Abstract

Background: Social media use is now ubiquitous, but the growth in social media communications has also made it a convenient digital platform for drug dealers selling controlled substances, opioids, and other illicit drugs. Previous studies and news investigations have reported the use of popular social media platforms as conduits for opioid sales. This study uses deep learning to detect illicit drug dealing on the image and video sharing platform Instagram.

Objective: The aim of this study was to develop and evaluate a machine learning approach to detect Instagram posts related to illegal internet drug dealing.

Methods: In this paper, we describe an approach to detect drug dealers by using a deep learning model on Instagram. We collected Instagram posts using a Web scraper between July 2018 and October 2018 and then compared our deep learning model against 3 different machine learning models (eg, random forest, decision tree, and support vector machine) to assess the performance and accuracy of the model. For our deep learning model, we used the long short-term memory unit in the recurrent neural network to learn the pattern of the text of drug dealing posts. We also manually annotated all posts collected to evaluate our model performance and to characterize drug selling conversations.

Results: From the 12,857 posts we collected, we detected 1228 drug dealer posts comprising 267 unique users. We used cross-validation to evaluate the 4 models, with our deep learning model reaching 95% on F1 score and performing better than the other 3 models. We also found that by removing the hashtags in the text, the model had better performance. Detected posts contained hashtags related to several drugs, including the controlled substance Xanax (1078/1228, 87.78%), oxycodone/OxyContin (321/1228, 26.14%), and illicit drugs lysergic acid diethylamide (213/1228, 17.34%) and 3,4-methylenedioxy-methamphetamine (94/1228, 7.65%). We also observed the use of communication applications for suspected drug trading through user comments.

Conclusions: Our approach using a combination of Web scraping and deep learning was able to detect illegal online drug sellers on Instagram, with high accuracy. Despite increased scrutiny by regulators and policymakers, the Instagram platform continues to host posts from drug dealers, in violation of federal law. Further action needs to be taken to ensure the safety of social media communities and help put an end to this illicit digital channel of sourcing.

(J Med Internet Res 2019;21(6):e13803) doi: 10.2196/13803

KEYWORDS

opioids; social media; narcotics; substance abuse; machine learning; internet; prescription drug abuse; artificial intelligence

Introduction

Background

In June 2018, the US Food and Drug Administration (FDA) held the *Online Opioid Summit*, a 1-day meeting seeking to generate momentum around the need to combat illicit internet sales of opioids [1]. In addition to federal agencies, several internet and social media companies were in attendance, including Google (which operates YouTube and Google+), Twitter, Facebook (which operates Instagram and WhatsApp), Pinterest, and other e-commerce, technology, and patient safety organizations [2]. The meeting was organized to facilitate cooperation among these stakeholders to address illicit internet opioid sourcing and diversion, a challenge that adds fuel to a national public health emergency that claims the lives of an average of 130 people daily from addiction-related overdose [3].

Importantly, federal law explicitly prohibits the internet sale of controlled substances as enforced by the 2008 Ryan Haight Online Pharmacy Consumer Protection Act (RHA) [4,5]. Named after a Californian adolescent who died in 2001 after overdosing on Vicodin purchased from an online drug seller without a prescription, the RHA was meant to curb the use of the internet as an alternative and convenient channel of sourcing [6]. However, since Mr Haight’s death, the internet ecosystem has rapidly proliferated and diversified, now populated by illegal internet pharmacies, social media posts from illegal sellers, and dark Web vendors, all who have been implicated in illegal online opioid sales [6-12].

Though illegal prescription drug sales are often found by users through search engine results and internet pharmacy advertisements (including spam email), popular social media platforms have emerged as a direct-to-consumer marketing tool for illegal sellers [2,12-15]. Previous research and investigative reporting have detected illegal opioid sales and drug dealing on several social media platforms, including Twitter, Facebook, and Instagram [7,12,16-21]. As Twitter provides a convenient way of accessing data through its application programming interface (API), many substance abuse infoveillance studies have focused on this microblogging platform [22-25]. Our own prior studies used both supervised machine learning classifiers and an unsupervised topic model to detect internet pharmacies selling opioids (including fentanyl) [6,7,21,26]. Others have primarily focused on analyzing Twitter messages for opioid and substance abuse behavior with manually annotated data, examining social circles of users, measuring user sentiment, using natural language processing, and using deep learning [22,23,27,28]. Other studies have used deep learning models to detect and describe adverse drug reactions via Twitter [29,30].

However, there are far fewer studies that have conducted infoveillance research on the Instagram platform, likely because of the difficulty of collecting Instagram data and the different data features that require additional data cleaning and processing. Instagram is an image and video sharing social media platform (reaching 1 billion monthly users in 2018) and is particularly popular among young adults (ie, 71% of those aged between 18 and 24 years), a critical demographic for substance use initiation [17]. Prior studies by Zhou et al have analyzed Instagram data primarily for substance abuse behavior and did not use deep learning models but instead used other machine learning approaches [28]. One relevant study by Yang and Luo used a model based on multitask learning that analyzed both images and text to track and classify drug abuse–related posts, including differentiating for drug dealers [17]. The study analyzed the user timelines of identified posts to differentiate drug dealers from users who exhibited drug use behavior and achieved a high classification accuracy of 88% [17].

Objective

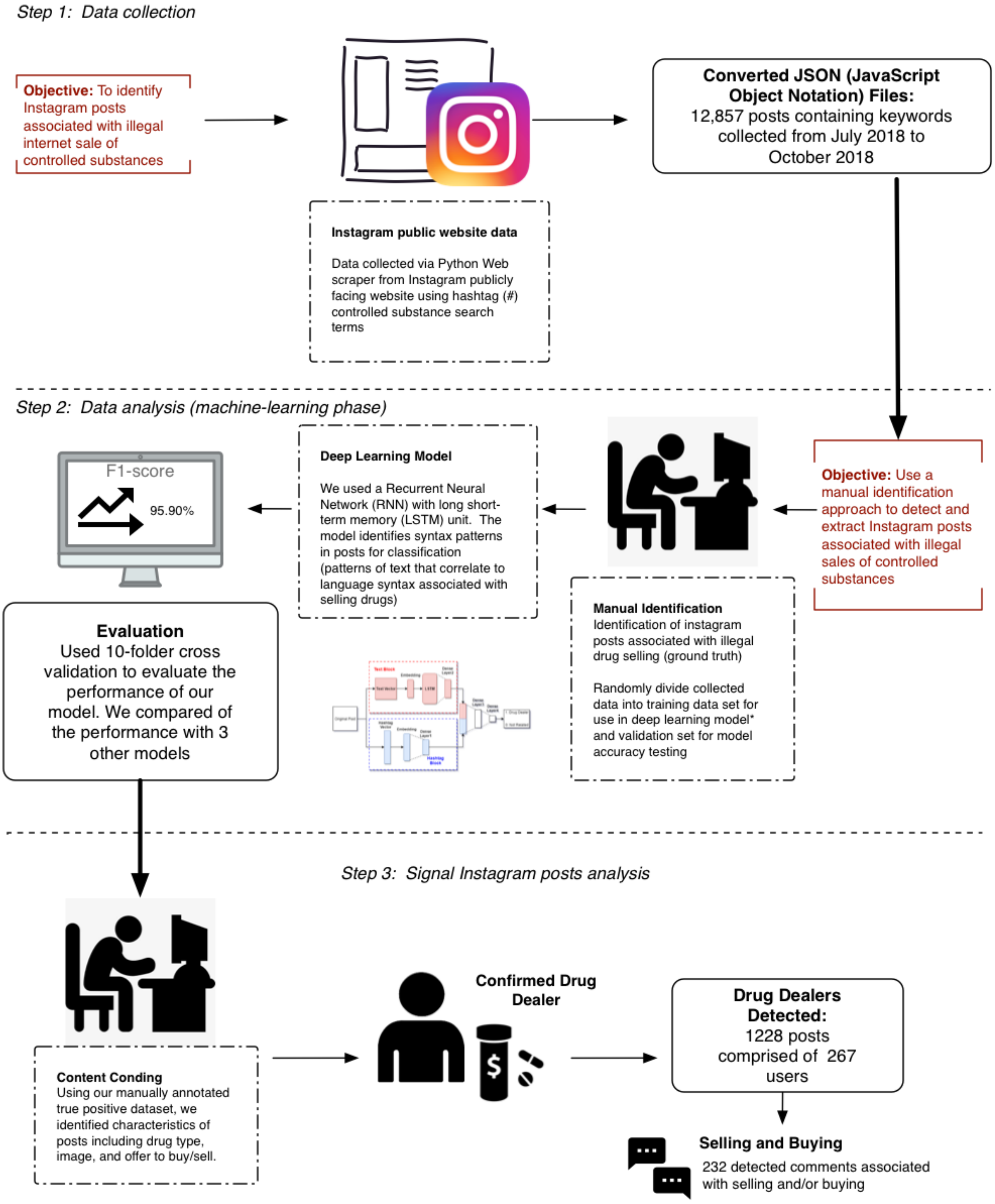
Building on these prior studies that have used different big data and machine learning approaches to detect substance abuse behavior and illegal drug selling on social media, this study describes the use and evaluation of a deep learning model to better automate the detection of illegal opioid and other illicit drug sales on Instagram. Our study focuses on detecting illegal drug selling posts (not accounts) from hashtag searches and using deep learning to analyze text from posts.

