Depression Detection on Social Media: A Classification Framework and Research Challenges and Opportunities

Abdulrahman Aldkheel (aaldkhee@uncc.edu)
The University of North Carolina at Charlotte
Lina Zhou
The University of North Carolina at Charlotte

Research Article

Keywords: social media, depression, depression detection, mental health, machine Learning

Posted Date: April 13th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2783567/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Healthcare Informatics Research on November 20th, 2023. See the published version at https://doi.org/10.1007/s41666-023-00152-3.

Abstract

Objective: Social media has become a safe space for discussing sensitive topics such as mental disorders. Depression dominates mental disorders globally, and accordingly, depression detection on social media has witnessed significant research advances. This study aims to review the current state-of-the-art research methods and propose a multidimensional framework to describe the current body of literature relating to detecting depression on social media.

Method: A study methodology involved selecting papers published between 2011 and 2022 that focused on detecting depression on social media. Three digital libraries were used to find relevant papers: Google Scholar, ACM digital library, and ResearchGate. In selecting literature, two fundamental elements were considered: identifying papers focusing on depression detection and including papers involving social media use.

Results: In total, 46 papers were reviewed. Multiple dimensions were analyzed, including input features, social media platforms, disorder and symptomatology, ground truth, and machine learning. Various types of input features were employed for depression detection, including textual, visual, behavioral, temporal, demographic, and spatial features. Among them, visual and spatial features have not been systematically reviewed to support mental health researchers in depression detection. Despite depression's fine-grained disorders, most studies focus on general depression.

Conclusion: Recent studies have shown that social media data can be leveraged to identify depressive symptoms. Nevertheless, further research is needed to address issues like depression validation, generalizability, causes identification, and privacy and ethical considerations. An interdisciplinary collaboration between mental health professionals and computer scientists may help detect depression on social media more effectively.

1 Introduction

Depression, or clinical depression, is a mood disorder that affects people of all ages. This condition can result in severe symptoms that adversely affect an individual's mood, thoughts, and ability to carry out basic daily activities, such as sleeping, eating, and working [1]. Depressive disorder is diagnosed when an individual has at least five symptoms of depression almost every day for at least two weeks. In most cases, depression will manifest as a persistent feeling of sadness or an inability to enjoy almost all activities [1].

Nowadays, depression has become the top diagnosed mental disorder worldwide [2]. It is also one of the most predominant mental health problems globally. According to the World Health Organization, depression affects around 350 million people [2]. The impact of depression on one's life can be very harmful and severe and dramatically change the individual's whole life. For instance, depression might affect work performance, physical health, mood, and relationships and even lead to suicidal thoughts [3]. According to the National Institute of Mental Health, depression is considered the second leading cause of death in people aged 15–29; and more than 17 million people in the U.S. suffer from at least one major depressive episode. The seriousness of depression lies in its possibility of prolongation, which may even last for years [4]. Multiple factors may cause depression, including biological, personality, interpersonal, environmental, or religious factors [5]. The ongoing pandemic COVID 19 has shown a significant impact on an individual's mental health and psychological status [6], [7], exacerbating depression around the world [6].

Social media has become a fertile space for sharing emotions and even psychological disorders that an individual may suffer during his/her life. Depressive users' posts can be viewed as an early warning, which emphasizes the importance of spotting symptoms early, as abnormal behavior concerns can be observed. The early detection and appropriate treatment of depression can help prevent and alleviate its symptoms, leading to an improved state of health [8]. In addition, early detection and interventions can reduce the emotional and financial worries associated with depression [8]. Detecting depression on social media can serve as a new monitoring method that assists physicians in determining the correct diagnoses of their patients. With the option of keeping anonymous on its various platforms, social media might become a favorable environment for users who suffer from depression and are unwilling to talk to clinic specialists because of stigma, individual discrimination, and lack of knowledge [9]. This literature survey aims to investigate state-of-the-art methods for detecting depression on social media.

