Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Review article

A Systematic review of the validity of screening depression through Facebook, Twitter, Instagram, and Snapchat

Jiin Kim [a], Zara A. Uddin [a], Yena Lee [a], Flora Nasri [a], Hartej Gill [a], Mehala Subramanieapillai [a], Renna Lee [a], Aleksandra Udovica [a], Lee Phan [a], Leanna Lui [a], Michelle Iacobucci [a], Rodrigo B. Mansur [a,d], Joshua D. Rosenblat [a,c], Roger S. McIntyre [b,c,d,e,f,g,*]

[a] *Mood Disorders Psychopharmacology Unit, University Health Network, 399 Bathurst Street, MP 9-325, Toronto, ON M5T 2S8, Canada*
[b] *Institute of Medical Science, University of Toronto, Toronto, ON, Canada*
[c] *Department of Pharmacology, University of Toronto, Toronto, ON, Canada*
[d] *Department of Psychiatry, University of Toronto, Toronto, ON, Canada*
[e] *Brain and Cognition Discovery Foundation, Toronto, ON, Canada*
[f] *Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
[g] *Institute for Health Innovation and Technology (iHealthtech), National University of Singapore, Singapore*

ARTICLE INFO

Keywords:
Social media
Facebook
Twitter
Instagram
Depression
Social media markers
Screening
Snapchat
Virtual
Ecological Momentary Assessment (EMA)

ABSTRACT

Background: The aim of this study was to determine the validity of using social media for depression screening.
Method: Article searches on PubMed and PsycINFO from database inception to August 20, 2019 were completed with a search string and filters.
Results: 15 articles made the inclusion criteria. Facebook, Twitter, and Instagram profiles of depressed people were distinguishable from nondepressed people shown by social media markers. Facebook studies showed that having fewer Facebook friends and mutual friends, posting frequently, and using fewer location tags positively correlated with depressive symptoms. Also, Facebook posts with explicit expression of depressive symptoms, use of personal pronouns, and words related to pain, depressive symptoms, aggressive emotions, and rumination predicted depression. Twitter studies showed that the use of "past focus" words, negative emotions and anger words, and fewer words per Tweet positively correlated with depression. Finally, Instagram studies showed that differences in follower patterns, photo posting and editing, and linguistic features between depressed people and nondepressed people could serve as a marker.
Limitations: The primary articles analyzed had different methods, which constricts the amount of comparisons that can be made. Further, only four social media platforms were explored.
Conclusion: Social media markers like number and content of Facebook messages, linguistic variability in tweets and tweet word count on Twitter, and number of followers, frequency of Instagram use and the content of messages on Instagram differed between depressed people and nondepressed people. Therefore, screening social media profiles on these platforms could be a valid way to detect depression.

1. Introduction

Social media outlets such as Facebook, Twitter, Instagram, and Snapchat are popular social media networking sites that people use to share their personal interests, thoughts, and moments. Facebook (https://facebook.com) is a website with over 1 billion registered accounts, and allows users to interact with pages, groups, and other users, by sharing, commenting, and clicking 'like' on posts and comments. Twitter (https://twitter.com) is a microblogging site that allows account holders to 'tweet' statements up to 280 characters, and 'retweet', (i.e., share) other tweets. On Instagram (https://instagram.com/), users can create, share, like, and comment on photos that may have captions attached. Finally, Snapchat (https://snapchat.com/) allows account holders to take pictures to send to others, and once the receiver views the photo, the photo disappears. By virtue of the nature of these platforms, an observer can create a schema about the user by looking at their

[*] Corresponding author at: University Health Network, 399 Bathurst Street, MP 9-325, Toronto, ON M5T 2S8, Canada.
E-mail address: roger.mcintyre@uhn.ca (R.S. McIntyre).

https://doi.org/10.1016/j.jad.2020.08.091
Received 28 May 2020; Received in revised form 1 August 2020; Accepted 21 August 2020
Available online 8 February 2021

profile. As such, examining the social media profiles of people may provide insight into their mental health status, including if they have depression or not.

Major Depressive Disorder (MDD) is a debilitating condition that affects over 300 million people worldwide (WHO, 2018). People with MDD may experience anhedonia, persistent sadness, concentration difficulties, disturbed eating and sleeping patterns, suicidal ideation, fatigue, and feelings of worthlessness (American Psychiatric Association, 2013). This disorder is complex and is influenced by biological, psychological, and sociological factors, including genetics, personality, gender, socioeconomic status, and stress levels (Depression 2019). Given the profound negative impact it has on people's lives, it is important to diagnose and treat depression early. One potential mechanism to more rapidly recognize depression is via social media.

Using social media to detect depression is valuable because it provides information about the mental health status of users that may otherwise be unavailable. For example, due to the stigma around depression and lack of accessibility of mental health services, users may be more comfortable and able to disclose personal mental health information - consciously or unconsciously - online rather than in-person (Ahmedani, 2011). Given the millions of social media users, depressed people may connect with other depressed people, and have online conversations that reveal their depressive nature. Furthermore, there is a benefit of passively detecting depressive symptoms using social media which could capture the current status of the person rather than having to rely on retrospective recall of the patient during a clinical assessment because depressed people exhibit more inaccuracy when recalling their negative affect (Ben-Zeev et al., 2009). Also, using social media can also be better than relying on self-reports, because bias can cause patients to underreport their symptoms, and this threatens the credibility of the data (Eaton et al., 2000).

Past studies have provided insight into how social media could be used for detecting depression by analyzing the pattern of social media use. For example, Radovic et al. (2017)Radovic et al. (2017) qualitatively examined the social media use of 23 adolescents who were diagnosed with depression and it was found that positive and negative social media use varies with mood. Positive use included searching for positive content (i.e. for entertainment, humor, content creation) or for social connection. Negative use included sharing risky behaviors, cyberbullying, and for making self-denigrating comparisons with others. In the context of treatment, these adolescents shifted their social media use patterns from what they perceived as negative to more positive use. As such, detecting a shift such as this on social media may be helpful for mental health workers to understand the mental state of the patient.

