Detecting Depression and Mental Illness on Social Media: An Integrative Review

Sharath Chandra Guntuku[1], David B. Yaden[1], Margaret L. Kern[2],
Lyle H. Ungar[1], Johannes C. Eichstaedt*[1]

[1]University of Pennsylvania, Philadelphia, PA
[2]The University of Melbourne, Melbourne, Australia

sharathg@sas.upenn.edu, dyaden@sas.upenn.edu, Margaret.Kern@unimelb.edu.au,
ungar@cis.upenn.edu, Johannes.penn@gmail.com

*Corresponding Author

Abstract

Although rates of diagnosing mental illness have improved over the past few decades, many cases remain undetected. Symptoms associated with mental illness are observable on Twitter, Facebook, and web forums, and automated methods are increasingly able to detect depression and other mental illnesses. In this paper, recent studies that aimed to predict mental illness using social media are reviewed. Mentally ill users have been identified using screening surveys, their public sharing of a diagnosis on Twitter, or by their membership in an online forum, and they were distinguishable from control users by patterns in their language and online activity. Automated detection methods may help to identify depressed or otherwise at-risk individuals through the large-scale passive monitoring of social media, and in the future may complement existing screening procedures.

Introduction

The widespread use of social media may provide opportunities to help reduce undiagnosed mental illness. A growing number of studies examine mental health within social media contexts, linking social media use and behavioral patterns with stress, anxiety, depression, suicidality, and other mental illnesses. The greatest number of studies of this kind focus on depression. Depression continues to be under-diagnosed, with roughly half the cases detected by primary care physicians **[1]** and only 13-49% receiving minimally adequate treatment **[2]**.

Automated analysis of social media potentially provides methods for early detection. If an automated process could detect elevated depression scores in a user, that individual could be targeted for a more thorough assessment, and provided with further resources, support, and treatment. Studies to date have either examined how the use of social media sites correlates with mental illness in users **[3]** or attempted to detect mental illness through analysis of the content created by users. This review focuses on the latter: studies aimed at predicting mental illness using social media. We first consider methods used to predict depression, and then consider four approaches that have been used in the literature. We compare the different approaches, provide direction for future studies, and consider ethical issues.

Prediction Methods

Automated analysis of social media is accomplished by building predictive models, which use 'features,' or variables that have been extracted from social media data. For example, commonly used features include users' language encoded as frequencies of each word, time of posts, and other variables (see Fig. 2). Features are then treated as independent variables in an algorithm (e.g., Linear Regression [**4**] with built in variable selection [**5**], or Support Vector Machines (SVM) **[6]** to predict the dependent variable of an outcome of interest (e.g., users' mental health). Predictive models are trained, using an algorithm, on part of the data (the training set) and then are evaluated on the other part (the test set) to avoid overfitting – a process called cross-validation. The prediction performances are then reported as one of several possible metrics (see Table 1).

Assessment Criteria

Several approaches have been studied for collecting social media data with associated information about the users' mental health. Participants are either recruited to take a depression survey and share their Facebook or Twitter data (section A below), or data is collected from existing public online sources (sections B, C, and D below; see Fig. 1). These sources include searching public Tweets for keywords to identify (and obtain all Tweets from) users who have shared their mental health diagnosis (section B), user language on mental illness related forums (section C), or through collecting public Tweets that mention mental illness keywords for annotation (section D). The approaches using public data (sections B, C, D) have the advantage

that much larger samples can, in principle, be collected faster and more cheaply than through the administration of surveys (see Table 1 for sample sizes), though survey-based assessment (section A) generally provides a higher degree of validity **[7]**.

We first compare studies that attempt to distinguish mentally ill users from neurotypical controls (Sections A and B). Table 1 summarizes the methodological details of these studies.