Methods

Overview

Our study comprised 3 phases: data collection, data processing, and model training. The goal of our study was to develop and evaluate a machine learning approach that has the best performance for identifying illicit opioid drug selling via Instagram (Facebook, Inc.). To accomplish this, we first collected a set of posts that contained suspicious drug selling behavior by conducting automated searches for opioid-related hashtags and posts. We then used posts detected in these hashtag searches as a training set for our deep learning model, so we could better enable detection of posts in the entire corpus of all messages collected (see Figure 1 for summary of methods). Importantly, Instagram is a platform that has different ways of presenting messages from its users. For other social media platforms such as Twitter and Facebook, the main content in the post is often the text accompanied by hashtags. However, on the basis of the method of searching for messages on Instagram (ie, users search for posts and topics by hashtags), the more hashtags a post contains, the easier it will be found. Therefore, a post for Instagram usually comprises 3 main pieces of content: text, hashtags, and an image. In this study, we will analyze the performance of models examining text and hashtags to determine what is the best approach to detecting illegal drug dealers.

Figure 1. Summary of study methodology.


Step 1: Data collection
Objective: To identify Instagram posts associated with illegal internet sale of controlled substances
Converted JSON (JavaScript Object Notation) Files:
12,857 posts containing keywords collected from July 2018 to October 2018
Instagram public website data
Data collected via Python Web scraper from Instagram publicly facing website using hashtag (#) controlled substance search terms
Step 2: Data analysis (machine-learning phase)
F1-score
95.90%
Deep Learning Model
We used a Recurrent Neural Network (RNN) with long short-term memory (LSTM) unit. The model identifies syntax patterns in posts for classification (patterns of text that correlate to language syntax associated with selling drugs)
Objective: Use a manual identification approach to detect and extract Instagram posts associated with illegal sales of controlled substances
Evaluation
Used 10-folder cross validation to evaluate the performance of our model. We compared the performance with 3 other models
Manual Identification
Identification of instagram posts associated with illegal drug selling (ground truth)
Randomly divide collected data into training data set for use in deep learning model* and validation set for model accuracy testing
Step 3: Signal Instagram posts analysis
Confirmed Drug Dealer
Drug Dealers Detected:
1228 posts comprised of 267 users
Content Conding
Using our manually annotated true positive dataset, we identified characteristics of posts including drug type, image, and offer to buy/sell.
Selling and Buying
232 detected comments associated with selling and/or buying

Data Collection Phase

Since July 2018, Instagram has disabled many functions of its public API and has set limitations on data collection. To collect a more representative dataset than what is available from its limited public API, we developed a method to Web scrape results from the Instagram platform website based on opioid keyword hashtags (a word or phrase preceded by a hash character, #) search results. Our Web scraper was built in the computer programming language Python and converted source code from the Instagram website into JavaScript Object Notation (JSON) data files. Data scraped from the website included the text of the Instagram post, comments to the post by Instagram users, and associated metadata (eg, date, time, and user profile information).

As search on Instagram is conducted using hashtags (eg, #subject), we chose an initial list of hashtags related to controlled substances identified via manual searches and then populated these opioid-related hashtags into our automated search and Web scraping data collection process. This allowed us to discover a greater number of opioid and drug-related hashtags that are present in the Instagram community. Our initial set of hashtag keywords included *xanaxangel*, *percocert* (Percocet misspelled), *adderrall* (Adderall misspelled), and *hydrocodoneacetaminophen* (misspelled.) These keywords relate to hashtags associated with common controlled substance drug names that were used on Instagram at the time of the study. Many of these keywords are misspelled as Facebook and Instagram currently disables search results for certain explicit opioid keywords in search queries [2]. However, alternative opioid hashtag keywords are relatively easy to find, including some that are derived from the platform's own suggested alternative hashtags when conducting searches [31].

Importantly, hashtags used by drug dealers are different depending on what type of drug(s) they are selling. Hence, it is important to expand the number and diversity of possible drug-related hashtags to collect a better sample of data to analyze. To increase the number of hashtags likely related to drug dealing, we examined search results for our initial set of hashtags using a 2-loop process during our automated search and Web scraping. In the first loop, we captured data from an individual post under a certain hashtag in the initial set of hashtags used. Our Web scraper continued to collect the JSON data from the source code for each of these posts until it reached a set limitation (in this study we ended our search loop when the post was older than 3 months). We then filtered out the hashtags from each post and chose other hashtags that contained keywords associated with controlled substances and illicit drugs by manually inspecting all hashtags collected in the loop. Hashtags that contained opioid and drug-related keywords or combinations thereof were then added to the hashtag list so they could be searched in a second loop that would go through all the hashtags identified. A full list of hashtags identified in this study is available in Multimedia Appendix 1.

Data collection occurred from July to October 2018, with hashtags limited to English language and with no set geographic or other filters for posts. In our data cleaning process, we also discarded duplicate posts that were replicated when collecting posts under different hashtags search loops, and then removed hashtags and stop words before textual analysis in the data processing phase described below.