Although Radovic et al. (2017) assessed the mental health status of their participants instead of relying on self-reports, they interviewed the subjects instead of directly analyzing their social media accounts (eg., by extracting the number of 'likes' and comments on their posts). Thus, their results are susceptible to errors from participants yielding to demand characteristics and impression management. Also, many studies have looked for potential markers of depression on social media, but few have cross-referenced these features with a validated depression screening tool (Moreno et al., 2012; Park et al., 2013). Although some studies did assess the depression of subjects using validated measures and directly extracted information from social media profiles, there has not been a systematic review done on all of the studies that abide by these conditions.

Current systematic review analyzed the studies done on this topic on Facebook, Twitter, Instagram, and Snapchat. Certain social media platforms were excluded, for example LinkedIn (https://linkedin.com), because it is a platform to build a professional portfolio rather than sharing users' sentiments. Herein, this paper will systematically review the literature for the following outcomes: whether social media markers can validly be used to screen for depression by cross-referencing with validated depression screening tools, and if so, what these markers are. For the purpose of this study, the operational definition of "depressive social media marker" is a marker that is deemed to be indicative of depression because it positively correlates with depressive symptoms of the social media users who produced the marker. These users were analyzed with a standardized scale of depression to verify the potential of social media being a screening tool for detecting depression of the users. Examples of markers are number and the content of messages on Facebook, linguistic variability in tweets and tweet word count on Twitter, and number of followers, frequency of Instagram use and the content of messages on Instagram.

2. Methods

Authors JK, ZU, RL, and AU systematically searched the PubMed and PsycINFO database for papers published from database inception to August 20, 2019, using a methodological filter for primary articles. The search string used was ("social media" OR Instagram OR Facebook OR Twitter OR Snapchat OR "social network") AND (depress*) AND (diagnose* OR screen* OR detect* OR predict*).

2.1. Inclusion criteria

The inclusion criteria for this review were as follows: 1) english language articles; 2) primary articles; 3) human studies; 4) the social media platforms that were analyzed in the articles had to be one of the following: Facebook, Twitter, Instagram or Snapchat; 5) the aim of the article was to analyze the Facebook, Twitter, Instagram, or Snapchat content of participants; 6) depression assessed using a validated and standardized scale; 7) the study design had to be observational and not experimental because the latter would be unethical; 8) the outcome had to be the detection of depression through social media markers.

2.2. Exclusion criteria

The exclusion criteria for this review were as follows: (1) the aim of the article was (a) to identify markers in mental illness other than depression through social media, (b) to identify markers for depression from a social media platform other than Facebook, Twitter, Instagram and Snapchat; (2) the article did not analyze the actual content of the social media platforms (e.g., used exclusively surveys or interviews); (3) the article was not primary and was without original observations (e.g. literature reviews); (3) the article did not use a validated and standardized scale when classifying participants as depressed.

2.3. Data extraction

The search returned 1,333 articles according to PRISMA guidelines (Figure 1). The filters of English, Journal Article, Date range of until August 20, 2019, and Human were applied in PubMed, and 486 articles remained. These same filters yielded 327 articles in PsycINFO. Two authors (JK and ZU) reviewed article titles/abstracts from PubMed identified by our search strategy, and two authors (RL and AU) reviewed article titles/abstracts from PsychINFO identified by the search strategy. After screening titles and abstracts, the authors and volunteers selected 25 articles for full-text review, of which 13 met the inclusion criteria and 12 were removed as they met the exclusion criteria. 14 articles met the inclusion criteria following a search of the references of included studies and one article was manually added from the University of Toronto Onesearch database. Therefore, 15 articles were selected.

3. Results

3.1. Descriptive data

The systematic search resulted in 15 papers that met the inclusion criteria. Many of the studies examined Facebook, Twitter and Instagram to detect depression markers. No Snapchat studies were found. The most

