



Detecting associations between dietary supplement intake and sentiments within mental disorder tweets

Health Informatics Journal

1–13

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1460458219867231

journals.sagepub.com/home/jhi**Yefeng Wang** 

University of Minnesota, USA

Yunpeng Zhao 

University of Florida, USA

Jianqiu Zhang

University of Minnesota, USA

Jiang Bian 

University of Florida, USA

Rui Zhang

University of Minnesota, USA

Abstract

Many patients with mental disorders take dietary supplement, but their use patterns remain unclear. In this study, we developed a method to detect signals of associations between dietary supplement intake and mental disorder in Twitter data. We developed an annotated dataset and trained a convolutional neural network classifier that can identify language use pattern of dietary supplement intake with an F1-score of 0.899, a precision of 0.900, and a recall of 0.900. Using the classifier, we discovered that melatonin and vitamin D were the most commonly used supplements among Twitter users who self-diagnosed mental disorders. Sentiment analysis using Linguistic Inquiry and Word Count has shown that among Twitter users who posted mental disorder self-diagnosis, users who indicated supplement intake are more active

Corresponding author:

Rui Zhang, Institute for Health Informatics and College of Pharmacy, University of Minnesota, 8-100 Phillips-Wangensteen Building, 516 Delaware Street SE, Minneapolis, MN 55455, USA.

Email: zhan1386@umn.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

and express more negative emotions and fewer positive emotions than those who have not mentioned supplement intake.

Keywords

dietary supplement, mental health, natural language processing, sentiment analysis, social media

Introduction

Mental disorders are the strongest contributor to disability worldwide.¹ The four major mental disorders that can lead to disability are major depression, bipolar disorder, schizophrenia, and obsessive-compulsive disorder (OCD). The incidence of mental disorders is high-in the United States, where 25 percent of US adults have been diagnosed with at least one mental disorder.² While mental disorders can seriously affect the quality of personal life and impose a significant burden on public health, their diagnosis and treatment are surprisingly difficult.³ Without reliable and quantifiable results from laboratory tests, the diagnosis of mental disorders has to rely on listening to patients' own recounts, observing patients' behaviors, and referencing mental status examination reports.⁴ Most mental disorders are typically treated with prescription drugs. For example, bipolar disorder is typically treated with lithium. However, these drugs are not always effective and may cause serious adverse reactions.^{1,2} Many mental disorder patients often take dietary supplements as an alternative treatment strategy.⁵ Therefore, learning what type of dietary supplements that mental disorder patients usually take and how these dietary supplements affect their daily life is an important first step for both patients and their providers toward adequate treatments and management of their mental disorders.

Approximately 69 percent of the public use some type of social media according to a 2018 Pew Research Center report.⁶ 24 percent of US adults use Twitter—a popular social media platform on which users interact with short messages known as “tweets.” Social media platforms like Twitter are not only changing how the public communicate with each other but also shifting the landscape of health communication. Twitter is used by both consumers and health professionals to obtain and share health information. Consumers want their voices to be heard and voluntarily share a critical mass of their health information including their personal health experience and express opinions toward health issues and health care services. The massive amount of Twitter data provides an opportunity to identify potential markers of mental disorders from users' tweets and develop a surveillance system to track their mental health status.⁷ De Choudhury et al.³ explored the possibility of predicting depression by analyzing Twitter data collected from major depressive disorder (MDD) patients who volunteered to participate in their study. Factors such as active hours on Twitter, the sentiment of the tweets, and the linguistic style of the tweets were all considered as predictive features of MDD. They successfully built a classifier based on these features to predict “whether an individual is vulnerable to depression” before the onset of MDD. Coppersmith et al.⁴ collected tweets that contain self-diagnosis of post-traumatic stress disorder (PTSD), depression, bipolar disorder, and seasonal affective disorder (SAD) and subsequently analyzed the tweeting history of these users. Significant differences in the language use patterns were found between patients with mental disorders and the control groups, which can be used as predictive markers of mental disorders. However, to the best of our knowledge, no studies have attempted to study the influence of dietary supplement intake on users' mental health conditions using Twitter data.