Data Processing Phase

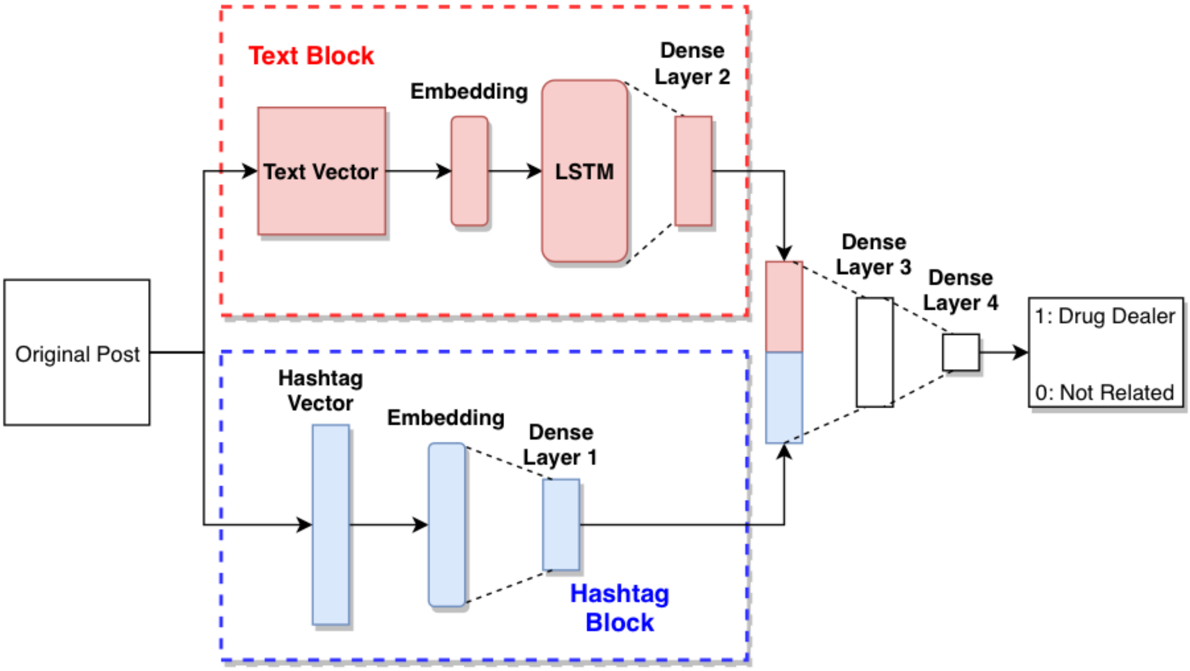
An Instagram post generally contains content with the user's text (ie, message) and also hashtags to self-curate the content and associate it with other user posts and comments. Unlike normal text, hashtags can be placed randomly in the post, which could affect the accuracy of the machine learning model used in this study, as our deep learning model is based on learning the pattern of the text. To address this, we took the following additional steps to process our data:

1. We eliminated duplicate results of posts that appeared in multiple hashtag searches.
2. From each post, we extracted the text and removed any hyperlinks, special characters (eg, emoji, !, @), and stop words (eg, hers, between, yourself). We used the Natural Language Toolkit package in Python to remove stop words. We did not exclude # in this study, as # represents the hashtags of keywords that we specifically wanted to analyze.
3. We then built a dictionary based on the words in all texts; each word had a corresponding index. Then we transformed the word into an index in each text.
4. We kept text that had more than 1 index after step 2 as original data; then we removed all the hashtags from the original data and removed the texts that had no remaining words left. These data are referred to as no hashtag data.
5. We eliminated the duplicate texts in each dataset after step 3. Hence, all the texts remaining were unique with a different pattern.
6. We then manually annotated all the posts detected (including text and images) to identify and classify posts that involved illegal drug dealing or selling to establish ground truth for model evaluation. We accomplished this by using a binary coding scheme of yes or no based on assessing whether a post contained text or image information about a prescription opioid, controlled substance, or other suspected illicit drug product and that the post also included contact information on how to trade or purchase the drug from the dealer. The second and the third author coded posts independently and achieved a high intercoder reliability for results (kappa=0.98). For inconsistent results, both authors met and reviewed the posts together with the last author (a subject matter expert in internet substance abuse behavior) and all authors conferred on the correct classification of the post.
7. Using the manually annotated ground truth data in step 6, we then evaluated the performance of a deep learning model to identify illegal drug sellers by comparing it with other machine learning models.

Model Training

In this study, we used 4 supervised models to analyze our Instagram data: decision tree, random forest (RF), support vector machine (SVM), and a deep learning model we developed [32-34]. The first 3 models are traditional machine learning models, which perform well for classification tasks [35-37]. For this study's deep learning model, we used a recurrent neural network with long short-term memory (LSTM) unit to study the pattern of text in a post. This model is well-suited for classifying, processing, and making predictions based on time series data and also exhibits high performance on speech text analysis [38-41].

Figure 2. Structure of deep learning model. Embedding layer: input_dim is 29,832, which is the size of the dictionary; the input_length for text is 50, input_length for hashtag is 15; the output_dim is 400. Long short-term memory layer: contains 800 units, with dropout=0.2, recurrent_dropout=0.2. Dense layer 1: unit=200, activation=sigmoid. Dense layer 2: unit=200, activation=sigmoid. Dense layer 3: unit=200, activation=sigmoid. Dense layer 4: unit=1, activation=sigmoid. Optimizer: Adam (learning rate set at 0.0001). Loss: Binary_crossentropy. LSTM: long short-term memory.



The structure of the model is shown in Figure 2. It contains 2 major parts, text block and hashtag block. The text block was used to store the information in the text only, the hashtag block was used to store the information from the hashtags only. When we input the data, we separated the original text into the text part and hashtags part, and for each part we generated a vector based on the dictionary we built in the data processing stage. The length of the text vector was 50 words, which is the average length of texts in our dataset (if the size of the vector is too large, it will increase the workload for the model; if it is too short, we may lose signal data). If the text was longer than 50 words, we only chose the first 50 words to maintain the efficiency of the model; if it was shorter, we filled the rest of the vector with 0. Then we used word embedding to generate the text into a word vector with dimensions (50, 400). The strategy for hashtags was the same, but the average length of hashtags was set to 15 characters.

The text block contains 1 layer of LSTM combined with 1 fully connected network (FCN) [42]. LSTM is a recurrent neural network that can be used to predict the next word based on the current word. Hence, it will learn the correlation between different words, which is the pattern of the text, during the training process. The FCN (dense layer) is used to learn features from the original data and to keep as much information as possible in another vector (in our study we kept the output from the former layer into a smaller dimension vector). The dense layers 1 and 2 in these 2 blocks will keep the signal from text to a much smaller vector. We then merged these 2 vectors together and used another 2 FCNs (dense layers 3 and 4) to generate the prediction.

Every time the model gives a prediction (possibility for 1=drug dealer and 0=not related; if the possibility is larger than 0.5, we consider it as drug dealer), the model will calculate the difference between the prediction and the label; the difference is referred to as the loss. If the loss is low, that means the model is performing well on prediction. There are many ways to calculate the loss; in our model, we used binary cross-entropy loss [43]:

$$\text{Loss}=-(y\times\log(p)+(1-y)\times\log(1-p))$$

This is the format for a binary class only; y represents the ground truth of each text (either 0=not drug dealer or 1=drug dealer), p represents the predicted probability that the text is of the class. The loss shows the difference between our prediction and ground truth.

After we get the loss, the model will run back propagation to update the parameters in LSTM and FCN to decrease the loss value [44]. The updating is based on the learning rate and optimizer we choose. The learning rate is used to keep the learning process in an acceptable range. If the learning rate is too small, it will take a longer time for the model to achieve the best parameters; if it is too large, it may never reach the local minimum for the loss. Therefore, an Adam optimizer algorithm was used for back propagation, which allows the learning rate to adapt based on the parameters which will make large updates for infrequent parameters and small updates for frequent parameters [45].

During the updating process, the model will learn how to distinguish what a drug dealer's post looks like. On the basis of this pattern, the model can identify syntax patterns in posts to classify them as signal posts (patterns of text that correlate to language syntax associated with selling drugs) and can then separate posts that contain relevant hashtag keywords, whose syntax is not related to drug selling.

The data were separated into a training set and validation set; the training set was used to update the model and the validation set was used to evaluate the performance of our model, which we discuss further in the Model Evaluation section. In the training process, the deep learning model will keep looping on learning the training set to reach the minimum of the loss. However, this process could cause overfit, which will make the model too sensitive to the pattern of the texts in the training set. To prevent overfit, we need another dataset that is different from the training set (ie, the validation set). If the loss of the validation set is increasing, it means that the model learned too much from the training set and the training needs to be stopped. In our model, we compare the most recent validation loss with all 4 previous validation losses; if it is larger (the loss is continually increasing), the model will stop training.

Ethics Approval and Consent to Participate

Ethics approval and consent to participate were not required for this study as data derived for this study were available from public sources and there were no interactions with social media users.

Data Availability

The pretrained model for this study is available on GitHub under the user account name Mathison under the file Instagram_drugdealer_detection.