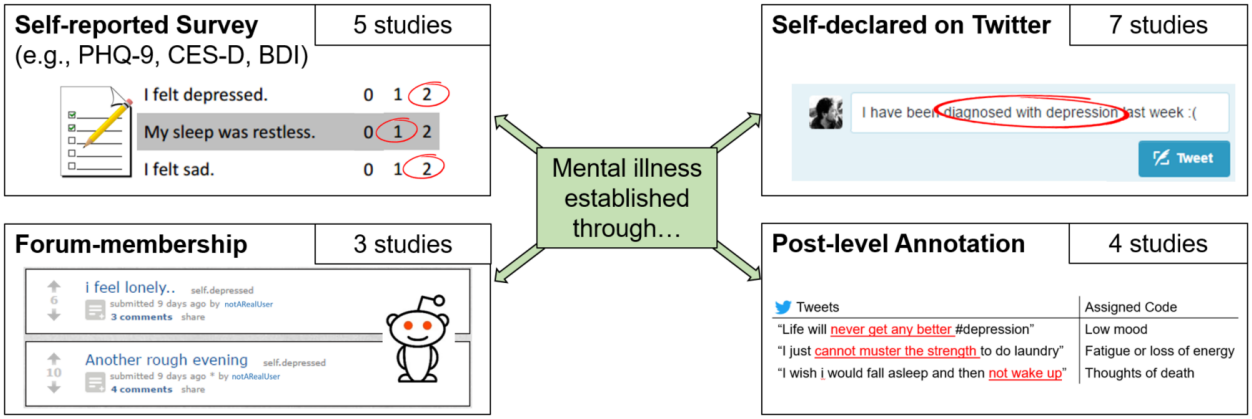

Figure 1. Data sources used in studies as assessment criteria to establish mental illness status. The number of studies selected for review in the present article is provided. The most commonly used self-reported screening surveys for depression include the PHQ-9 = Patient Health Questionnaire **[7]**, CES-D = Centers for Epidemiological Studies Depression Scale Revised [**9**], BDI = Beck Depression Inventory [**10]**

A. Prediction Based on Survey Responses

Psychometric self-report surveys for mental illness have a high degree of validity and reliability (e.g., see, **[7])**. In psychological and epidemiological research, self-report surveys are second only to clinical interviews, which no social media study to date has used as an outcome measure. We discuss five studies that predict survey-assessed depression status by collecting participants' responses to depression surveys in conjunction with their social media data.

The most cited study used Twitter activity to examine network and language data preceding a recent episode of depression **[8]**. The presence of depression was established through participants reporting the occurrence and recent date of a depressive episode, combined with scores on the Center for Epidemiologic Studies Depression Scale Revised (CES-D, [**9**] and Beck's Depression Inventory (BDI, [**10]**). This study revealed several distinctions in posting activity by depressed users, including: diurnal cycles, more negative emotion, less social interaction, more self-focus, and mentioning depression-related terms throughout the year preceding depression onset.

[11] predicted user depression and post-traumatic stress-disorder (PTSD) status from text and Twitter meta-data that preceded a reported first episode (see Fig. 2 for examples of meta-data) with relatively high Areas under the Receiver Operating Characteristic (ROC) curve (AUCs) of .87 (depression) and .89 (PTSD). Data were aggregated to weeks, which somewhat outperformed aggregation to days, and modelled as longitudinal trajectories of activity patterns that differentiated healthy from mentally ill users.

[**12**] predicted depression from Twitter data in a Japanese sample, using the CES-D as their assessment criterion. Using tweets from the most recent 6–16 weeks preceding the administration of the CES-D was sufficient for recognizing depression; predictions derived from data across a longer period were less accurate.

While most studies have used Twitter, **[13]** used Facebook status updates for the prediction. Mothers self-reported a specific postpartum depression (PPD) episode and completed a screening survey. A model using demographics, Facebook activity, and content of posts before childbirth accounted for 35.5% of the (within-sample[1]) variance in PPD status.