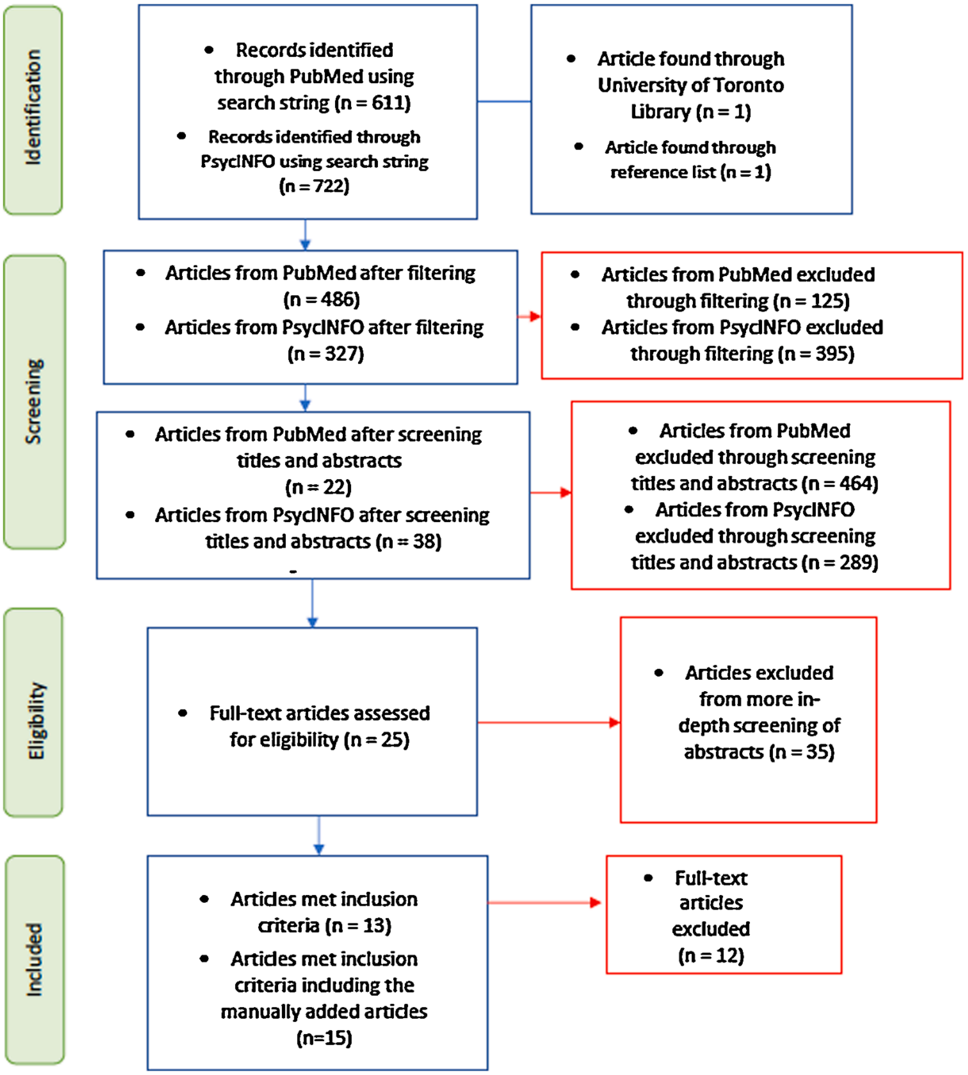
Identification
Records identified through PubMed using search string (n = 611)
Records identified through PsycINFO using search string (n = 722)
Article found through University of Toronto Library (n = 1)
Article found through reference list (n = 1)
Screening
Articles from PubMed after filtering (n = 486)
Articles from PsycINFO after filtering (n = 327)
Articles from PubMed excluded through filtering (n = 125)
Articles from PsycINFO excluded through filtering (n = 395)
Articles from PubMed after screening titles and abstracts (n = 22)
Articles from PsycINFO after screening titles and abstracts (n = 38)
Articles from PubMed excluded through screening titles and abstracts (n = 464)
Articles from PsycINFO excluded through screening titles and abstracts (n = 289)
Eligibility
Full-text articles assessed for eligibility (n = 25)
Articles excluded from more in-depth screening of abstracts (n = 35)
Included
Articles met inclusion criteria (n = 13)
Articles met inclusion criteria including the manually added articles (n=15)
Full-text articles excluded (n = 12)

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) Study Selection Flow Diagram.

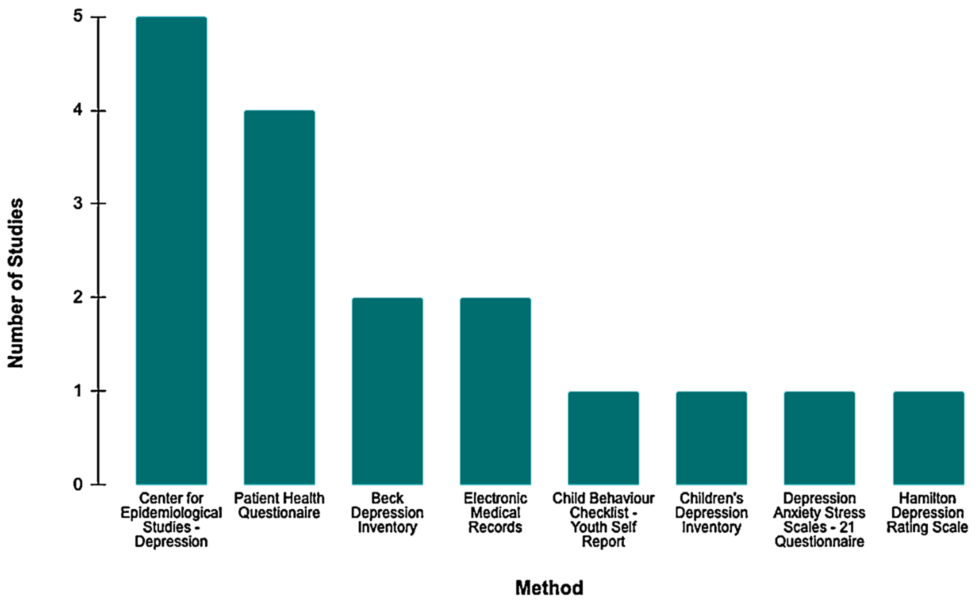

Fig. 2. Method of Depression Assessment.

commonly used method of depression assessment was Center for Epidemiological Studies-Depression (CES-D) (Figure 2).

3.2. Quality assessment

The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies from the National Heart, Lung and Blood Institute (NHLBl) was used to evaluate the quality of the articles (National Heart, Lung and Blood Institute 2020). The assessments are available in Table 1. This tool contains 14 closed-ended questions that can be answered with either a "yes", "no", or "other". The "other" option means that the answer to the question cannot be determined, is not applicable, or is not reported. A greater number of "yes" responses for a study corresponds to a higher quality.

3.3. Facebook

We included 9 papers that looked at the validity of using social media markers on Facebook to predict depression (Ehrenreich et al., 2016; Eichstaedt et al., 2018; Monreno et al., 2012; Negriff, 2019; Ophir et al., 2019; Park et al., 2013; Seabrook et al., 2018; Settanni & Marengo, 2015; Smith et al., 2017). These are summarized in Table 2. The common markers that were found in these papers were about numberand the content of messages.

3.3.1. Numbers

Smith et al. (2017) found that a diagnosis of depression was positively correlated with increased Facebook postings. However, another study found that this relationship only held for adolescent girls and not boys (Ehrenreich et al., 2016). Additionally, two studies (n=2) found that depressive symptoms negatively correlated with the number of Facebook friends (Negriff, 2019; Park et al., 2013). Negriff (2019) also found that the strength of network connections (i.e., number of mutual friends) was correlated with fewer depressive symptoms. Furthermore, it has been found that people with depression use the location tag feature less often, and to an insignificant degree, the 'like' button (Park et al., 2013).