Results

Model Evaluation

We collected a total of 12,857 Instagram posts over a 3-month period (July 19, 2018 to October 18, 2018) that we used for analysis. There were a total of 1228 drug dealer posts based on the manual annotation we conducted, which comprised 267 unique users. As of October 18, 2018, 206 of these posts were still active and viewable on the Instagram platform. As the volume of the target posts is low compared with the total dataset, we used 10-folder cross-validation to evaluate the performance of our model. This method is used to estimate how the model is expected to perform when facing limited samples. We shuffled each dataset (original data and no hashtag data) randomly and separated them into 10 folders—9 of them are the training set (in the deep learning model we used 70% as the training set and 30% as the validation set) and the rest are the test set used to assess the area under the curve (AUC), precision, recall, and the F1 score of the model. When calculating the AUC, we used the whole dataset rather than the cross-validation. There were a total of 10 iterations; for each iteration, we chose a different folder as the test set. Hence, this allowed us to ensure that each text could be used to make prediction. For each iteration, the model was reset. After we finished all iterations, we could calculate the average score for each model.

For the deep learning model, to prevent the problem of overfit, we separated the original training set into 2 parts, 70% of the set as the training set and 30% as the validation set. This allowed us to ensure that the model we generate from each iteration can have the best performance on the test set.

The results from our model evaluation are separated into 3 parts according to the data we preprocessed (Table 1): text with hashtags, hashtags only, and text without hashtags.

On the basis of this evaluation, the deep learning model has the best performance compared with the other 3 models based on the F1 score. However, the precision of the deep learning model does not show a better result than RF or SVM, suggesting that deep learning is not more effective at filtering for false positives.

Table 1. Performance for each model based on variations of text and hashtag use.

Performance measure	Decision tree	Random forest	Support vector machine	Study model
Text with hashtags, %				
Precision	95.05	96.00	96.86	94.81
Recall	82.15	86.08	81.21	91.42
F1 score	88.13	90.77	88.35	93.09
Area under the curve	96.67	96.85	97.18	98.12
Hashtags only, %				
Precision	86.22	94.14	95.39	89.60
Recall	86.50	87.13	84.24	88.89
F1 score	86.36	90.50	89.47	89.24
Area under the curve	95.95	95.23	95.43	94.32
Text without hashtags, %				
Precision	88.49	97.07	97.80	93.60
Recall	93.08	91.31	89.32	98.31
F1 score	90.73	94.11	93.37	95.90
Area under the curve	95.56	94.85	93.49	99.12

The possible reason is that drug dealers on Instagram usually have similar distinct text formats of selling drugs, which includes providing contact information with a phone number or email address. Therefore, once the pattern of the text is established, the deep learning model can identify these posts with greater ease. This is also shown in the results; the recall of the deep learning model is much higher than in other models. However, the text patterns of other nonrelated posts are not similar, so the deep learning model may be confused by random patterns which makes its precision, compared with other models, lower.

Table 1 shows that the model performance on hashtags only does not improve when compared with performance on texts with hashtags, especially for the deep learning model. However, when we use the text without the hashtags, all the scores increased except for the recall of the deep learning model compared with performance on texts with hashtags (Table 1). These performance results suggest that removing hashtags can increase the accuracy of the model, but may increase false positives for the deep learning model. This may occur when the user is engaging in nondrug related sales, but uses the same text pattern with different hashtags than drug dealers, as removal of hashtags makes these texts patterns appear more similar to posts from actual drug dealers.

Text Analysis of Drug Dealer Posts

When manually annotating the text contained in the 1228 drug selling posts, we identified 2 forms of communication by users: (1) the use of hashtags (#) in front of illicit drug-related keywords to self-curate content and (2) descriptive language of drug selling–related activity in combination with related images posted by users. Each post can contain a combination of types of text (eg, hashtags and/or descriptive text), images or videos, and other metadata associated with the Instagram user. The vast majority of posts (1196/1228, 97.39%) had hashtags, whereas 32 out of 1228 posts (2.61%) only had descriptive language with no hashtags, which we suspect were posts modified after data collection occurred and before manual annotation.

The majority of hashtags detected in signal posts were related to the controlled substance Xanax (Alprazolam, a nonopioid controlled substance; 1078/1228 posts, 87.78%), followed by oxycodone/OxyContin (321/1228, 26.14%) and illicit drugs lysergic acid diethylamide (LSD; 213/1228, 17.34%) and 3,4-methylenedioxymethamphetamine (MDMA; 94/1228, 7.65%; see Table 2 for summary).

Table 2. Number of posts related to controlled substance hashtags (N=1228).

Drug name[a]	Hashtag[b]	Posts, n (%)[c]
Xanax	#xanax	802 (65.3)
	#xanaxfamily	530 (43.1)
	#2mgxanax	321 (26.1)
	#zanax	112 (9.1)
	#greenxanax	84 (6.8)
	Total	1078 (87.8)
Oxycodone/OxyContin	#oxycodone	30 (2.4)
	#oxycodine	261 (21.3)
	#oxy80s	213 (17.3)
	#oxycontin	215 (17.5)
	#oxicotin	233 (18.9)
	#oxicodone	212 (17.2)
	Total	321 (26.1)
Lysergic acid diethylamide	#LSD25	138 (11.2)
	#LSDtabs	130 (10.5)
	Total	213 (17.3)
3,4-Methylenedioxy-methamphetamine	#mdmapills	50 (4.1)
	#mdmaforsale	21 (1.7)
	#mdmazing	40 (3.3)
	#mdmaonline	21 (1.7)
	Total	94 (7.6)

[a] Drug name column relates to drug detected in the image and text of the post.

[b] Hashtag refers to the presence of a hashtag in a post detected.

[c] Posts is the number of posts with the hashtag and the percentage of total posts that contained the hashtag.

When manually annotating posts that contained hashtags, we found that the use of hashtags can be categorized into 3 major groups: (1) hashtags with controlled substance drug names and other illicit drug names, (2) slang or other codewords used in the Instagram user community for specific controlled substances and substance use behavior, and (3) other keywords and codewords describing selling and other promotional behavior (eg, shipping and selling). Examples for each of these categories are provided below:

- Drug name: Generic or brand name of a drug in a hashtag (eg, #xanax, #oxycodone, #oxycontin, #LSD, #MDMA)
- Codewords: Codewords for controlled substances, including (1) misspelled drug name, for example, “#zanax”, (2) extended drug name, for example, “#mdmaforsale”, “#2mgxanax”, and (3) street name of drug, for example, “#whitebar”
- Keywords related to sale or shipping, for example, “#forsale”, “#shipping”, etc

When assessing posts with descriptive text describing actions or behaviors of drug dealers, we were able to classify posts into 2 additional categories: (1) drug sale promoting language, for example, “interested in placing order without prescription”, “order now for quick delivery” and (2) contact information of purported drug seller. On the basis of these 2 categories, Instagram drug dealers appear to clearly indicate in their descriptive text of their posts an offer for sale and also provide other users with information on how to contact the seller. Contact information generally included an email address, phone number, or user account information about a communication-based application or mobile app. Among posts containing communication applications or apps, Wickr (445/1228, 36.23%), Telegram (245/1228, 19.95%), Kik app (225/1228, 18.32%), and WhatsApp (188/1228, 15.30%) were utilized. In some cases, a drug seller might also include descriptive text in the post referring to their account profile to reference contact information instead of providing it in the text of the post.

From the 1228 detected posts we analyzed, 232 of these posts explicitly included an offer for sale and offer to buy by the users. In this case, there was an Instagram post or comments within a post from a user offering to sell drug(s) (with contact information) and a comment from another user that asked for more information or offered to buy the drug.