[14] and colleagues used questions from a personality survey to determine users' continuous depression scores across a larger sample of Facebook users (N = 28,749) than used in other studies (which typically range in the low hundreds). This study observed seasonal fluctuations of depression, finding that people were more depressed during winter months. This study also provided a shortlist of the words, phrases and *topics* (clusters of semantically coherent words) most associated with depression.

Survey responses provide the most reliable ground-truth data for predictive models in this emerging literature. However, the costs required for this method have motivated the use of more publically accessible assessment criteria, such as those described in the next three sections.

B. Prediction Based on Self-Declared Mental Health Status

A number of studies use publicly accessible data. "Self-declared" mental illness diagnosis on Twitter (identified through statements such as "I was diagnosed with depression today") is one such source of publicly-available data. We review seven studies of this kind. Helping to facilitate studies of this kind, a Computational Linguistics and Clinical Psychology (CLPsych) workshop was started in 2014 to foster cooperation between clinical psychologists and computer scientists. "*Shared tasks*" were designed to explore and compare different solutions to the same prediction

[1] i.e., not using cross-validation

problem on the same data set.

In the 2015 CLPsych workshop, participants were asked to predict if a user had PTSD or depression based on self-declared diagnoses on Twitter (PTSD *n* = 246, depression *n* = 327, with the same number of age- and gender-matched controls) [**15**]. Participating teams built language *topic models* (e.g., an anxiety topic contained the words: *feel, worry, stress, study, time, hard*) **[16]**, sought to identify words most associated with PTSD and depression status **[17],** considered sequences of characters as features [**15**], and applied a rule-based approach to build relative counts of N-grams present in PTSD and depression statuses of all users **[18]**. The latter resulted in the highest prediction performance. All approaches found that it was harder to distinguish between PTSD and depression versus detecting the presence of either condition (compared to controls), suggesting overlap in the language associated with both conditions.

On a shared dataset similar to the 2015 CLPsych workshop, the prediction of anxiety was improved by taking gender into account in addition to 10 comorbid conditions **[19]** . Other studies have used psychological dictionaries (Linguistic Inquiry and Word Count; LIWC) [**20]** to characterize differences between mental illness conditions [**21]**, with some success. On the same dataset, **[17]** observed that estimating the age of users adequately identified users who had self-declared a PTSD diagnosis, and that the language predictive of depression and PTSD had large overlap with the language predictive of personality. This suggests that users with particular personality or demographic profiles chose to share their mental health diagnosis on Twitter, and thus that the results of these studies (mostly, prediction accuracies) may not generalize to other sources of autobiographical text.

C. Prediction Based on Forum Membership

Online forums and discussion websites are a second source of publicly-available text related to mental health. They offer a space in which users can ask for advice, receive and provide emotional support, and generally discuss stigmatized mental health problems openly. We review three such studies here.

In **[22],** forum (reddit) posts were used to study the mental well-being of U.S. university students. A prediction model was trained on data gathered from reddit mental health support communities and applied to the posts collected from 109 university subreddits to estimate the level of distress at the universities. The proportion of mental health posts increased over the course of the academic year, particularly for universities with quarter-based, rather than semester-based, schedules. In **[23],** the language of 16 subreddits covering a range of mental health problems was characterized using LIWC and other markers of sentence complexity.

[24] examined posts of a group of reddit users who posted about mental health concerns and then shifted to discuss suicidal ideation in the future. Several features predicted this shift:

heightened self-focus, poor linguistic style matching with the community, reduced social engagement, and expressions of hopelessness, anxiety, impulsiveness, and loneliness.