3.3.2. Content of posts

The content of Facebook posts may also help reveal the presence of depression. Monreno et al. (2012) studied the status updates of random students from a university for depressive symptoms. For example, a status that read, "I am feeling down," would be considered as expressing a depressive symptom. They found that there was a significant positive correlation between depressive symptoms expressed in the status updates and the student's score on the Patient Health Questionnaire 9 (PHQ-9). Similarly, another study found that adolescents who explicitly expressed distress in their posts had a higher Beck Depression Inventory II (BDI-II) score than those who did not (Ophir et al., 2019). Finally, Ehrenreich et al. (2016) found that the posts of adolescent girls expressing depressive symptoms were more likely to pertain to negative affect, somatic complaints, and cries for support. The responses of their peers were also more likely to contain negative affect, and offers of support.

3.3.3. Language used in posts

On a deeper level, choice of words can predict depression. All of the following studies used the Linguistic Inquiry and Word Count (LIWC) software to conduct their analyses. One study found that the presence of depression was positively correlated with the expression of negative emotions on Facebook (Settanni & Marengo, 2015). Additionally, Eichstaedt et al. (2018) found that depressed people tend to use more first-person singular pronouns, namely *I, my* and *me*. The other language markers observed by this study includes words related to pain (*hurt, bad, head, surgery, pain, hospital*), depressive symptoms (*tears, cry, pain, miss, much, baby, lost, alone),* aggressive emotions (*smh, fuck, fuckin, hate, ugh),* and rumination (*mind, alot, lot, scared, worry, upset).* Finally, it has been found that there is a significant inconsistency in the proportion of negative emotions in the status updates of participants expressing depressive symptoms (Seabrook et al., 2018).

3.4. Twitter

We examined 4 papers that tested whether Twitter contents could be used for detecting depression (Park et al., 2012; Reece et al., 2017; Sasso et al., 2019; Seabrook et al., 2018). These are summarized in Table 2. The common markers that were found in these papers were about linguistic variability in tweets and tweet word count.

3.4.1. Past focus

Sasso et al. (2019) conducted a longitudinal study on undergraduates and used LIWC software to analyze Twitter feeds. They specifically looked for "past focus" (e.g., "learned," "remember") words because these "past focus" words are associated with ruminative brooding, which is a symptom of depression (American Psychiatric Association, 2013). The Beck Depression Inventory-I (BDI-l) was used to further assess whether the participants had depressive symptoms or not. The study found that participants who were tweeting with a past focus were more likely to show an increase in cognitive vulnerability and depressive symptoms than participants who did not use a past focus for their tweets.

3.4.2. Sentiments

Another study by Park et al., (2012) examined whether the use of sentiment words were different between depressed users identified by scoring higher than 25 on Center for Epidemiological Studies-Depression (CES-D) scale and a typical user. The study found that there was no notable difference in positive emotion, positive feeling, and optimism across the two groups. Users expressed a similar level of these sentiments, irrespective of their depression level. Negative emotions and anger however, had a different pattern, the usage of words related to anger was significantly higher in the depressed group in comparison to the control group. In addition, Reece et al. (2017) found that labMT happiness scores were strongly negatively correlated with depression. This study also found that Affective Norms for English Words (ANEW) and LIWC sentiment-related variables were also predictors. For example, arousing words as evaluated by ANEW predicted depression. Furthermore, another study found a unique pattern such that greater negative emotion word variability was significantly associated with lower depression severity (Seabrook et al., 2018).

3.4.3. Number of Tweets

Reece et al. (2017) found that the average number of words per Tweet (the word count) was negatively correlated with depression. However, they did not find a significant correlation between the frequency of tweets and depression.

3.5. Instagram

This study examined 3 papers that explored whether Instagram contents could be used to detect depression (Lup et al., 2015; Reece & Danforth, 2017; Ricard et al., 2018). These are summarized in Table 2. The common markers that were found in these papers were about the number of followers, frequency of Instagram use and the content of messages.

3.5.1. Followers and Instagram use

Lup et al. (2015) examined the association between depressive symptoms, Instagram use, and amount of strangers one follows. Results showed that there was a marginal positive correlation between number of hours using Instagram and depressive symptoms, and the number of strangers followed slightly moderated this relationship. Specifically, depressive symptoms increased with Instagram use if the user followed a

Table 1
Quality Evaluation of Studies with the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies.

Criteria	Paper														
	Monreno et al., 2012	Park et al., 2013	Ehrenreich, & Underwood, 2016	Smith et al., 2017	Negriff, 2019	Eichstaedt et al., 2018	Ophir, Asterhan, & Schwarz, 2019	Settanni & Marengo, 2015	Seabrook et al., 2018	Sasso et al., 2019	Reece et al., 2017	Park et al., 2012	Ricard, Marsch, Crosier, & Hassanpour, 2018	Lup et al., 2015	Reece & Danforth, 2017
1. Was the research question or objective in this paper clearly stated?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
2. Was the study population clearly specified and defined?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N
3. Was the participation rate of eligible persons at least 50%?	Y	O	N	N	Y	Y	N	N	N	N	N	N	N	N	N
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?	Y	Y	Y	N	Y	N	Y	Y	N	Y	N	Y	N	Y	N
5. Was a sample size justification, power description, or variance and effect estimates provided?	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?	N	N	Y	N	Y	N	N	N	N	Y	N	N	N	O	N
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?	O	O	Y	O	Y	O	O	O	O	Y	O	O	Y	O	O
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?	N	N	N	N	N	N	N	N	Y	Y	N	N	N	N	N
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

(*continued on next page*)

Table 1 (*continued*)

Criteria	Paper														
	Monreno et al., 2012	Park et al., 2013	Ehrenreich, & Underwood, 2016	Smith et al., 2017	Negriff, 2019	Eichstaedt et al., 2018	Ophir, Asterhan, & Schwarz, 2019	Settanni & Marengo, 2015	Seabrook et al., 2018	Sasso et al., 2019	Reece et al., 2017	Park et al., 2012	Ricard, Marsch, Crosier, & Hassanpour, 2018	Lup et al., 2015	Reece & Danforth, 2017
10. Was the exposure(s) assessed more than once over time?	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
12. Were the outcome assessors blinded to the exposure status of participants?	Y	N	O	Y	O	Y	Y	O	O	O	O	O	O	O	O
13. Was loss to follow-up after baseline 20% or less?	Y	Y	Y	O	N	O	N	O	O	O	O	O	O	O	O
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome (s)?	N	N	N	N	N	Y	N	N	Y	Y	Y	Y	N	Y	N

Y=Yes, N=No, O=Other (CD, cannot determine; NA, not applicable; NR, not reported)

Table 2
Articles Detecting Depression on Facebook, Twitter and Instagram.