Motivated to fill this knowledge gap, we designed the study to detect the associations between mental disorders and dietary supplement intake using Twitter data. We hypothesized that the co-occurrence of dietary supplements and mental disorder hashtags within a user's tweeting history

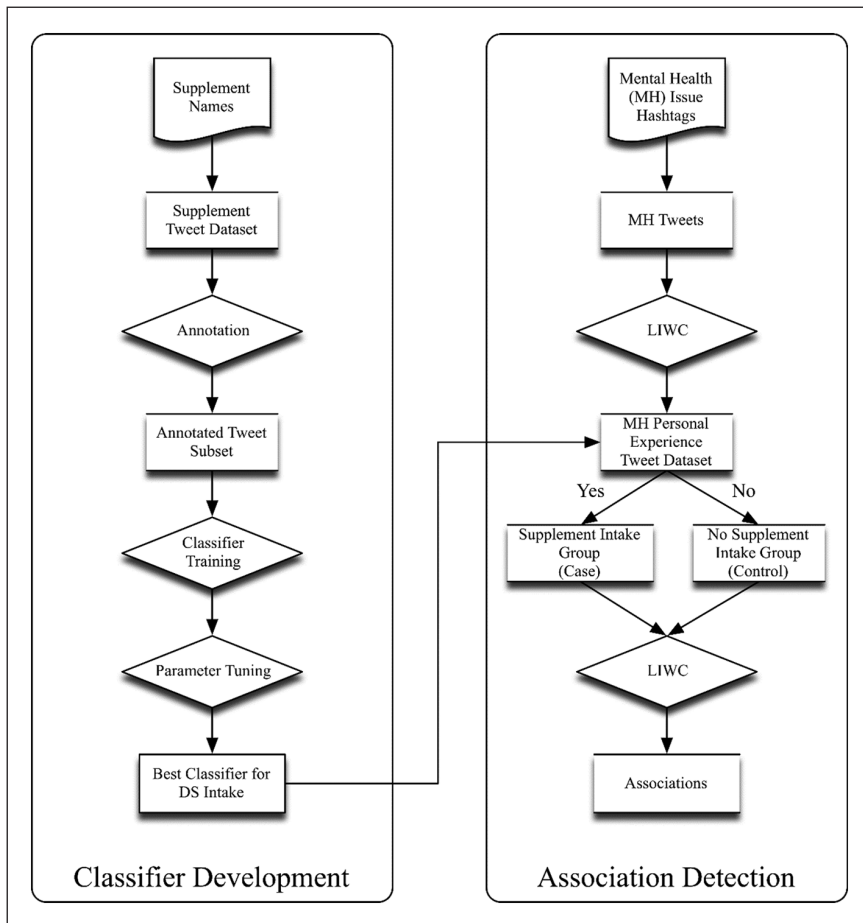


Figure 1. An overview of the steps and methods used in this study.

can be regarded as a signal of the associations between dietary supplement intake and mental disorders. The aim of this article is twofold: (1) to identify the most commonly used dietary supplements among mental disorder patients and (2) to analyze the associations between dietary supplement intakes and mental disorders.

Methods

The overview of the steps and methods used in this study is shown in Figure 1.

Data collection

Collecting tweets with mentions of dietary supplement names. We have compiled a list of 31 dietary supplements^{2,5} that are commonly used to improve health conditions of mental disorder patients (Table 1). Due to the noisy nature of tweet data, dietary supplements were usually mentioned in various forms in user tweets. Vector representation of words derived from a massive corpus can provide a measure of semantic similarity and group similar words,⁸ which can help identify potential

Table 1. List of supplements used to collect the supplement dataset.

5-HTP	Niacin	Vitamin A
Biotin	Pantothenic acid	Vitamin B12
Citicoline	Pyridoxine	Vitamin C
DHEA	Rhodiola	Vitamin D
Fish oil	Riboflavin	Vitamin D2
Folic acid	SAMe	Vitamin D3
GABA	St. John's Wort	Vitamin E
Gingko	Theanine	Vitamin K1
Inositol	Thiamine	Vitamin K2
Kava	Tryptophan	
Melatonin	Valerian	

synonyms. To retrieve relevant synonyms in tweets for these supplements, we adapted four word vectors pre-trained with various corpora, including (1) the Word2Vec vectors trained on Google News dataset (100 billion tokens),⁹ (2) the fastText vectors trained on Common Crawl (600 billion tokens) data,¹⁰ (3) the GloVe vectors trained on Common Crawl (840 billion tokens) data,¹¹ and (4) the Word2Vec vectors trained on Wikipedia and PubMed (5.5 billion tokens) data.¹² In this study, we quantified the semantic similarity between two words by the cosine similarity between their word vectors. A higher cosine similarity between two word vectors indicates that they are semantically more related. However, it does not guarantee that one word is a synonym to the other. For example, one word could also be an antonym, a hypernym, or a hyponym of the other word.¹³ Therefore, we manually evaluated the results returned by pre-trained word vectors to find the synonyms.