Ref.	Year	Dataset			Section	Mental Illness Criteria	Features (predictors)						Outcome Type	Model	Metric	Performance
		Platform	N (users)	Cases (conditions; base rate [BR])			n-grams	LIWC	Sentiment	Topics	Metadata	Others				
[8]	2013	Twitter	476	Depression = 171 (BR = 36%)	A	survey (CESD + BDI)		Y	Y		Y	Social Network	Binary	PCA, SVM w/ RBF Kernel	Accuracy	.72
[13]	2014	Facebook	165	Post-partum Depression = 28 (BR = 17%)	A	survey (PHQ-9)		Y	Y		Y	User Activity, Social Capital	Binary	Logistic Regression	pseudo-R2**	.36
[14]	2014	Facebook	28,749	(continous Depression score)	A	survey (Personality)	Y	Y		Y			Continuous	Ridge Regression	Correlation	.38
[12]	2015	Twitter	209	Depression = 81 (BR = 39%)	A	survey (CESD)	Y	Y	Y	Y	Y	User Activity	Binary	SVM	Accuracy	.69
[11]	2016	Twitter	378	Depression = 105 (BR = 28%) PTSD = 63 (BR = 17%)	A	survey (CESD)		Y	Y		Y	Time-Series, LabMT	Binary	Random Forests	AUC	Depression = .87 PTSD = .89
[41]	2014	Twitter	5,972	PTSD = 244 (BR = 4%)	B	self-declared	Y	Y					Binary	(not reported)	ROC	(AUC not reported)
[43]	2014	Twitter	21,866	11,866 (across 4 Conditions, BR = 54%)	B	self-declared	Y	Y	Y		Y	User Activity	Binary	Log linear classifier	Precision*	Depression = .48 Bipolar = .64 PTSD = .67 SAD = .42
[17]	2015	Twitter	1,957	Depression = 483 (BR = 25%) PTSD = 370 (BR = 19%)	B	self-declared	Y	Y	Y	Y		Age, Gender, Personality	Binary	Logistic Regression	AUC	Depression = .85 PTSD = .91
[21]	2015	Twitter	4,026	2,013 (across 10 Conditions, BR = 50%)	B	self-declared	Y	Y					Binary	(not reported)	Precision*	Depression = .48 Bipolar = .63 Anxiety = .85 Eating Dis. = .76
[42]	2016	Twitter	250	Suicide Attempt = 125 (BR = 50%)	B	self-declared	Y		Y		Y	User Activity	Binary	(not reported)	Precision*	.70
[44]	2016	Twitter	900	Depression = 326 (BR = 36%)	B	self-declared	Y						Binary	Naive Bayes	AUC	.70
[19]	2017	Twitter	9,611	4820 (across 8 Conditions, BR = 50%)	B	self-declared	Y					Gender	Multi-Task	Neural Network	AUC	Depression = .76 Bipolar = .75 Depression = .76 Suicide Attempt = .83

Table 1: Prediction performances achieved by different mental illness studies reviewed in this paper. The relevant dataset, features, and prediction settings are provided. AUC: Area Under the Receiver Operating Characteristic (ROC) Curve; Precision: fraction of cases ruled positive that are truly positive; Accuracy: fraction of cases that are correctly labeled by the model; SVM: Support Vector Machines; PCA: Principal Component Analysis; RBF - Radial Basis Function *Precision with 10% False Alarms; **within-sample (not cross-validated); ***using the Depression facet of the Neuroticism factor measured by the International Personality Item Pool (IPIP) proxy to the NEO-PI-R Personality Inventory **[39]**. Studies highlighted in green report AUCs; AUCs are not base rate dependent and can be compared across studies.