Citation	Sample Size	Sample Characteristics	Method of Depression Assessment on Social Media	Survey of Depression Assessment & Cut-off	Synopsis of Findings
Facebook					
Moreno et al., 2012	307	Aged 18 - 2054% Female	Facebook profile status updates evaluated by trained coders for depressive symptoms	Patient Health Questionnaire (PHQ-9); 5 and greater indicated depression	Depression symptoms displayed were positively correlated with PHQ-9 score ($p = 0.018$). Although insignificant, $p = 0.056$, people who displayed depressive symptoms were more likely than non displayers to score in one of: mild, moderate or severe depression.
Park et al., 2013	55	Aged 19 - 3627% female	Facebook features such as number of Facebook friends, number of likes, etc	-Center for Epidemiological Studies-Depression (CES-D); 16 for probable depression, and 25 for definite depression -Hamilton Depression Rating Scale (HAM-D) through interview with psychiatrist to participants with a 25+ score on the CES-D (which were only 2); cut-off was 7	More Facebook friends ($p = 0.08$) and location tags ($p = 0.045$) correlated with a less likelihood for depression. To an insignificant degree, depressed people tend to use the 'like' button less than nondepressed people.
Ehrenreich, & Underwood, 2016	125	Aged 1845% female	Facebook coding from the content (like Facebook status updates, wall posts, pictures, and comments) captured by a software application	Child Behaviour Checklist - Youth Self Report (CBCL-YSR)	Positive correlation for girls with depression symptoms and number of posts.
Smith et al., 2017	695	Mean age = 28.6 yearsSD = 8.974% female	Used Facebook app to extract the amount and the content of status updates, and the number of friends	Electronic Medical Records (EMR) that confirmed a diagnosis of depression	A diagnosis of depression was positively correlated with increased Facebook postings (p= .001).
Negriff, 2019	133	Aged 9 - 13 57% female	Facebook application that extracted list of friends, list of mutual friends, and timeline data of participants which included posts, comments, and likes	Children's Depression Inventory	Depressive symptoms negatively correlated with Facebook friends, and network ties.
Eichstaedt et al., 2018	683	Mean age = 29.9 yearsSD = 8.5776.7% female	Facebook application that extracted statuses	Electronic Medical Records (EMR) that confirmed a diagnosis of depression	Language predictors can detect people with depression with a fair degree of accuracy.
Ophir, Asterhan, & Schwarz, 2019	86	Aged 13 - 18 51.2% female	Facebook content analyzed by judges	Beck Depression Inventory II (BDI); cut-off score of 13	Participants who explicitly expressed distress in their posts had a higher BDI score than those who did not ($p = 0.050$).
Settanni & Marengo, 2015	201	Mean age = 28.4 yearsSD = 7.3 66% female	Automated text analysis of participant Facebook statuses and comments. Analysis included emoticons	adapted version of Depression Anxiety Stress Scales (DASS) - 21 questionnaire	Participants with a higher level of depression tended to express more negative emotions on Facebook than nondepressed people.
Seabrook et al., 2018	29	Aged 19 - 4559% female	Application called *MoodPrism* that extracted Facebook statuses, which were then analyzed. Used *LIWC 2007*	Patient Health Questionnaire (PHQ-9)	Negative emotion word instability significantly correlated with depression severity ($p = .02$).
Twitter					
Sasso et al., 2019	105	Mean Age = 19 years 57% females	Twitter Content	Beck Depression Inventory (BDI); α = 0.87 at baseline and α = 0.92 atfollow-up	The study found that participants who were tweeting with a past focus were more likely to show an increase in cognitive vulnerability and depressive symptoms than participants who did not use a past focus for their tweets (coefficient = 1.21, $t = 2.43$, $p = .02$, CI 0.21–2.21; Model R2 = 0.34, F [5, 55] = 5.55, $p < .001$).
Reece et al., 2017	204	Aged 19-6342% female	Collected user data from both the survey on MTurk and participants' Twitter history	Center for Epidemiological Studies-Depression (CES-D); cut off 21	The study found that the model was successful at discriminating between depressed and healthy content and compared favorably to general practitioners' average success rates in diagnosing depression, albeit in a separate population.
Park et al., 2012	69	Aged 17-42 40% female	Twitter Content	Center for Epidemiological Studies-Depression (CES-D); cut off 22	Compared to the normal group, the usage of words related to anger was significantly higher in the depressed group.
Seabrook et al., 2018	49	Aged 16-5765% female	The average proportion of positive and negative emotion words used, within-person variability, and instability were computed	Patient Health Questionnaire (PHQ-9)	A different pattern emerged on Twitter where greater negative emotion word variability indicated lower depression severity (r(49)=−.34, p=.01, 95% CI −0.58 to 0.09).
Instagram					
Ricard, Marsch, Crosier, & Hassanpour, 2018	749	Mean age = 26.7 yearsSD: 7.2968.8% female	Instagram profiles	Patient Health Questionnaire-8(PHQ-8) assessment questionnaire; cutoff of "15"	The 2 models, the first trained on only community-generated data (area under curve (AUC=0.71) and the second trained on a combination of user-generated and

(continued on next page)