For each supplement, 160 synonym candidates were generated by collecting the top 40 words ranked by the cosine similarity from each of the four word vectors. Each candidate word was then used as a query keyword in Twitter search Application Programming Interface (API), and the first retrieved 20 tweets were manually checked for synonyms. Both the candidate and the new synonyms found in the tweet samples were added to the final synonym list, containing 332 supplement keywords in total.

The tweets that contain at least one supplement keyword between 28 May 2018 to 29 June 2018 were collected with a crawler we developed previously¹⁴ using the Twitter streaming API. We denote this tweet dataset as the "supplement dataset," which we used to develop methods for identifying Twitter users' dietary supplement intake.

Collecting tweets by mental health hashtags. In order to identify Twitter users with mental disorders, we collected tweets from 1 January 2015 to 31 December 2017 that contain 25 hashtags related to depression, anxiety, PTSD, OCD, schizophrenia, and bipolar disorder. Linguistic Inquiry and Word Count (LIWC)¹⁵ was used to analyze the first-person narrative components in each tweet. LIWC is a popular text analysis tool to reveal people's thoughts, feelings, personality, and sentiments based on a given text written by the user. It has a set of pre-defined word categories that correspond to different psychometric properties, for example, different affect processes including positive and negative emotions, anger, and anxiety. LIWC calculates the percentage of the words that belongs to each pre-defined word category, and thus can provide clues on the user's mental status (e.g. sentiment) and whether a tweet is about the user's personal experience.

For example, a high-percentage of first-person pronoun use in a tweet implies that the tweet is describing personal experience. If 10 percent of the words in a tweet belong to the first-person pronoun word subcategory according to LIWC, the user of this tweet is considered to have

self-diagnosed mental disorders.¹⁶ The likelihood of simultaneous mentions of mental disorders and dietary supplement within the same tweet is low. Therefore, we have collected the most recent 3200 tweets (limited by Twitter API) of these users to detect the associations between dietary supplement intake and mental disorders. We will denote this tweet dataset as “mental health dataset,” used for analyzing the association between supplement intake and mental health.

Tweet preprocessing. Only English tweets were used to train the Tweet2Vec model to obtain the vector representation of the tweets. The English tweets can be readily picked out according to the *lang* field of the Twitter API. All the hyperlinks, user handles (in the form of @username), emojis, and non-ASCII characters were removed from the tweets. Repeating punctuation marks were reduced to a single one; elongated words were transformed into the original form, for example, “haaaappppppy” was transformed into “happy”; contractions were expanded to the full form, for example, “can’t” was transformed into “cannot.” The tweets were all lowercased. Retweets and duplicates were removed. After preprocessing, 234,053 tweets remained in the supplement dataset, while 5,291,991 tweets from 3757 users remained in the mental health dataset.

Annotation

In this study, the recognition of supplement intake information within a tweet was framed as a binary classification problem, that is, containing supplement intake information or not. In order to train a classifier, an annotated tweet dataset was generated with a two-step procedure.

First, we randomly selected 2000 tweets from supplement dataset and labeled the tweet as “supplement intake” and “no supplement intake.”¹⁷ A “supplement intake” tweet should describe the following:

1. The user personally took the supplement. For example, *“i took melatonin at 10 then slept at 12 and woke up like a little less than 3 hours later and im so out of it but i need to wake up in 2 hours and honestly i just want to perish.”*
2. The user was going to take the supplement in the near future. For example, *“im boutta go get some melatonin pills.”*
3. The user had taken the supplement in the past or was taking the supplement on a regular basis. For example, *“I developed insomnia while doing chemo. Melatonin worked wonders for me,”* and *“Apparently melatonin stops being effective after you take it for long periods of time and it looks like I’m never going to fall asleep ever again.”*

A “no supplement intake” tweet, on the contrary, usually contains the following:

1. The user is giving other users suggestions on taking dietary supplement. For example, *“@username But remember, don’t forget to eat your vitamin and don’t overwork yourself. Kay?”*
2. The tweet is an advertisement for a supplement product. For example, *“How about a flash sale??? 100% daily vitamin, konjac sponge, toothpaste, and 2 Aluminum Chlorohydrate free antiperspirants. Only \$27.50 til midnight tonight!!!”*
3. The tweet is a single mention of a supplement, which provides too few clues to imply supplement use. For example, *“@username Melatonin!”*
4. Other irrelevant tweets. For example, the abbreviation “EPA” could both stand for the omega-3 fatty acid supplement “Eicosapentaenoic acid” and “United States Environmental Protection Agency.” The tweets that matched the latter would be irrelevant.