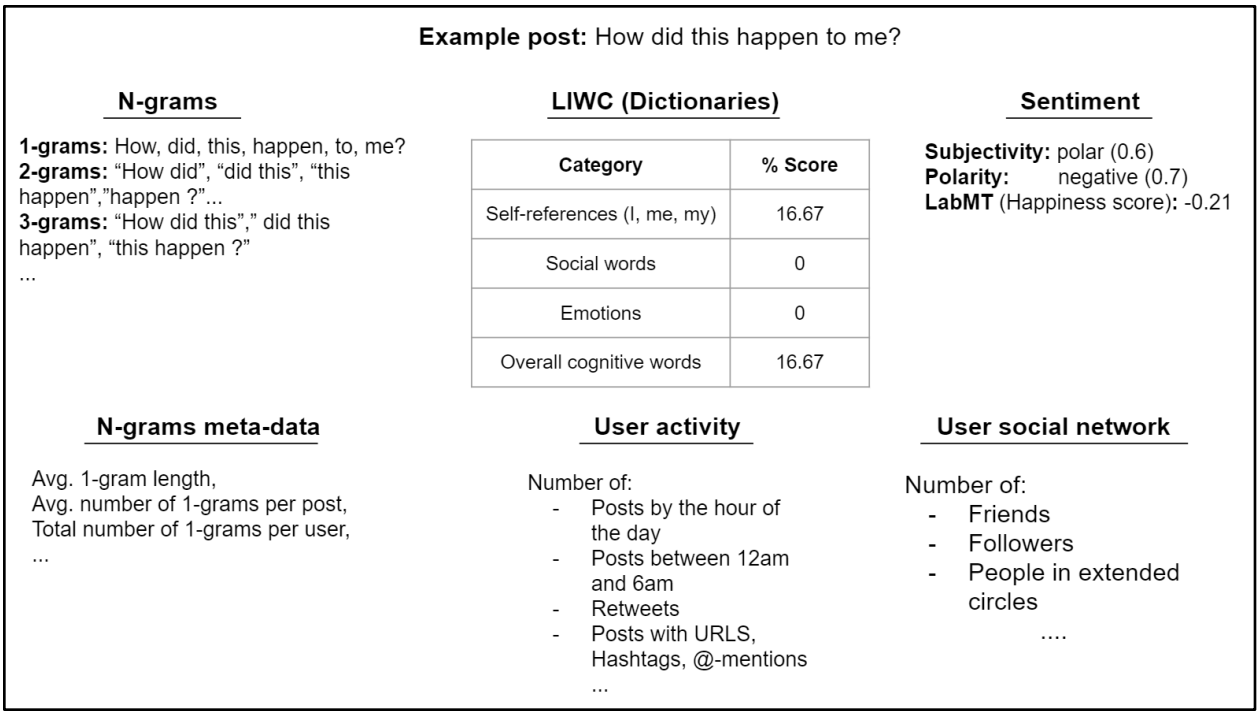

Figure 2. Examples of features included in the different feature sets referenced in Table 1. LIWC: Linguistic Inquiry and Word Count **[20]**, LabMT: Language Assessment by Mechanical Turk **[40]**

D. Prediction based on Annotated Posts

A third source of publicly-available text involves manually examining and annotating Tweets that contain mental health keywords. Annotators code social media posts according to pre-established (a priori, theory-driven) or bottom-up (determined from the data) classifications **[25, 26]**; annotations can be predicted from the language of posts.

Most annotation studies on depression focus on identifying posts in which users are discussing their own experience with depression **[27]**. Annotators are provided with guidelines on how to recognize a broad range of symptoms of depression **[28]** that are derived from clinical assessment manuals such as the DSM-5 **[29],** or a reduced set of symptoms, such as *depressed mood*, *disturbed sleep* and *fatigue* **[30]**. Annotation has also been used to differentiate between mentions of mental illness for the purpose of stigmatization or insult as opposed to voicing support or sharing useful information with those suffering from a mental illness **[25].** In general, annotations of posts are a complementary (but labor-intensive) method that can reveal life circumstances associated with mental illness (e.g., occupational and educational problems, or the weather **[28]**) not captured by traditional depression diagnostic criteria **[29]**.

Comparison of Studies across Data Sources

Our review has described four sources of data used to study and detect depression through social media. Here we compare these sources.