Manual Image Annotation of Drug Dealer Posts

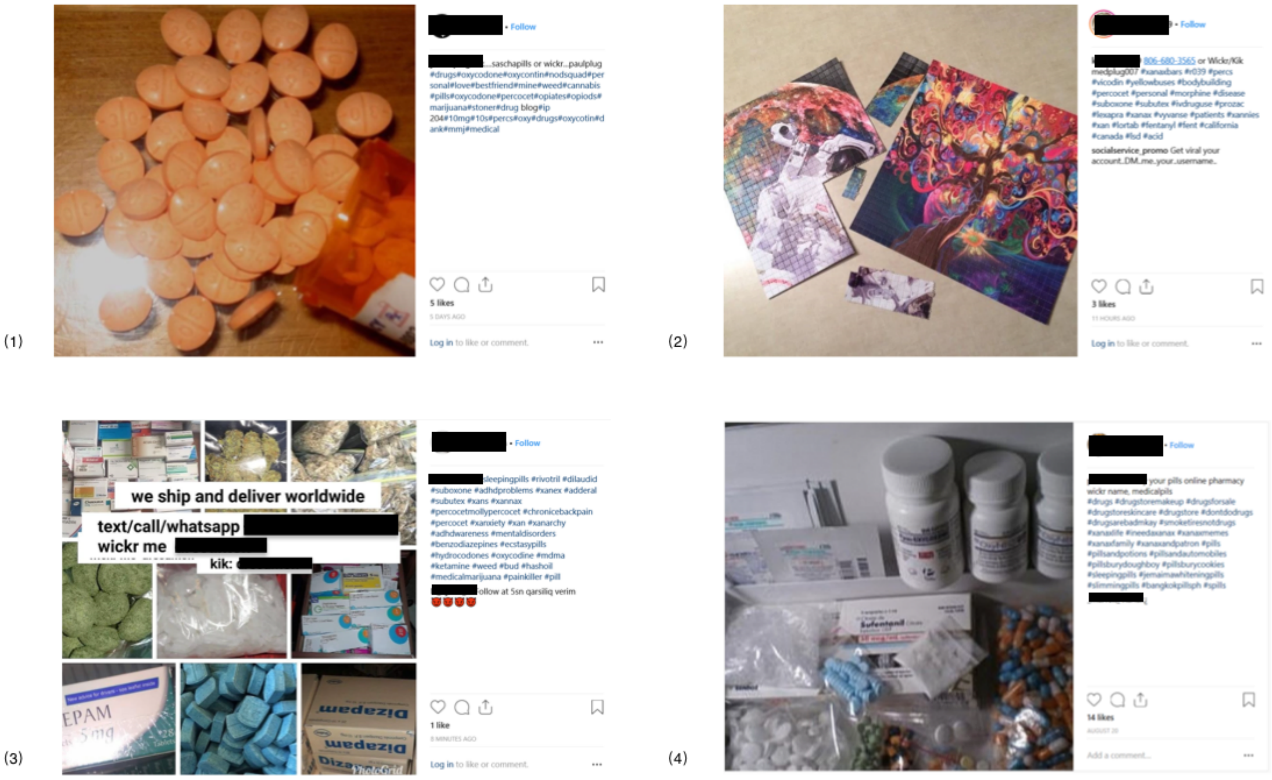
In addition to text analysis, we also collected a total of 260 pictures from the 1228 posts analyzed. These images were manually annotated to determine if they were related to controlled substances or illicit drugs (primarily coded for whether they included a picture of a controlled substance or suspected illicit drug product) or if they were unrelated images. A total of 252 images (252/260, 96.9%) included pictures or images of different types of drugs that can be categorized into 5 different categories.

In the first category which included prescription-controlled substances only, there were 175 posts (175/260, 67.3%) comprised of the following controlled substances: Xanax (41 posts), Alprazolam (34 posts), oxycodone/OxyContin (25 posts), Adderall (17 posts), and amphetamine and dextroamphetamine (14 posts). In the second category that contained images of illicit drugs only, there were 60 images comprising LSD (19 posts, including blotter paper soaked in LSD), ecstasy/MDMA (14 posts), cannabis (13 posts), and magic mushrooms (7 posts; see Figure 3).

In the third category, there were 15 images that contained both a prescription drug and illicit drug. In the fourth category, there were 5 images we could not classify, but which we suspected as either illicit drugs or other drug-related manufacturing materials (see Figure 4.)

Finally, in the fifth category, separate from images of drug products, we also detected images that included typed or written information on a physical medium communicating the drug dealers’ contact information (see Figure 5).

Figure 3. Examples of Instagram posts of illegal drug sale categories (user information and text from post have removed). (1) A post of prescription drugs; (2) a post of lysergic acid diethylamide (LSD); (3) a post with written contact information imbedded in the image; and (4) post with multiple drug types.



(1)

[redacted] • Follow

[redacted] ...saschapills or wickr...paulplug #drugs#oxycodone#oxycontin#nodsquad#personal#love#bestfriend#mine#weed#cannabis#pills#oxycodone#percocet#opiates#opiods#marijuana#stoner#drug blog#ip 204#10mg#10s#percs#oxy#drugs#oxycotin#dank#mmj#medical

5 likes

5 DAYS AGO

Log in to like or comment.

(2)

[redacted] • Follow

[redacted] 806-680-3565 or Wickr/Kik medplug007 #xanaxbars #r039 #percs #vicodin #yellowbuses #bodybuilding #percocet #personal #morphine #disease #suboxone #subutex #ivdruguse #prozac #lexapra #xanax #vyvanse #patients #xannies #xan #lortab #fentanyl #fent #california #canada #lsd #acid

socialservice_promo Get viral your account..DM..me..your..username..

3 likes

11 HOURS AGO

Log in to like or comment.

(3)

we ship and deliver worldwide

text/call/whatsapp [redacted]

wickr me [redacted]

kik: [redacted]

[redacted] • Follow

[redacted] sleepingpills #rivotril #dilaudid #suboxone #adhdproblems #xanex #adderal #subutex #xans #xannax #percocetmollypercocet #chronicebackpain #percocet #xanxiety #xan #xanarchy #adhdwareness #mentaldisorders #benzodiazepines #ecstasypills #hydrocodones #oxycodine #mdma #ketamine #weed #bud #hashoil #medicalmarijuana #painkiller #pill

[redacted] ollow at 5sn qarsiliq verim 😈😈😈😈

1 like

8 MINUTES AGO

Log in to like or comment.

(4)

[redacted] • Follow

[redacted] your pills online pharmacy wickr name, medicalpills #drugs #drugstoremakeup #drugsforsale #drugstoreskincare #drugstore #dontdodrugs #drugsarebadmkay #smoketiresnotdrugs #xanaxlife #ineedaxanax #xanaxmemes #xanaxfamily #xanaxandpatron #pills #pillsandpotions #pillsandautomobiles #pillsburydoughboy #pillsburycookies #sleepingpills #jemaimawhiteningpills #slimmingpills #bangkokpillsph #spills

[redacted]

14 likes

AUGUST 20

Add a comment...

Figure 4. Examples of Instagram posts with suspected drug products (user information removed). The 5 unclassified pictures include (1) clear capsules with white crystalline granules, (2) cups with pink liquid, (3) blue and white capsules with no drug identification, (4) plastic bags of blue crystals, and (5) a bag of white crystals with a label "B".