Two annotators reviewed these 2000 tweets and the inter-rater agreement was calculated based on the first 1000 tweets. Approximately 90 percent of these tweets belong to the “no supplement intake” class. The class imbalance has to be dealt with before training the classifier, and thus, a second step was taken to find more “supplement intake” tweets from the supplement dataset.

We hypothesized that the tweets which belong to the “supplement intake” class present similar patterns of language use. Inspired by the concept of word embeddings, we managed to obtain a character-level 512-dimension vector representation of the entire supplement dataset using an improved version of the Tweet2Vec model.¹⁸ The similarity between the tweets can be readily quantified by calculating the cosine similarity between their vector representations. We retrieved the 20 most similar tweets for each of the 217 tweets that were labeled as “supplement intake” in the initial annotation and annotated 3000 of them. Different tweets may share the same candidate as the most similar tweets; therefore, duplicates were removed. The resulting dataset was combined with the annotated dataset from the first step of annotation, and the final annotated dataset containing 4210 tweets, among which 993 tweets belong to the “supplement intake” class, was used to train the classifier.

Development and evaluation of supplement intake classifiers

We used a convolutional neural network (CNN) classifier to identify “supplement intake” tweets from the mental health dataset. Random forest, support vector machine (SVM), and multi-layer perceptron (MLP) served as the baseline for performance comparison. The optimal hyperparameters that give the highest F1-score were determined by 10-fold cross validation on the supplement intake dataset for each classifier. We tuned the number of trees, the number of features to consider when looking for the best split, and the minimum number of samples required to be at a leaf node for the random forest classifier; the kernel type (linear and radial-basis function) and the penalty parameter of the error term C for the SVM classifier; and the number of hidden layers for the MLP classifier.

Analyzing association between supplement intake and mental health

The best classifier was applied to the mental health dataset. Applying this classifier to the mental health dataset yields a subset where all the tweets were classified as “supplement intake.” We matched the subset to the supplement name list used in Data Collection section and found 173 users who posted both a mental health issue hashtag and a supplement in their tweeting history. We called these 173 users as “case group.”

The distributions of dietary supplements, mental health issue hashtags, and the pair of dietary supplement and health issue hashtags were explored. The 25 mental health issue hashtags can be classified into seven categories: anxiety, depression, bipolar disorder, PTSD, OCD, schizophrenia, and general mental health issues where no specific mental disorders were mentioned. Table 2 listed some hashtag examples.

In order to analyze how dietary supplement intake affects the emotions of mental disorder patients, we also performed sentiment analysis with LIWC. We randomly selected 692 (4 times of 173) users from the remaining mental health dataset who only posted one or more mental health issue hashtags but no supplement in their tweeting history as the “control group.” We compared the word count percentage of positive emotion and negative emotion word subcategories in the tweets between the case and the control group. For negative emotions, we specifically compared emotions of anxiety, stress, and anger. A two-sample independent t test was performed to verify if the mean

Table 2. Examples of mental health issue hashtags.

Mental disorder category	Example hashtags
Anxiety	#anxiety, #anxietyproblems
Depression	#depressed, #depression
Bipolar disorder	#bipolar, #bipolar disorder
PTSD	#PTSD, #CPTSD
OCD	#OCD
Schizophrenia	#schizophrenia
General mental health issues	#mentalillness, #mentalhealthmatters, #sicknotweak

PTSD: post-traumatic stress disorder; OCD: obsessive-compulsive disorder.

Table 3. Performance and the optimal hyperparameters of the classifiers used in this study.

Classifier	<i>F</i> -score	Precision	Recall	Optimal hyperparameters
Random Forest	0.765	0.788	0.744	300 trees, bi-gram model at least 1 sample at leaf nodes
SVM	0.781	0.779	0.785	Linear kernel, tri-gram model, $C = 1$
MLP 1 Hidden Layer	0.858	0.863	0.866	8 epochs, 32-tweet batches
MLP 2 Hidden Layers	0.869	0.871	0.875	10 epochs, 32-tweet batches
MLP 3 Hidden Layers	0.861	0.864	0.869	10 epochs, 32-tweet batches
CNN Classifier	0.899	0.900	0.900	7 epochs, 32-tweet batches, convolution filters of size ranging from 3 to 8

SVM: support vector machine; MLP: multi-layer perceptron; CNN: convolutional neural network.

word count percentage of each emotion category is significantly different. We further compared sentiments between case group and control group within each mental disorder category.