Ease of collection & sample biases. While validated and reliable screening surveys (section A) are the closest to clinical practice, they are costly to administer at large scale and are often completed by self-selecting crowdworkers (e.g., on MTurk) which introduces a variety of sampling biases **[31]**. The approaches using publicly accessible data (section B, C and D) have larger samples, but incur additional sample biases by relying on users to share their diagnosis publicly (e.g., see **[17]**) or join a forum, and it is unlikely that users unaware of their diagnosis would be captured.

Prediction performances. The lower the base rate of mentally ill users in a study sample, the harder the prediction task. While U.S. prevalence rates are below 10% [8], many studies opt for a more equal class balance (closer to base rates of 50%). Performance metrics like precision and accuracy depend on base rates; AUCs do not and are thus more comparable across studies. AUCs reported in the studies reviewed above (sections A and B) range from moderate (.70) to high (.91; see Table 1).

How do these AUCs compare with clinical baselines? Using clinical (as opposed to self-selected online) samples and gold-standard structured clinical interviews as the criterion, Mitchell et al. **[32]** estimated the ability of primary care physicians to detect depression as meta-analytic Bayesian case-finding AUCs for different countries, which range from AUC = .62 in Australia and .65 in the U.S. to AUC = .74 in the Netherlands. These AUCs are matched or exceeded by the AUCs reported in the studies reviewed above (see Table 1). On the other hand, screening inventories (such as the Patient Health Questionnaire (PHQ) **[7]** and Hospital Anxiety and Depression Scale (HADS **[33]**) obtain high AUCs of around .90 against structured clinical interviews (e.g., **[34]**). This suggests that social media-based screening may reach prediction performance somewhere between unaided clinician assessment and screening surveys; however, no study to date has assessed social-media-based prediction against structured clinical interviews.

Recommendations for Future Studies

The greatest potential value of social media analysis may be the detection of otherwise undiagnosed cases. However, studies to date have not explicitly focused on successfully identifying people unaware of their mental health status.

In screening for depression, multi-stage screening strategies have been recommended **[35, 32]** as a means to alleviate the relatively low sensitivity (around 50%) and high false positive rate associated with assessments by non-psychiatric physicians **[1, 32]** or short screening inventories **[35]**. Social-media based screening may eventually provide an additional step in a mental health screening strategy. Studies are needed that integrate social media data collection with gold-standard structured clinical interviews and other screening strategies in ecologically valid samples to test the incremental benefit of social media based screening and distinguishing between mental health conditions **[15, 21].**

Self-reported surveys and clinical diagnoses provide snapshots in time. Online social media data may "fill in the gaps" with ongoing in-the-moment measures of a broad range of people's thoughts and feelings. However, as depressed users may cease generating social media content [**36**], alternative uninterrupted data streams such as text messages and sensor data should also be tested for ongoing monitoring applications **[37]**.

Ethical Questions

The feasibility of social-media-based assessment of mental illness raises numerous ethical questions. Privacy is an ongoing concern. Employers and insurance companies, for example, may use these against the interests of those suffering from mental illness. As mental illnesses carry social stigma and may engender discrimination, data protection and ownership frameworks are needed to ensure users are not harmed **[38]**. Few users realize the amount of mental-health-related information that can be gleaned from their digital traces. Transparency about which health indicators are derived by whom and why is critical.

From a mental health perspective, clear guidelines on mandated reporting are needed. There are open questions around the impact of misclassifications, and how derived mental health indicators can be responsibly integrated into systems of care [**36**]. Discussions around these issues should include clinicians, computer scientists, lawyers, ethicists, policy makers, and individuals from different socioeconomic and cultural backgrounds who suffer from mental illness.

Conclusion

The studies reviewed here suggest that depression and other mental illnesses are detectable on several online environments, but the generalizability of these studies to broader samples and gold standard clinical criteria has not been established. Advances in natural language processing and machine learning are making the prospect of large-scale screening of social media for at-risk individuals a near-future possibility. Ethical and legal questions about data ownership and protection, as well as clinical and operational questions about integration into

systems of care should be addressed with urgency.