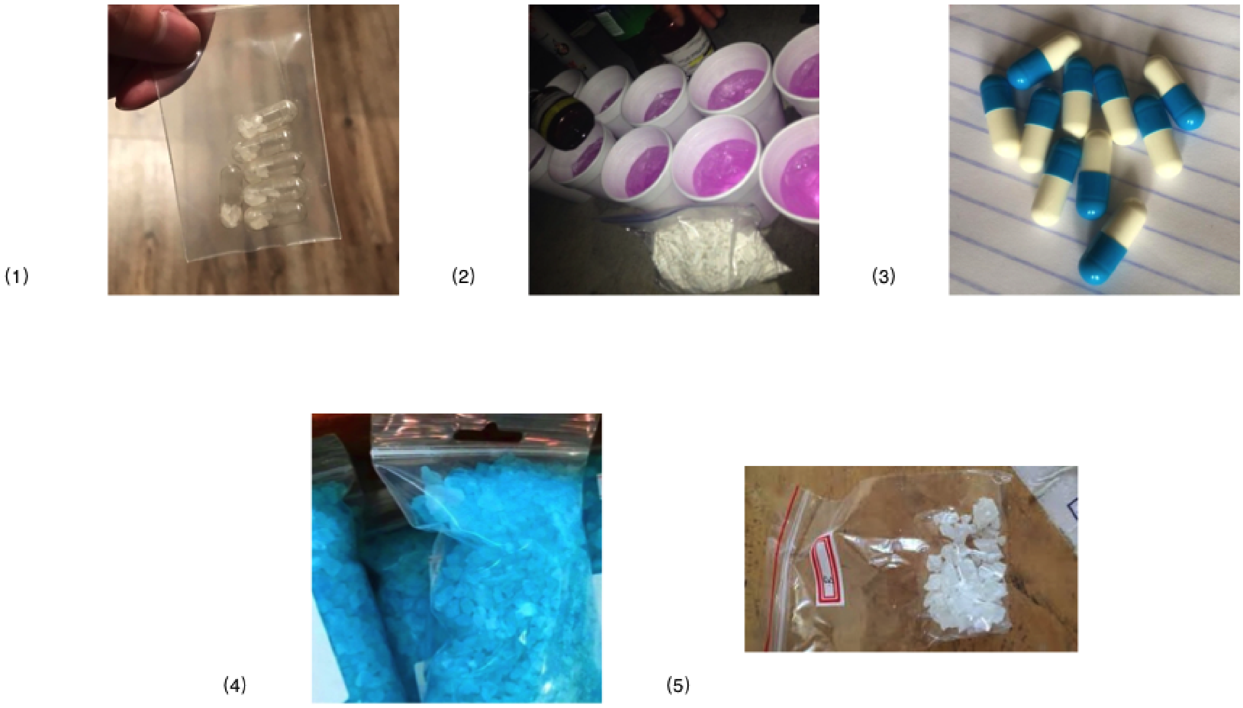
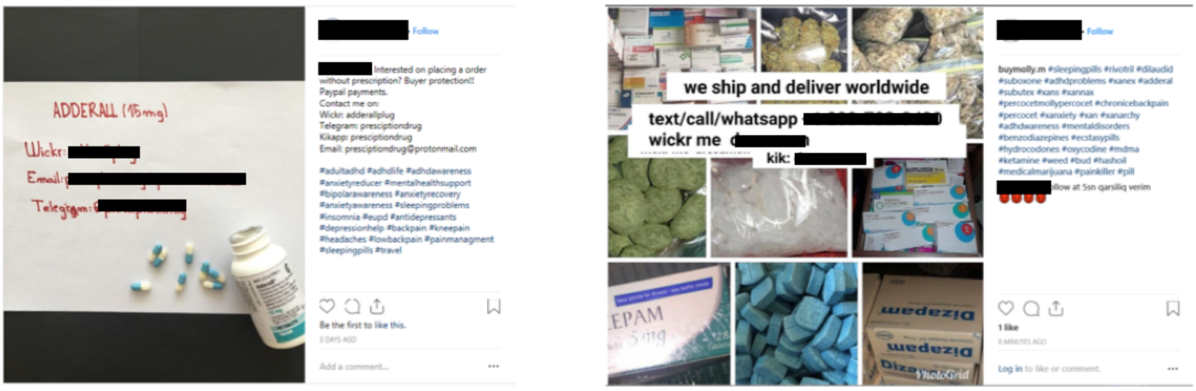


Figure 5. Examples of Instagram posts with written contact information. There were 60 images that included either typed or written contact information.



[redacted] Interested on placing a order without prescription? Buyer protection!!
Paypal payments.
Contact me on:
Wickr: adderallplug
Telegram: presciptiondrug
Kikapp: presciptiondrug
Email: presciptiondrug@protonmail.com

#adultadhd #adhdlife #adhdawareness #anxietyreducer #mentalhealthsupport #bipolarawareness #anxietyrecovery #anxietyawareness #sleepingproblems #insomnia #eupd #antidepressants #depressionhelp #backpain #kneepain #headaches #lowbackpain #painmanagment #sleepingpills #travel

Be the first to like this.
3 DAYS AGO
Add a comment...

buymolly.m #sleepingpills #rivotril #dilaudid #suboxone #adhdproblems #xanex #adderal #subutex #xans #xannax #percocetmollypercocet #chronicebackpain #percocet #xanxiety #xan #xanarchy #adhdwareness #mentaldisorders #benzodiazepines #ecstasypills #hydrocodones #oxycodine #mdma #ketamine #weed #bud #hashoil #medicalmarijuana #painkiller #pill

[redacted] llow at 5sn qarsiliq verim

1 like
8 MINUTES AGO
Log in to like or comment.

Discussion

Principal Findings

Over a 3-month period, we used a combination of hashtag keyword searches, Web scraping, deep learning, and manual annotation to detect and characterize 1228 posts from 267 unique users that we suspected were associated with illegal drug dealing on Instagram. Several prescription-controlled substances, illicit drugs, and other suspected drug products were detected as being offered for sale purportedly by drug sellers/dealers. In addition, we observed Instagram users having conversations via comments purporting to both offer a sale of a drug product and receiving an offer to buy from another user (ie, generally a comment from a seller to another user to contact them directly via a third-party communication application to enter into a drug-related transaction.)

These initial results are alarming and generally conform to existing nonscientific investigational news reports on the subject, in addition to published research on illegal drug sales on this and other social media platforms such as Twitter [6,7,17]. As a result of emerging evidence linking the risks of illicit drug access and diversion via social media technology there has been increased public attention and scrutiny. Congressional bodies, including the House Energy and Commerce Committee, have called on social media company executives to explain why their platforms facilitate this illegal activity [46,47]. This includes sharp inquiry from Congressman David McKinley (West Virginia), whose state has been heavily impacted by the opioid epidemic. Mr McKinley questioned both Twitter co-founder and Chief Executive Officer Jack Dorsey and Facebook Chairman and Chief Executive Officer Mark Zuckerberg in 2018 congressional testimony on what steps their platforms were taking to remove illegal opioid sellers and how they will protect the public [47,48]. Mr Zuckerberg responded that the enormity of data makes it hard to monitor content and that *artificial intelligence* approaches to *proactively* find content were needed [47].

In April 2018, Facebook and Instagram took action by blocking opioid-related hashtag searches on their platform and reportedly suspending accounts [16,20]. However, our study indicates that drug sellers continue to populate Instagram despite these actions and that these communities have changed their use of hashtags possibly to avoid detection. Hence, to carry out the legislative intent of the RHA to promote patient safety and prevent substance abuse behavior, there is clear need for innovative technology solutions that have high accuracy and are scalable and can help all parties (including technology companies, regulators, and law enforcement) detect, classify, and take action against digital drug dealers.

Study Limitations

Our study has certain limitations. First, this study was limited to a short period of data collection and we did not purposely sample Instagram accounts or consider user characteristics such as age, gender, or other demographics. Hence, study results are not generalizable and are not necessarily representative of the Instagram user community. In addition, demographic data are not always readily available in the metadata or the user account information or the post, and if available, may not be accurate. Future studies should assess which specific user communities may be at higher risk for illegal drug sourcing online. We also did not engage with users or verify if drugs purportedly being sold were actually available or sold to other users. Given this limitation, we cannot say with certainty that these drugs were actually being sold. Conducting test purchases of controlled substances and other illicit drugs is prohibited by federal law. However, drug dealers often post pictures of drug products to demonstrate to users that they have availability and we did not observe comments reporting scams or failed drug buys. We also did not use multimodal or synchronous approaches to develop a classification model based on both text and images as used by Yang and Luo, an approach that could improve performance of the model and should be explored in future studies. In addition, though data were collected and analyzed within 3 months of collection, we relied on manual annotation to establish validity of results and evaluate the performance of our model. This time lag because of manual annotation may have resulted in some posts being removed or modified before analysis or manual annotation. Specifically, drug dealers may self-delete posts after they have completed a transaction. Future studies should continue the iterative process of establishing training datasets to inform machine learning models that can more quickly and accurately detect illicit drug dealing.