Results

Annotation

After the two annotators had resolved the disagreement on the first 1000 annotations, the inter-rater agreement measured by Cohen's Kappa was 0.967. Among the initial 2000 annotations, only 217 tweets were related to supplement intake (10.85%). We used the Tweet2Vec model to find the 20 most similar tweets to each of these "supplement intake" tweets and annotated 3000 of them. In all, 1049 tweets were related to supplement intake in this Tweet2Vec expanded subset (34.97%). This indicated that the Tweet2Vec model was successful in up-sampling the "supplement intake" tweets from the supplement dataset. The expanded subset was combined with the initial annotations, and a final annotated tweet dataset containing 4210 unique tweets was prepared for classifier training, among which 993 were "supplement intake" tweets (23.6%).

Performances of classifiers

Table 3 showed the *F*-score, precision, recall, and optimal hyperparameters of the random forest, SVM, MLP, and CNN classifiers. For random forest and SVM classifiers, we used term frequency-inverse document frequency weighted *n*-gram features for word representations. For MLP and

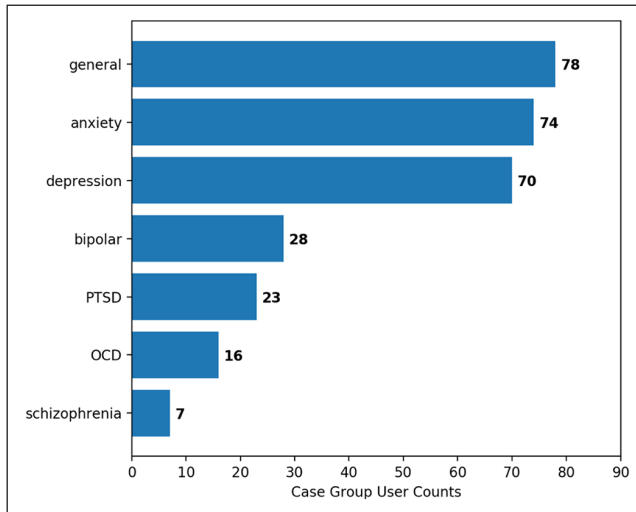


Figure 2. The distribution of mental health issue hashtags among the case group users.

CNN classifier, we trained word embeddings on-the-fly in the network architecture. The 10-fold cross-validation results have shown that the CNN classifier has yielded the best F1-score.

Distribution of the signals of association between supplement intakes and mental disorders

Figure 2 presented the distribution of mental health issue hashtags among the case group users. Besides the general mental health issue mentions, anxiety and depression are the two mental disorders that are frequently mentioned by 74 and 70 users, respectively.

Figure 3 presented the distribution of supplement mentions among the case group users. Melatonin, taken by 116 users, was the most frequently used supplement, followed by vitamin D, vitamin C, and fish oil.

Table 4 listed the 10 most common associations between mental health disorders and supplements mentioned by the case group users.

Figure 4 presented the results of LIWC sentiment analysis of the tweets posted by the case and the control group. Violin plots were used instead of boxplots to show the distribution and the mean of the word count percentage in each emotion category. A significant difference was found in the emotion category tagged with an asterisk ($\alpha=0.05$, $p < 10^{-3}$). Figure 4 demonstrated that the case group expressed fewer positive emotions and more negative emotions in general than the control group. Among the negative emotion, the case group expressed more anxiety than the control group, while the control group expressed more anger than the case group. The sentiment of the tweets posted by each mental disorder category within the case group was also compared to the control group. A significant difference was found in the emotion category tagged with an asterisk for each group ($\alpha=0.05$). The p values that are larger than 10^{-3} were shown in the labels. All mental disorder groups have expressed fewer positive emotions than the control group. Especially, all mental disorder groups have expressed more anxiety than the control group. However, the control group can express as much anger and sadness as the case groups, and sometimes even more anger than the case groups.