Table 2 (*continued*)

Citation	Sample Size	Sample Characteristics	Method of Depression Assessment on Social Media	Survey of Depression Assessment & Cut-off	Synopsis of Findings
					community-generated data (AUC=0.72), had statistically significant performances for predicting depression based on the Mann-Whitney U test (*p*= .03 and *p*= .02, respectively).
Lup et al., 2015	117	Aged 18–29 84% female	Instagram use, strangers followed	Center for Epidemiological Studies-Depression (CES-D); and the Social Comparison Rating Scale; itemswere summed (a = 0.93)	Depressive symptoms increased with Instagram use if the user followed a high number of strangers, but if they followed fewer strangers, then Instagram use and depressive symptoms were unrelated.
Reece & Danforth, 2017	166	Ages: 19-55	Asked a different set of MTurk crowdworkers to rate the Instagram photographs collected	Center for Epidemiological Studies-Depression (CES-D); excluded participants with CES-D scores of 22 or higher	Instagram posts received more comments, it was more likely the posts were posted by depressed participants, however the opposite was true for the number of likes received. Depressed participants were more likely to apply Instagram filters when they posted their photos and when they did they favored the 'Inkwell' filter. In contrast, healthy participants disproportionately used the Valencia filter Lastly, they found that depressed participants' photos appeared more sad and less happy than the healthy participants.

high number of strangers, but if they followed fewer strangers, Instagram use and depressive symptoms were unrelated.

3.5.2. *Instagram posts*

The variability in photos that users post on instagram can also detect depression. Reece & Danforth (2017) examined Instagram data of people who had a history of depression. They applied machine learning tools to identify markers of depression. 43,950 Instagram photos were used to collect statistical features such as color analysis, metadata components and algorithmic face detection. They found that this model outperformed general practitioners' average unassisted diagnostic success rate for depression. Some of the patterns that they found was that if the Instagram posts received more comments, it was more likely the posts were posted by depressed participants, however the opposite was true for the number of likes received. In terms of photo editing, depressed participants were more likely to apply Instagram filtersand when they did they favored the 'Inkwell' filter, where the filter converts images to a black and white color schema. In contrast, healthy participants disproportionately used the "Valencia" filter where it brightens the tint of photos. Lastly, they found that depressed participants' photos appeared more sad and less happy than healthy participants.

3.5.3. *Community generated comments*

The study by Ricard et al. (2018) analyzed user-generated content, such as Instagram posts, and community-generated content, such as the comments made by community members on an individual's posts. The predictive model included linguistic features extracted from instagram post captions and comments, including multiple sentiment scores, emoji sentiment analysis results and meta-variables such as the number of likes and average comment length. The researchers found that community generated comments on Instagram posts can be used to identify moderate to severe depression, but user-generated data did not yield significant predictive value.

3.6. *Snapchat*

No studies on Snapchat were identified.

4. Discussion

4.1. *Facebook*

From the studies analyzed, it appears that screening social media profiles on Facebook, Instagram, and Twitter, could be a valid way to assess depression. Facebook studies show that having fewer Facebook friends and mutual friends, posting frequently, and using fewer location tags are positively correlated with depressive symptoms. If the content of Facebook posts explicitly express depressive symptoms, then this may also indicate that the poster has depression. Finally, more subtle language choices like an increased use of personal pronouns, and words related to pain, depressive symptoms, aggressive emotions, and rumination may also predict depression.

4.2. *Twitter*

For Twitter, we found that depression can possibly be detected by conducting linguistic analysis by examining the types of words, linguistic style, content and expression. Specifically, Twitter users who use "past focus" words such as "learned" or "remembered" were associated with ruminative brooding, which is a symptom of depression. Also, negative emotions and anger words were related to having depression. In addition, having fewer words per Tweet was positively correlated with depression. Reece et al. (2017) contends this result is expected as depression often results in reduced communication.

4.3. *Instagram*

For Instagram, differences in follower patterns, photo posting and editing, and linguistic features between depressed people and nondepressed people could serve as a marker for detecting depression. One interesting finding was that an increase in the number of hours using Instagram was associated with greater depressive symptoms, but this was only true for those who followed a high number of strangers (Lup et al., 2015). This effect might be due to comparing themselves with strangers that they follow. Specifically, since the users do not know the strangers they follow well, this might lead to greater negative social comparisons than if they knew the person better. This is because the user will not be able to integrate other knowledge into the perception they have of the stranger. For example, if the user sees an expensive car that a stranger owns that they cannot afford, the comparison of the relative affluence might lead to greater sense of social and financial deprivation. In contrast, if the user is friends with the person they follow, seeing the same post of the car may not lead to this same sense of deprivation, because the user may know that their friend bought the car second hand. Thus, following fewer strangers can reduce misconceptions and misinterpretations, and thus decrease the chance of negative social comparisons. Alternatively, following more strangers can increase this chance, which may worsen depressive symptoms.

4.4. Application

There can be an application developed to detect depression on social media, that takes into account all of the studies reviewed in this paper. Past studies have already proven the effectiveness of an application such as this. For example, Reece et al. (2017) created a model with supervised learning algorithms by extracting predictive features such as affect, linguistic style and context from participant tweets (N=279,951). The study found that the model was successful at discriminating between depressed and healthy content and compared favorably to general practitioners' average success rates in diagnosing depression. In fact, results held even when the analysis was limited to content posted before first depression diagnosis. If the AI detects depression, resources can be displayed to the user through the social media site. There would be several benefits to this. For one, users will be notified if they might be depressed, and along with the resources provided, can get the help they need early. Secondly, it would provide public health data that can uncover the extent of depression in different demographic groups, and thus be informative in public policy and funding decisions. Finally, the application will be able to provide insight into whether a treatment is working or not, based on if it no longer detects depression from a profile.