Conclusions

In this study, we evaluated a deep learning model to detect drug dealers on Instgram. The deep learning model based on LSTM performed the best compared with the other 3 models evaluated. Furthermore, we compared the deep learning model's performance with hashtags against messages with only text in Instagram posts and demonstrated that the model yields better results from text without the hashtags, despite the risk of including false positives.

The results of our study further indicate that despite increased scrutiny by regulators and policymakers, the popular social media platform Instagram continues to act as a conduit for opioid, controlled substance, and illicit drug access, a direct violation of the RHA. Importantly, users have active conversations about selling and buying drugs, meaning that these social media posts act as digital marketplaces for drug dealing. Further action is needed to protect the public but needs to be carried out through meaningful collaboration and coordination involving partnership between technology companies, researchers, regulators, law enforcement, and impacted user communities to ensure that the opioid epidemic is not exacerbated by the digital risk environment.

Acknowledgments

TM would like to thank the US FDA for their invitation to present research associated with this study at the 2018 Online Opioid Summit and the FDA's leadership in this area.

Authors' Contributions

All authors jointly collected the data, designed the study, conducted the data analyses, and wrote the manuscript. All authors contributed to the formulation, drafting, completion, and approval of the final manuscript.

Conflicts of Interest

TM has received funding a pilot grant to research social media, illicit online pharmacies, and antibiotics from the Alliance for Safe Online Pharmacies, a 501(c)(4) social welfare organization engaged on the issue of illicit online pharmacies. TM is also the cofounder of S-3 Research, LLC, a company that received funding through the National Institutes of Health, National Institute on Drug Abuse—Start a SUD Startup Challenge award, and is also supported by an Accelerating Innovations to Market award provided by the University of California, San Diego—Jacobs School of Engineering Institute for the Global Entrepreneur. S-3 Research, LLC, has not received any nonpublic or industry funds related to its activity. TM and JL also received a pilot research grant from Google LLC to detect and characterize illegal drug dealing on YouTube. TM has previously received travel support from the World Health Organization and the US FDA to attend and participate in technical meetings regarding pharmaceutical supply chain and online drug safety. This includes travel support to attend and present at the FDA Online Opioid Summit as referenced in this paper. He has also acted as an expert consultant for the US Department of Justice on pharmaceutical diversion matters. None of these organizations had any involvement or input into the conception, design, collection, planning, conduct, analysis, interpretation, writing, and discussion of this work. Other coauthors have no conflicts or financial interests to disclose.

Multimedia Appendix 1

List of hashtag terms used.

[PDF File (Adobe PDF File), 114KB-Multimedia Appendix 1]

References

1. Food and Drug Administration. 2018. FDA Invites Internet Stakeholders to Opioid Summit URL: https://www.fda.gov/drugs/newsevents/ucm608633.htm [accessed 2019-02-19] [WebCite Cache ID 76JE4mYcD]
2. Mackey TK. Opioids and the internet: convergence of technology and policy to address the illicit online sales of opioids. Health Serv Insights 2018;11:1178632918800995 [FREE Full text] [doi: 10.1177/1178632918800995] [Medline: 30245569]
3. House of Representatives, Committee on Energy and Commerce. Centers for Disease Control and Prevention. 2019 Feb 19. Wide-ranging online data for epidemiologic research (WONDER) URL: https://wonder.cdc.gov/ [accessed 2019-02-18] [WebCite Cache ID 76JWOwzKf]
4. Mackey TK, Liang BA, Strathdee SA. Digital social media, youth, and nonmedical use of prescription drugs: the need for reform. J Med Internet Res 2013 Jul 26;15(7):e143 [FREE Full text] [doi: 10.2196/jmir.2464] [Medline: 23892156]
5. Congress.gov.: Congress.gov; 2008. Ryan Haight Online Pharmacy Consumer Protection Act of 2008 URL: https://www.congress.gov/110/plaws/publ425/PLAW-110publ425.pdf [accessed 2019-02-19] [WebCite Cache ID 76JEKQcEd]
6. Mackey T, Kalyanam J, Klugman J, Kuzmenko E, Gupta R. Solution to detect, classify, and report illicit online marketing and sales of controlled substances via Twitter: using machine learning and web forensics to combat digital opioid access. J Med Internet Res 2018 Dec 27;20(4):e10029 [FREE Full text] [doi: 10.2196/10029] [Medline: 29613851]
7. Mackey TK, Kalyanam J, Katsuki T, Lanckriet G. Twitter-based detection of illegal online sale of prescription opioid. Am J Public Health 2017 Dec;107(12):1910-1915. [doi: 10.2105/AJPH.2017.303994] [Medline: 29048960]