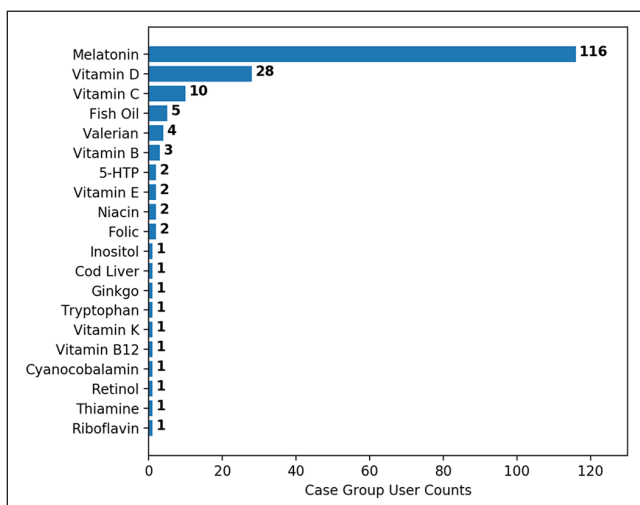


Figure 3. Distribution of supplement mentions among the case group users.

Table 4. Most common associations between mental health disorders and supplements.

Mental disorders	Supplements	Number of case group users mentioning the association
Anxiety	Melatonin	30
General	Melatonin	22
Depression	Melatonin	21
Bipolar disorder	Melatonin	7
OCD	Melatonin	7
PTSD	Melatonin	6
Anxiety	Vitamin D	4
Anxiety	Vitamin C	2
PTSD	Vitamin D	2
Schizophrenia	Melatonin	2

OCD: obsessive-compulsive disorder; PTSD: post-traumatic stress disorder.

Discussion

Associations between supplement intake and mental disorders

In this study, we define the associations between supplement intake and mental disorders from two perspectives. First, we explored the most common supplements that Twitter users who self-diagnosed mental disorders would take. Second, we analyzed the sentiment of the tweets between case group and control group (previously defined). Since this study only relies on the information from the tweets, we hypothesized that the mental health issue hashtag can serve as an approximate to a self-diagnosis of mental disorder for the patients.

As shown in Figure 3 and Table 4, the most common supplement taken by the users who self-diagnosed mental disorders was melatonin. Melatonin was mentioned by 116 case group users, and they have self-reported mental disorders including anxiety, depression, bipolar disorder, PTSD,

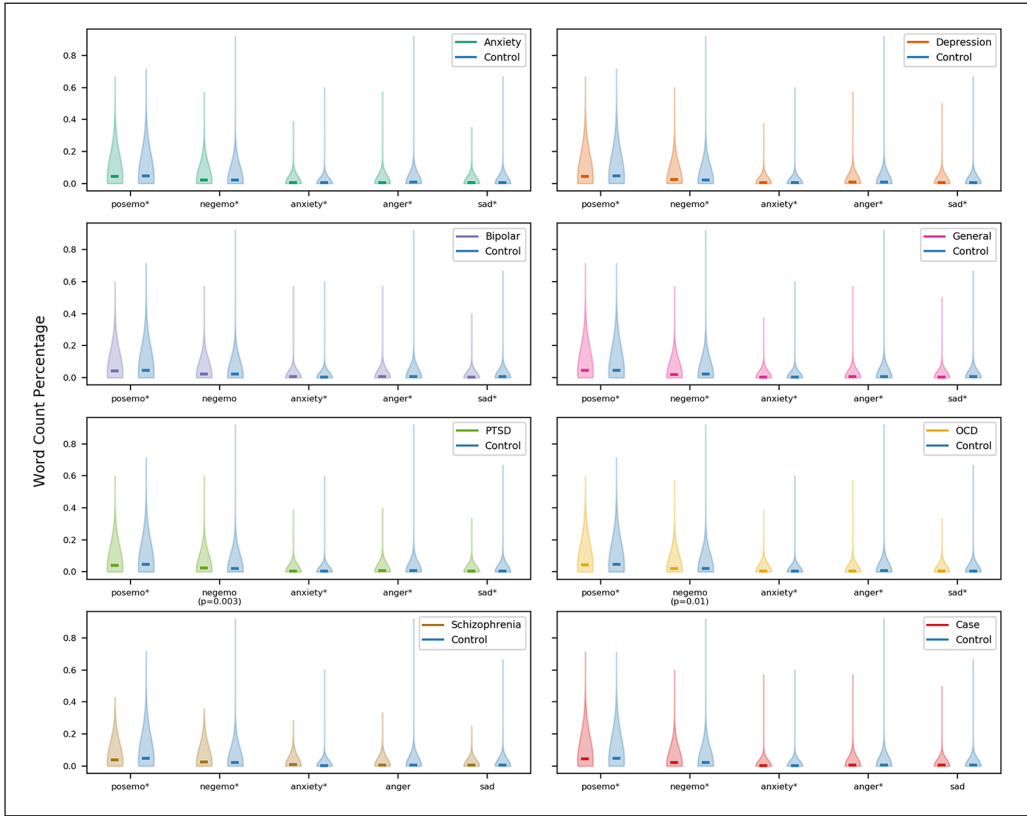


Figure 4. The results of LIWC sentiment analysis of the tweets posted by the case group users of each mental disorder category. The y-axis in each subplot corresponds to the word count percentage of each emotion category “posemo” is shorthand for positive emotions, while “negemo” is for negative emotions. Each subgroup was compared to the control group, respectively. The last subplot shows the sentiment difference between the entire case group and the control group.