This systematic review can enhance models that have been created to detect depression through social media by providing information on additional markers that can be included in them. For example, Ricard, Marsch, Crosier, & Hassanpour (2018) created a model that uses community-generated content, such as comments made by followers, and user-generated content, to detect depression among social media users. The study found that the model had statistically significant performances for predicting depression based on the Mann-Whitney U test. Ricard et al.'s (2018) study provided new insights into the next generation of population-level mental illness risk assessment and intervention delivery. Including other social media markers that have been discussed in this systematic review, such as photo editing filters, can further increase the accuracy of the model for detecting depression. However, a robust experimental design is needed for such an enhanced model to be a valid assessment tool for depression, especially considering the limitations of this review.

Unfortunately, there are several limitations to a model that can automatically detect depression online. Firstly, some may believe that it is unethical to monitor personal user accounts for signs of depression (Mikal et al., 2016). Additionally, the application may be prone to errors insofar as identifying incorrectly people with depression who do not meet criteria for depression (i.e. false positives). Although this would not likely have a significant negative impact, it may cause users some stress to be labeled as depressed. The application may also incorrectly identify depressed people as no longer depressed, giving mental health workers a false impression of their people.

4.5. Limitations and future directions

There were several limitations to our study, mostly which have to do with the lack of consistency in the methods across the included articles, which consequently limits the comparisons that can be made between them. Firstly, the studies analyzed had participants of varying demographic backgrounds. For example, a study by Negriff (2019) had participants' age ranging from 9 to 13 while other studies had participants with age older than 13. Therefore, the finding of Negriff (2019) cannot be generalized to the overall finding in this systematic review especially when it is not clear if the depressive social media markers shown in adolescents would also transfer to adults with the lack of research in this area. Also, the studies did not consider the cultural differences that people might have based on their ethnicity. Specifically, the effect of having certain social media platforms being banned in some countries (i.e. Facebook is banned in China) could have affected the results because there would have been no studies on these platforms in these areas, and thus would have not been included in our systematic review.

Similarly, the studies in this review were conducted at different times between the years 2012 and 2019. The way people used social media earlier in this time span is substantially different than how it was used later on (Koiranen et al., 2019), since earlier on social media was newer. This is another reason why the results of this study may not be generalizable.

Moreover, among studies that used the same depression scale, the cut-offs were not always the same. For example, Monreno et al. (2012) used a cut-off of 5, but a cut - off between 8 to 11 has been found to be more valid for PHQ-9 to detect major depressive disorder (Manea et al., 2012). Therefore, the method of evaluating for depression was not the gold standard for some studies.

Further, the p values reported in the analyzed studies may be misleading since in each study many factors were extracted from social media profiles. Due to the higher statistical familywise error rate, it is possible that some of these factors were significantly correlated to depression scores, even though there was no relationship. Also, effect sizes were not reported in any of the studies. Thus, from a statistical standpoint, the results of this review cannot be conclusive.

Additionally, different methods of screening in each study might have affected the overall results. Specifically, some of the studies used an application to extract account data, and others manually analyzed social media accounts. In fact, the latter technique is likely prone to more subjection than the former. Also, some of the studies screened only public profiles, while other screened accounts regardless of privacy settings. For example, a study by Manea et al., (2012), only looked at public Facebook profiles, thus possibly skewing their results because there may be differences between users who use their social media privately and publicly.

A limitation present across all studies is that the social media content might not be able to capture everything about the person and it might not reflect the mood they seem to portray on the questionnaires because people craft their images on their social media (Chua & Chang, 2016). Additionally, people might be afraid to share all their authentic thoughts or moods due to the stigma around mental illness and this might question the practical aspect of using social media contents to detect depression in users.

Due to all the limitations discussed, it cannot be concluded that social media is a valid way to assess depression. However, this study does support the potential for this. Future studies should analyze the differences in depression expression taking demographic variables into account. They should calculate effect size, and use a standardized method to detect depression both on the social media platform and through a validated questionnaire. Finally, more studies should look at how to detect depression on other popular social media platforms such as Snapchat, Reddit, Whatsapp, and YouTube.

With enough validated information, a predictive tool can be developed and integrated into social media sites. This tool would recognize depressive symptoms and advise the user to seek medical assistance. It may also be used to notify the user's loved ones or doctor, so that they can check-in with them. However, it will be crucial to understand that the tool will not be able to diagnose depression, since people can exaggerate depressive symptoms, and because the tool can be prone to errors.

5. Conclusion

Candidates for depression show different patterns of social media use. It is valuable to find social media markers for depression as social media sites are increasingly popular outlets for user thoughts and feelings. Current literature has shown that depression could be detected through social media contents. Specifically, all three social network sites, Facebook, Twitter, and Instagram showed that they could be avenues of depression assessment. Social media markers found in Facebook studies were related to friends, location tag and posts with explicit expression of depression symptoms. For Twitter studies, markers much as use of words and number of words were shown. Lastly, for Instagram, follower pattern, photo posting and editing and linguistic features served as potential markers for depression. Some limitations of this systematic review is that the primary articles analyzed had different demographics of participants and methods used to diagnose depression, which constricts the amount of comparisons and generalization that can be made. Further, only four social media platforms were explored and there are possibilities that different social markers for depression could be found in other social media platforms.

Role of the Funding Source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Jiin Kim: Conceptualization, Methodology, Investigation, Writing - original draft. **Zara A. Uddin:** Conceptualization, Methodology, Investigation, Writing - original draft. **Yena Lee:** Writing - review & editing, Supervision. **Flora Nasri:** Supervision. **Hartej Gill:** Writing - review & editing. **Mehala Subramanieapillai:** Supervision, Validation. **Renna Lee:** Conceptualization, Methodology. **Aleksandra Udovica:** Conceptualization, Methodology. **Lee Phan:** Writing - review & editing. **Leanna Lui:** Writing - review & editing. **Michelle Iacobucci:** Supervision, Validation. **Rodrigo B. Mansur:** Supervision, Validation. **Joshua D. Rosenblat:** Supervision, Validation. **Roger S. McIntyre:** Supervision, Validation, Project administration.