8. Dasgupta N, Freifeld C, Brownstein JS, Menone CM, Surratt HL, Poppish L, et al. Crowdsourcing black market prices for prescription opioids. J Med Internet Res 2013;15(8):e178 [FREE Full text] [doi: 10.2196/jmir.2810] [Medline: 23956042]
9. Pergolizzi JJ, LeQuang JB, Taylor RJ, Raffa R, NEMA Research Group. The "Darknet": the new street for street drugs. J Clin Pharm Ther 2017 Dec;42(6):790-792. [doi: 10.1111/jcpt.12628] [Medline: 28921578]
10. Broséus J, Rhumorbarbe D, Mireault C, Ouellette V, Crispino F, Décary-Hétu D. Studying illicit drug trafficking on Darknet markets: structure and organisation from a Canadian perspective. Forensic Sci Int 2016 Jul;264:7-14. [doi: 10.1016/j.forsciint.2016.02.045] [Medline: 26978791]
11. Drug Enforcement Administration (DEA). Drug Enforcement Administration. Counterfeit Prescription Pills Containing Fentanyls: A Global Threat URL: https://www.dea.gov/sites/default/files/docs/Counterfeit%2520Prescription%2520Pills.pdf [accessed 2019-02-19] [WebCite Cache ID 76JXRyq3H]
12. Lytvynenko J. Buzzfeednews. 2018 Jun 1. Social Networks Are Losing A Deadly Battle With Illegal Online Pharmacies URL: https://www.buzzfeednews.com/article/janelytvynenko/social-networks-are-losing-a-deadly-battle-with-illegal [accessed 2019-02-19] [WebCite Cache ID 76JTBFyAp]
13. Smith A, Anderson M. Pew Internet. 2018 Mar 1. Social Media Use in 2018 URL: http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/ [accessed 2019-02-19] [WebCite Cache ID 76JTO6qoa]
14. United States Government Accountability Office. 2013 Jul. Internet Pharmacies Federal Agencies and States Face Challenges Combating Rogue Sites, Particularly Those Abroad URL: http://www.gao.gov/assets/660/655751.pdf [accessed 2019-02-19] [WebCite Cache ID 76JTgOf4I]
15. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
16. Constine J. Techcrunch. 2018. Facebook cracks down on opioid dealers after years of neglect URL: https://techcrunch.com/2018/08/16/facebook-opioid-searches/ [accessed 2019-02-19] [WebCite Cache ID 76JTu0QWS]
17. Yang X, Luo J. Tracking illicit drug dealing and abuse on Instagram using multimodal analysis. ACM Trans Intell Syst Technol 2017 Jul 17;8(4):1-15. [doi: 10.1145/3011871]
18. Lapowsky I. WIRED. Tech companies deflect blame for opioid crisis ahead of FDA Summit URL: https://www.wired.com/story/tech-companies-deflect-blame-for-opioid-crisis-fda-summit/ [accessed 2019-02-19] [WebCite Cache ID 76JU9rYup]
19. Dwoskin E. The Washington Post. 2018 Sep 25. Instagram has a drug problem. Its algorithms make it worse URL: https://tinyurl.com/y65rvu6w [accessed 2019-02-19] [WebCite Cache ID 76JUMWJbR]
20. Tiku N. WIRED. 2018 Apr 6. One woman got facebook to police opioid sales on instagram URL: https://www.wired.com/story/one-woman-got-facebook-to-police-opioid-sales-on-instagram/ [accessed 2019-02-19] [WebCite Cache ID 76JUWUVmi]
21. Mackey TK, Kalyanam J. Detection of illicit online sales of fentanyls via Twitter. F1000Res 2017;6:1937 [FREE Full text] [doi: 10.12688/f1000research.12914.1] [Medline: 29259769]
22. Hanson C, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. J Med Internet Res 2013 Apr 17;15(4):e62 [FREE Full text] [doi: 10.2196/jmir.2503] [Medline: 23594933]
23. Hu H, Phan N, Geller J, Iezzi S, Vo H, Dou D. arXiv. 2019 Apr 3. An Ensemble Deep Learning Model for Drug Abuse Detection in Sparse Twitter-Sphere URL: https://arxiv.org/pdf/1904.02062.pdf [accessed 2019-06-05] [WebCite Cache ID 78ucr60Ib]
24. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-related posts on Instagram. Prev Sci 2016 Aug;17(6):710-720. [doi: 10.1007/s11121-016-0669-9] [Medline: 27262456]
25. Cavazos-Rehg P, Krauss M, Fisher S, Salyer P, Grucza R, Bierut L. Twitter chatter about marijuana. J Adolesc Health 2015 Feb;56(2):139-145 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.270] [Medline: 25620299]
26. Katsuki T, Mackey TK, Cuomo R. Establishing a link between prescription drug abuse and illicit online pharmacies: analysis of Twitter data. J Med Internet Res 2015 Dec 16;17(12):e280 [FREE Full text] [doi: 10.2196/jmir.5144] [Medline: 26677966]
27. Hanson C, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013 Sep 6;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
28. Zhou Y, Sani N, Lee C, Luo J. arXiv. 2016. Understanding Illicit Drug Use Behaviors by Mining Social Media URL: https://arxiv.org/pdf/1604.07096.pdf [accessed 2019-06-05] [WebCite Cache ID 78ueLIogr]
29. Huynh T, He Y, Willis A, Rüger S. Adverse Drug Reaction Classification With Deep Neural Networks. 2016 Presented at: COLING 2016; December 13-16, 2016; Osaka, Japan p. 877-887 URL: https://coling2016.anlp.jp/doc/main.pdf
30. Wu C, Wu F, Liu J, Wu S, Huang Y, Xie X. Detecting Tweets Mentioning Drug Name and Adverse Drug Reaction with Hierarchical Tweet Representation and Multi-Head Self-Attention. 2018 Presented at: 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task. 2018; October 31–November 1, 2018; Brussels, Belgium p. 34-37 URL: https://www.aclweb.org/anthology/W18-5909
31. Constine J. TechCrunch. 2018. Dealers remain on Instagram as it pushes drug searchers to treatment URL: https://techcrunch.com/2018/09/12/instagram-drugs/ [accessed 2019-06-05] [WebCite Cache ID 78ufyAbLT]
32. Quinlan J. Induction of decision trees. Mach Learn 1986 Mar;1(1):81-106. [doi: 10.1007/BF00116251]
33. Breiman L. Random forests. Mach Learn 2001 Oct;45(1):5-32 [FREE Full text]

34. Wang L. Support Vector Machines: Theory And Applications (Studies In Fuzziness And Soft Computing). Berlin: Springer; 2005.
35. Joachims T. Text categorization with Support Vector Machines: Learning with many relevant features. In: Machine Learning: ECML-98. Berlin, Heidelberg: Springer-Verlag New York, LLC; 1998:137-142.
36. Màrquez L, Rodríguez H. Part-of-speech tagging using decision trees. In: Machine Learning: ECML-98. Berlin, Heidelberg: Springer-Verlag New York, LLC; 1998:25-36.
37. Xu B, Guo X, Ye Y, Cheng J. An improved random forest classifier for text categorization. J Comput 2012 Dec 1;7(12):2913-2920 [FREE Full text] [doi: 10.4304/jcp.7.12.2913-2920]
38. Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. Recurrent Neural Network Based Language Model. 2010 Sep Presented at: INTERSPEECH 2010; September 26-30, 2010; Makuhari.
39. Sundermeyer M, Schlüter R, Ney H. LSTM Neural Networks for Language Modeling. 2012 Sep Presented at: INTERSPEECH 2012; September 9-13, 2012; Portland.
40. Chen Z, Watanabe S, Erdogan H, Hershey J. Speech Enhancement and Recognition Using Multi-Task Learning of Long Short-Term Memory Recurrent Neural Networks. 2015 Presented at: INTERSPEECH 2015; September 6-10, 2015; Dresden.
41. Weninger F, Erdogan H, Watanabe S, Vincent E, Roux J, Hershey J, et al. Speech enhancement with LSTM recurrent neural networksits application to noise-robust ASR. In: Latent Variable Analysis and Signal Separation. Cham: Springer, Cham; 2015:91-99.
42. Lin M, Chen Q, Yan S. arXiv. 2013. Network in network URL: https://arxiv.org/pdf/1312.4400.pdf [accessed 2019-06-05] [WebCite Cache ID 78uliSakV]
43. de Boer P, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. Ann Oper Res 2005 Feb;134(1):19-67. [doi: 10.1007/s10479-005-5724-z]
44. Vogl TP, Mangis JK, Rigler AK, Zink WT, Alkon DI. Accelerating the convergence of the back-propagation method. Biol Cybern 1988 Sep 1;59(4-5):257-263. [doi: 10.1007/BF00332914]
45. Kingma D, Ba J. arXiv. 2014 Dec. Adam: A method for stochastic optimization URL: https://arxiv.org/pdf/1412.6980.pdf [accessed 2019-06-05] [WebCite Cache ID 78um3S6jI]
46. Portman R, Carper C. United State Senate. Combatting the opioid crisisxploiting Vulnerabilities in international mail URL: https://tinyurl.com/y5ky63b3 [accessed 2019-02-19] [WebCite Cache ID 76JUsthpf]
47. Zamora. House of Representatives, Committee on Energy and Commerce. Washington, D.C; 2018 Apr 11. Facebook: Transparency and use of consumer DATA URL: https://docs.house.gov/meetings/IF/IF00/20180411/108090/HHRG-115-IF00-Transcript-20180411.pdf [accessed 2019-02-19] [WebCite Cache ID 76JW9H39W]
48. O'Brien S. CNN Business. 2018 Apr 11. Mark Zuckerberg takes heat over illegal opioid sales on Facebook URL: https://tinyurl.com/y5h4bx6u [accessed 2019-02-17] [WebCite Cache ID 76JVEI7Cp]

Abbreviations

API: application programming interface
AUC: area under the curve
FCN: fully connected network
FDA: Food and Drug Administration
JSON: JavaScript Object Notation
LSD: lysergic acid diethylamide
LSTM: long short-term memory
MDMA: 3,4-methylenedioxy-methamphetamine
RF: random forest
RHA: Ryan Haight Online Pharmacy Consumer Protection Act, 2008
SVM: support vector machine

Edited by G Eysenbach; submitted 22.02.19; peer-reviewed by A Nara, J Kim, N Okazaki, N Miyoshi, Z Huang, JT Pollettini; comments to author 20.03.19; revised version received 12.05.19; accepted 27.05.19; published 15.06.19

Please cite as:
Li J, Xu Q, Shah N, Mackey TK
A Machine Learning Approach for the Detection and Characterization of Illicit Drug Dealers on Instagram: Model Evaluation Study
J Med Internet Res 2019;21(6):e13803
URL: http://www.jmir.org/2019/6/e13803/
doi: 10.2196/13803
PMID: 31199298