OCD, and other mental health illness in general. One of the most important indications of melatonin is sleep disorder.¹⁹ This suggests that most mental disorder patients, no matter what their specific mental disorder is, experienced difficulty in sleeping normally. The second common supplement taken by the case group users was vitamin D. Vitamin D is closely related to depression. Studies have found that the lower level of vitamin D in the body leads to greater risk of depression;²⁰ also, vitamin D receptors were located in the same brain region that is associated with depression.²¹ As shown in Table 4, users who self-diagnosed depression most frequently mentioned about vitamin D. This suggests that our method was able to find signals of associations between supplement intake and mental disorders.

We also compared the sentiment of the tweets between case group and control group regarding the supplement intakes. As shown in Figure 4, for every type of mental disorder patient, the case group expressed fewer positive emotions and more negative emotions than the control group. Especially, the case group expressed more anxiety than the control group. We calculated the average number of tweets posted by the case group and the control group. The case group contains 173

users and 338,788 tweets, which averaged to 1958 tweets per user, while the control group contains 692 users and 629,274 tweets in total, which averaged 909 tweets per user. This suggests that the mental disorder patients who were taking supplements are more active on Twitter than those who were taking none. Therefore, the case group is likely to express negative emotions more often due to their need for solutions to their mental health conditions, and dietary supplement is one solution that may work.

Limitations and future work

The major limitation of this study is the quality of the Twitter data. For one thing, the signal-to-noise ratio of the tweets in terms of supplement intake information is low. During the annotation stage of the study, we found that only about 10 percent of the tweets are descriptions about personal supplement intake of the users. Suggestions to others, advertisement, and health-related blog post links did not contribute to supplement intake information at all, but they are hard to avoid as they all contained the supplement keywords. We also searched for synonyms and aliases of the supplement names before collecting the supplement tweet dataset. Some of the aliases, especially the abbreviations, are ambiguous and thus contribute to the noise. For example, one active ingredient of the fish oil, eicosapentaenoic acid (EPA) can also stand for “Environmental Protection Agency.” It happens that during the time of our data collection, a scandal about the Environmental Protection Agency was one of the hot topics on Twitter, and thus, most of the tweets that matched the keyword “EPA” were irrelevant. For another, it is hard to determine a user’s mental disorder status just by the user’s tweets alone, for Twitter is not a personal health record system. In this study, we adopted a makeshift method where we used mental health issue hashtags and first-person narrative to approximate the self-diagnosis of mental health disorder. While the method did help us shrink the number of users from which we would pick the case group, it cannot guarantee that the self-diagnosis is accurate. Even if the self-diagnosis is accurate, the user can simply choose not to share their supplement intake information on Twitter, which there is nothing we can do about.

Despite the limitations, we were still able to detect some signals of associations between dietary supplement intake and mental health disorders. Since our Tweet2Vec model was trained on character level, it can be readily applied to other types of medical text, for example, clinical notes where the noise ratio is lower, and we can use the same methodology to detect signals between dietary supplement intake and mental health disorders from other text sources in the future.

Conclusion

In this study, we have built an annotated tweet dataset containing 4210 unique tweets and compared several classifiers to identify dietary supplement intake. The best classifier was a CNN classifier that achieves an F1-score of 0.899, a precision of 0.900, and a recall of 0.900 on identifying the language use pattern of “supplement intake” tweets. The classifier was then applied on the Twitter dataset with mental health hashtags. We have found that melatonin was the most commonly used supplement by patients of different type of mental disorders, including anxiety, depression, bipolar disorder, PTSD, OCD, and other types of mental disorders. Vitamin D was also frequently used by anxiety and depression patients. We also analyzed the sentiment of the tweets and found that the users who reported both mental disorders and supplement intake were more active on Twitter and expressed fewer positive emotions and more negative emotions, and especially more anxiety than those who only reported mental disorders but no supplement intake.

Acknowledgements

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Science Foundation.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research reported in this publication was supported by the National Center for Complementary & Integrative Health and the Office of Dietary Supplements (ODS) of the National Institutes of Health under Award Number R01AT009457 (PI: Zhang) and NSF Award #1734134 (PI: Bian).