There have been a few related survey studies [10]–[14]; however, they differ in terms of research focus, reviewed dimensions, and/or user population. For instance, [12] covered a large amount of literature, but it did not provide comparisons between different studies in terms of methods and selected features for depression detection. In addition, the study did not discuss the role of machine learning in the prediction of depression in social media users and was focused on the population of immigrant college students. A recent study [14] surveyed only text mining methods for depression detection on social media. Similarly, another study [11] focused on the assessment criteria of depression and its detection methods on text-based platforms. Nonetheless, none of the studies has considered the role of image data in depression detection, which has been observed to be effective in representing users' emotions more vividly. Psychologists have notably shown that imagery can be an effective way to convey difficult emotions [15]. Additionally, none of the above studies has discussed the effects of temporal and spatial features in depression detection on social media. The temporal features can be used to understand the behavioral patterns of user posting, and the spatial features are useful for interpreting depression with unique demographic characteristics of specific neighborhoods or communities, as well as for understanding an individual's geographic behavior when they are depressed.

This study aims to fill the gaps from the literature as mentioned above by conducting a systematic review. Our major contributions are the following. First, we aim for both depth and breadth in this review. Specifically, we investigate detecting depression on social media from diverse perspectives, including social media platforms, mental disorder and symptomatology, ground truth, input features, and machine learning techniques. Meanwhile, we examine depression detection methods in great depth to understand state-of-the-art techniques and performance. Second, we discuss the role of image data in depression detection on social media by summarizing the important visual features that contribute to the detection process. Third, our study confirms that some linguistic and emotional features (e.g., negative emotions, first-person pronouns, depression terms) have consistently shown to have strong associations with depression. On the other hand, we also reveal inconsistent findings regarding the effects of user behaviors such as user behavior and engagement on detecting depression on social media across different studies; and more importantly, we provide explanations for the differences from multiple aspects, such as extracted features, user expression behavior, and depressive type, and level.

2 Method

We select papers that focus on depression detection on social media published during the period (2011–2022). To ensure wide coverage of this topic, we use three digital libraries for publications search, including Google Scholar, ACM digital library, and ResearchGate. Our literature selection strategy comprises two fundamental elements: a) detection of depression, and b) use of social media. Accordingly, we use a combination of related keywords: depression, detection, mental health, and social media or social network site as search queries. Then, we exclude those papers that did not use social media as primary source but used offline environment (e.g., interview, survey) or other types of an online environment (e.g., smartphone apps) as primary source of their data collection (n = 21). Additionally, we do not include publications detecting other types of disorders on social media (e.g., eating disorders, stress) (n = 88). Further, we exclude studies that collect data from non-English speaking users (e.g., Japanese, Chinese, etc.) (n = 8). Finally, we obtain a set of 46 papers. Figure 1 describes the steps of our paper selection process.

Based on an analysis of the publication count, this line of research has experienced rapid growth in the recent 12 years and peaked in 2018 (See Fig. 2).

3 Depression Detection Process

The objective of depression detection in this study is to evaluate whether or not a social media user is depressed based on their social media data. Based on the detection method, the process of detecting depression on social media can be classified into two main categories: a) machine learning based, and b) non-machine learning based methods (e.g., manual). Accordingly, the two types of detection processes are depicted separately in Fig. 3 and introduced separately in the remainder of this section.

3.1 Detection using machine learning

The detection process for machine-learning methods consists of four main stages: data collection, data preprocessing, feature extraction, and classification.

1. Data collection: This phase aims to gather information on the attributes of interest, contributing to answering research questions. Data can be collected using a variety of methods, including API, web scrapping, or manual observation [16], and the sources of data include self-declared posts,

comments, user profiles, and metadata. Additionally, social media data can come in multiple modalities, such as text and images.