Declaration of Competing Interest

Roger S. McIntyre has received research grant support from the Stanley Medical Research Institute and the Canadian Institutes of Health Research/Global Alliance for Chronic Diseases/National Natural Science Foundation of China and speaker/consultation fees from Lundbeck, Janssen, Shire, Purdue, Pfizer, Otsuka, Allergan, Takeda, Neurocrine, Sunovion, and Minerva. Dr. Joshua D. Rosenblat has received research grant support from the Canadian Cancer Society, Canadian Psychiatric Association, American Psychiatric Association, American Society of Psychopharmacology, University of Toronto, University Health Network Centre for Mental Health, Joseph M. West Family Memorial Fund and Timeposters Fellowship and industry funding for speaker/consultation/research fees from Janssen, Allergan, Lundbeck, Sunovion and COMPASS. He is the medical director of a private clinic providing intravenous ketamine infusions and intranasal esketamine for depression. Author Yena Lee received salary support from the Global Alliance for Chronic Diseases/Canadian Institutes of Health Research (CIHR)/National Natural Science Foundation of China's Mental Health Team Grant and the CIHR Frederick Banting and Charles Best Canada Graduate Scholarship; personal fees from Champignon Brands. All other authors have no conflicts of interest to disclose.

References

Ahmedani, B.K., 2011. Mental health stigma: Society, individuals, and the profession. J. Social Work Values Ethics 8 (2), 4, 1.

Ben-Zeev, D., Young, M.A., Madsen, J.W., 2009. Retrospective recall of affect in clinically depressed individuals and controls. Cognit. Emotion 23 (5), 1021–1040.

Chua, T.H.H., Chang, L, 2016. Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. Comput. Hum. Behav. 55, 190–197.

"Depression." World Health Organization. World Health Organization, December 4, 2019. https://www.who.int/news-room/fact-sheets/detail/depression.

Eaton, W.W., Neufeld, K., Chen, L.S., Cai, G., 2000. A comparison of self-report and clinical diagnostic interviews for depression: diagnostic interview schedule and schedules for clinical assessment in neuropsychiatry in the Baltimore epidemiologic catchment area follow-up. Arch. Gen. Psychiatry 57 (3), 217–222.

Ehrenreich, S.E., Underwood, M.K., 2016. Adolescents' internalizing symptoms as predictors of the content of their Facebook communication and responses received from peers. Transl. Issues Psychol. Sci. 2 (3), 227.

Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preoţiuc-Pietro, D., …, Schwartz, H.A., 2018. Facebook language predicts depression in medical records. Proc. Natl. Acad. Sci. 115 (44), 11203–11208.

Koiranen, I., Keipi, T., Koivula, A., Räsänen, P., 2019. Changing patterns of social media use? A population-level study of Finland. Universal Access Inf. Soc. 1–15.

Lup, K., Trub, L., Rosenthal, L., 2015. Instagram# instasad?: exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. Cyberpsychol. Behav. Social Netw. 18 (5), 247–252.

Manea, L., Gilbody, S., McMillan, D., 2012. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ 184 (3), E191–E196.

Mikal, J., Hurst, S., Conway, M., 2016. Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. BMC Med. Ethics 17 (1), 22.

Moreno, M.A., Christakis, D.A., Egan, K.G., Jelenchick, L.A., Cox, E., Young, H., Becker, T., 2012. A pilot evaluation of associations between displayed depression references on Facebook and self-reported depression using a clinical scale. J. Behav. Health Serv. Res. 39 (3), 295–304.

National Heart, Lung and Blood Institute (n.d.). *Study Quality Assessment Tools*. Retrieved May 25, 2020, from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools?fbclid=IwAR3ZOnJNHrQ-YWiPnSbTFNbFiAM-GQ4Ycvnem2PH8bbYL51zUZnOOUExLVU.

Negriff, S., 2019. Depressive symptoms predict characteristics of online social networks. J. Adolesc. Health.

Ophir, Y., Asterhan, C.S., Schwarz, B.B., 2019. The digital footprints of adolescent depression, social rejection and victimization of bullying on Facebook. Comput. Hum. Behav. 91, 62–71.

Park, M., Cha, C., & Cha, M. (2012). Depressive moods of users portrayed in Twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)* (Vol. 2012, pp. 1-8).

Park, S., Lee, S.W., Kwak, J., Cha, M., Jeong, B., 2013. Activities on Facebook reveal the depressive state of users. J. Med. Internet Res. 15 (10), e217.

Radovic, A., Gmelin, T., Stein, B.D., Miller, E., 2017. Depressed adolescents' positive and negative use of social media. J. Adolesc. 55, 5–15.

Reece, A.G., Danforth, C.M., 2017. Instagram photos reveal predictive markers of depression. EPJ Data Sci. 6 (1), 1–12.

Reece, A.G., Reagan, A.J., Lix, K.L., Dodds, P.S., Danforth, C.M., Langer, E.J., 2017. Forecasting the onset and course of mental illness with Twitter data. Sci. Rep. 7 (1), 13006.

Ricard, B.J., Marsch, L.A., Crosier, B., Hassanpour, S., 2018. Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram. J. Med. Internet Res. 20 (12), e11817.

Sasso, M.P., Giovanetti, A.K., Schied, A.L., Burke, H.H., Haeffel, G.J., 2019. # Sad: Twitter content predicts changes in cognitive vulnerability and depressive symptoms. Cognit. Therapy Res. 43 (4), 657–665.

Seabrook, E.M., Kern, M.L., Fulcher, B.D., Rickard, N.S., 2018. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J. Med. Internet Res. 20 (5), e168.

Settanni, M., Marengo, D., 2015. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Front. Psychol. 6, 1045.

Smith, R.J., Crutchley, P., Schwartz, H.A., Ungar, L., Shofer, F., Padrez, K.A., Merchant, R.M., 2017. Variations in facebook posting patterns across validated patient health conditions: a prospective cohort study. J. Med. Internet Res. 19 (1), e7.