ORCID iDs

Yefeng Wang  <https://orcid.org/0000-0003-2153-8187>

Yunpeng Zhao  <https://orcid.org/0000-0002-5771-3373>

Jiang Bian  <https://orcid.org/0000-0002-2238-5429>

References

1. De Jonge P, Wardenaar KJ, Hoenders HR, et al. Complementary and alternative medicine contacts by persons with mental disorders in 25 countries: results from the World Mental Health Surveys. *Epidemiol Psychiatr Sci* 2018; 27(6): 552–567.
2. Lakhani SE and Vieira KF. Nutritional therapies for mental disorders. *Nutr J* 2008; 7: 2.
3. De Choudhury M, Gamon M, Counts S, et al. Predicting depression via social media. In: *Proceedings of the seventh international AAAI conference on weblogs and social media (ICWSM13)*, Cambridge, MA, 8–11 July 2013, pp. 1–10. Menlo Park, CA: AAAI.
4. Coppersmith G, Dredze M and Harman C. Quantifying mental health signals in Twitter. In: *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, Baltimore, MD, 23–25 June 2014, pp. 51–60. Stroudsburg, PA: ACL.
5. Cauffield JS and Forbes HJ. Dietary supplements used in the treatment of depression, anxiety, and sleep disorders. *Lippincotts Prim Care Pract* 1999; 3(3): 290–304.
6. Perrin A. Social media fact sheet, <https://www.pewinternet.org/fact-sheet/social-media/> (2018, accessed 16 July 2018).
7. Reece AG, Reagan AJ, Lix KL, et al. Forecasting the onset and course of mental illness with Twitter data. *Sci Rep* 2017; 7(1): 13006.
8. Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. In: *NIPS'13 proceedings of the 26th international conference on neural information processing systems, Lake Tahoe, NV*, 5–10 December 2013, pp. 3111–3119. New York: Curran Associates, Inc.
9. Word2Vec—Google Code Archive, <https://code.google.com/archive/p/word2vec/> (2013, accessed 16 July 2018).
10. English word vectors fastText, <https://fasttext.cc/docs/en/english-vectors.html> (2017, accessed 16 July 2018).
11. Pennington J, Socher R and Manning CD. GloVe: global vectors for word representation, <https://nlp.stanford.edu/projects/glove/> (2014, accessed 16 July 2018).
12. Biomedical natural language processing, <http://bio.nlpplab.org/> (2013, accessed 16 July 2018).
13. Chen Z, He Z and Liu X. An exploration of semantic relations in neural word embeddings using extrinsic knowledge. In: *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*, Kansas City, MO, 13–15 November 2017, pp. 1246–1251. New York: IEEE.

14. Hicks A, Hogan WR, Rutherford M, et al. Mining Twitter as a first step toward assessing the adequacy of gender identification terms on intake forms. *AMIA Annu Symp Proc* 2015; 2015: 611–620.
15. Pennebaker JW, Boyd RL, Jordan K, et al. The development and psychometric properties of LIWC2015, September 2015, <https://repositories.lib.utexas.edu/handle/2152/31333>
16. Rude S, Gortner EM and Pennebaker J. Language use of depressed and depression-vulnerable college students. *Cogn Emot* 2004; 18(8): 1121–1133.
17. Klein A, Sarker A, Rouhizadeh M, et al. Detecting personal medication intake in Twitter: an annotated corpus and baseline classification system. In: *BioNLP 2017*, Vancouver, BC, Canada, 4 August 2017, pp. 136–142. Stroudsburg, PA: Association for Computational Linguistics.
18. Vosoughi S, Vijayaraghavan P and Roy D. Tweet2vec: learning tweet embeddings using character-level CNN-LSTM encoder-decoder. In: *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval*, Pisa, 17–21 July 2016, pp. 1041–1044. New York: ACM.
19. Auld F, Maschauer EL, Morrison I, et al. Evidence for the efficacy of melatonin in the treatment of primary adult sleep disorders. *Sleep Med Rev* 2017; 34: 10–22.
20. Penckofer S, Kouba J, Byrn M, et al. Vitamin D and depression: where is all the sunshine? *Issues Ment Health Nurs* 2010; 31(6): 385–393.
21. Eyles DW, Smith S, Kinobe R, et al. Distribution of the vitamin D receptor and 1 α -hydroxylase in human brain. *J Chem Neuroanat* 2005; 29(1): 21–30.