Acknowledgements

The authors thank Courtney Hagan for her help with editing the manuscript. This work was supported by a grant from the Templeton Religion Trust (ID #TRT0048)

References

1 Cepoiu M, McCusker J, Cole MG, Sewitch M, Belzile E, Ciampi A: Recognition of depression by non-psychiatric physicians-a systematic literature review and meta-analysis. Journal of General Internal Medicine, 23(1), 25-36.

2 Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler R: Twelve-month use of mental health services in the United States: Results from the National Comorbidity Survey Replication. Archives of General Psychiatry, 62(6), 629-640.

3 Seabrook EM, Kern ML, Rickard NS: Social networking sites, depression, and anxiety: a systematic review. *JMIR Mental Health*, *3*(4), e50.

4 Neter J, Kutner MH, Nachtsheim CJ, Wasserman W: Applied linear statistical models. 1996. Chicago: Irwin.

5 Tibshirani R: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, 267-288.

6 Cortes C, Vapnik V: Support-vector networks. Machine Learning, 20(3), 273-297.

7 Löwe, B., Kroenke, K., Herzog, W., & Gräfe, K. (2004). Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). Journal of affective disorders, 81(1), 61-66.

8 De Choudhury M, Gamon M, Counts S, Horvitz E: Predicting Depression via Social Media. In Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, 2013.
** This classic study demonstrates different diurnal cycles in posting activity between depressed and non-depressed users, and that use of depression-related language (e.g., negative emotion, depression terms, focus on the self) increased across the year preceding depression onset.

9 Radloff LS: The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1(3), 385-401.

10 Beck AT, Steer RA, Brown GK: Beck depression inventory II. San Antonio, 78(2), 490-8.

11 Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ: Forecasting the onset and course of mental illness with Twitter data. *arXiv preprint arXiv:1608.07740* (2016).
12 Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H: Recognizing Depression from Twitter Activity. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM 2015, 3187-3196.
13 De Choudhury M, Counts S, Horvitz EJ, Hoff A: Characterizing and predicting postpartum depression from shared facebook data. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, ACM 2014, 626-638.
14 Schwartz HA, Eichstaedt J, Kern M, Park G, Sap M, Stillwell D, Kosinski M, Ungar L: Towards assessing changes in degree of depression through facebook. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, ACL 2014, 118-125.
15 Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M. CLPsych 2015 shared task: Depression and PTSD on Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality 2015 Jun 5 (pp. 31-39).
** This paper summarizes the rationale behind Shared Tasks in which multiple teams of researchers benchmark different prediction methods on the same data set, and reports prediction performances obtained in the 2015 shared task in the identification of Twitter users self-declared to suffer from depression or PTSD vs. controls.

16 Resnik P, Armstrong W, Claudino L, Nguyen T, Nguyen V, Boyd-Graber, J: Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, ACL 2015, 99-107.
17 Preotiuc-Pietro D, Eichstaedt J, Park G, Sap M, Smith L, Tobolsky V, Schwartz HA, Ungar L: The role of personality, age and gender in tweeting about mental illnesses. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology, ACL 2015, 21-31.
* This study showed that language predictive of a self-declared diagnosis of depression and PTSD had a large overlap with the language predictive of demographics and personality, suggesting that it may be users with a particular personality or demographic profile who choose to share their mental health diagnosis on Twitter.

18 Pedersen T: Screening Twitter users for depression and PTSD with lexical decision lists. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality 2015.
19 Benton A, Mitchell M, Hovy D: Multi-Task Learning for Mental Health using Social Media Text, Proceedings of European Chapter of the Association for Computational Linguistics 2017.
20 Pennebaker JW, Booth RJ, Francis ME: Linguistic inquiry and word count: LIWC [Computer software]. *Austin, TX: liwc. net* (2007).
21 Coppersmith G, Dredze M, Harman C, Hollingshead K: From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology, NAACL 2015, 1-11.
22 Bagroy S, Kumaraguru P, De Choudhury M: A Social Media Based Index of Mental Well-Being in College Campuses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,* 2017.
* This study demonstrates the potential use of social media for monitoring the mental health of university communities. Mental well-being seems to be lower in universities with more females, lower tuition, and those that are public and located in rural or suburban areas**.**

23 Gkotsis, G., Oellrich, A., Hubbard, T. J., Dobson, R. J., Liakata, M., Velupillai, S., & Dutta, R. (2016). The language of mental health problems in social media. In *Third Computational Linguistics and Clinical Psychology Workshop (NAACL)* (pp. 63-73).
24 De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M: Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2098-2110.
25 Hwang JD, Hollingshead K: Crazy Mad Nutters: The Language of Mental Health, In Proceedings of the 3rd Workshop on Computational Linguistics and Clinical Psychology, ACL 2016, 52-62.
26 Kern ML, Park G, Eichstaedt JC, Schwartz HA, Sap M, Smith LK, Ungar LH: Gaining insights from social media language: Methodologies and challenges. Psychological Methods 2016, 21:4, 507.
27 Cavazos-Rehg PA, Krauss M, Sowles S, Connolly S, Rosas C, Bharadwaj M, Bierut L: A Content Analysis of Depression-Related Tweets. Comput Human Behav 2016, 54:351-357.
28 Mowery DL, Bryan C, Conway M: Towards Developing an Annotation Scheme for Depressive Disorder Symptoms: A Preliminary Study using Twitter Data. In Proceedings of 2nd

Workshop on Computational Linguistics and Clinical Psychology-From Linguistic Signal to Clinical Reality 2015, 89-99.
29 American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). American Psychiatric Publishing.
30 Mowery D, Bryan C, Conway M: Feature Studies to Inform the Classification of Depressive Symptoms from Twitter Data for Population Health. 2017 arXiv preprint arXiv:1701.08229.
31 Arditte, K. A., Çek, D., Shaw, A. M., & Timpano, K. R. (2016). The importance of assessing clinical phenomena in Mechanical Turk research. *Psychological assessment*, *28*(6), 684.
32 Mitchell AJ, Rao S, Vaze A: International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. British Journal of General Practice, 61(583).
33 Zigmond AS, Snaith RP: The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica, 67(6), 361-370.
34 Löwe, B., Spitzer, R. L., Gräfe, K., Kroenke, K., Quenter, A., Zipfel, S., ... & Herzog, W. (2004). Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. Journal of affective disorders, 78(2), 131-140.
35 Nease DE, Malouin JM: Depression screening: A practical strategy. Journal of Family Practice, 52(2), 118-126.
36 Inkster B, Stillwell D, Kosinski M, Jones P: A decade into Facebook: where is psychiatry in the digital age? The Lancet Psychiatry 2016, 3:1087-1090.
37 Mohr DC, Zhang M, Schueller S: Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annual Review of Clinical Psychology. 2017 Apr;13(1).
38 Inkster B, Stillwell D, Kosinski M, Jones P: A decade into Facebook: where is psychiatry in the digital age? The Lancet Psychiatry 2016, 3:1087-1090.
39 Goldberg LR: A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. Personality psychology in Europe 1999, 7:7-28.
40 Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one*, *8*(5), e64417.
41 Coppersmith G, Harman C, Dredze M: Measuring Post Traumatic Stress Disorder in Twitter. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, 2014, 579-582.

42 Coppersmith G, Ngo K, Leary R, Wood A: Exploratory analysis of social media prior to a suicide attempt. In Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology, 2016, 106-117.
43 Coopersmith G, Dredze M, Harman C: Quantifying mental health signals in Twitter. Workshop on Computational Linguistics and Clinical Psychology, ACL 2014, 51-60.
44 Nadeem M: Identifying depression on Twitter, arXiv preprint arXiv:1607.07384 (2016).