2. Data preprocessing: The objective of this step is to transform the raw data into an understandable form. Usually, raw data may come with the issues of inconsistency, missing values, and errors. These issues may be addressed using one or more of the four strategies: data cleaning (e.g., addressing missing values, data inconsistencies, and duplications), data integration (integrating data from multiple sources), data transformation (converting raw data into a required form to be compatible with model, normalization, aggregation, generalization), and data reduction (getting a reduced representation of data that yields the same analytic findings). There are other preprocessing tasks that are data-dependent. For instance, text data may go through the preprocessing steps including URL and stop words removal, stemming, and lemmatization.
3. Feature extraction and representation: It is the process of extracting the indicative features from the preprocessed data. Different types of social media features may contribute to detecting depression, such as linguistic, behavioral, visual, emotional, and demographic features [16].
4. Classification: This step builds models using the extracted features as the input and classifies each social media instance into the depression or non-depression class, or levels of depression. Thus, depression detection can be a binary, multiclass, or multi-label classification problem. Common classification techniques include Support Vector Machine, Logistic Regression, Decision Tree, Naive Bayes, etc.

3.2 Detection without machine learning

The manual detection process consists of five main steps: data collection, measures and unit selection, data coding, validation, and results analysis.

1. Data collection: Similar to what we described above in Section 3.1.
2. Measures and unit selection: It captures the primary units that researchers intend to analyze and code. In the context of depression detection on social media, the units could be posts, comments, likes, captions, etc.
3. Data coding: This process consists of three sub-phases. a) Developing a coding scheme: The development of coding scheme can be done inductively and/or deductively. Researchers might create a codebook based on theories from prior literature and then refine it during training. b) Preliminary coding: This refers to pre-testing the coding scheme by using a sample of data. It is essential to ensure consistency and validity during this step. c) Formal coding: The coding process can be conducted on the entire dataset. With the guide of the codebook, researchers can go through data units and assign each relevant data under the proper theme(s).
4. Validation: This phase refers to data checking and validation to ensure the data reliability and accuracy. This can be done in many ways, such as involving multiple coders, studying the consistency of measurement over time, etc.
5. Result analysis: In this final phase, researchers review and finalize their work, indicate the most important patterns, determine the observable factors of detailed information, examine correlation,

review the inferences, and present the results in an understandable format like graphs, matrices, etc. For instance, what are the most common themes that have been observed? which symptoms are most frequently mentioned in the datasets? which gender has more exposure to depression? what age group is highly vulnerable to depression?

4 A Classification Framework For Detecting Depression On Social Media

Based on our review of the selected literature on detecting depression from social media, we propose a multi-dimensional classification schema of the literature. The schema consists of five main dimensions: depressive disorders and symptomatology, social media platforms, depression features, ground truth, and machine learning techniques (see Fig. 4). In general, research designs and investigations revolve around the detection of depression based on textual, visual, behavioral, and other cues extracted from social media data. In our categorization, we highlight the basic elements that are used for depression detection (i.e., data sources, features, machine learning techniques).

4.1 Types of depressive disorders

Depression comes in many different forms, which vary in terms of the severity of symptoms, among others. It is noted from our literature review that the majority of the studies have examined depression in general. In other words, researchers used the term depression without providing a specific definition to guide the identification of positive samples. Only some studies have focused on specific types or levels of depressive disorders (See Fig. 5), such as bipolar depression [17], major depressive disorder (MDD) [18]–[21], postpartum depression [22], [23], suicidal ideation risk caused by depression [24], [25], and different severity levels of depression [26]–[28].

4.2 Social media platforms

Researchers have adopted different social media platforms as the primary data sources for depression detection research (See Fig. 6). The figure shows that most of the studies (69%) collected data from Twitter and/or Facebook. This is because they are among the most popular social media platforms [29], which have a large base of users who might share their personal experiences of mental disorders. It is also because these platforms support multimodal communications such as text, image, and video, empowering users with diverse means of self-expression. In addition, Reddit accounts for a significant percentage (17%) of the reviewed studies, where user profiles are anonymous, and users can seek emotional support, social support, advice, etc. Further, Instagram has a presence among the reviewed studies (12%) owing to its affordance of visual features that can shed light on the psychological state of a social media user.

4.3 Ground truth for depression

Establishing the ground truth for depression is a critical issue for depression detection research on social media. We grouped the existing methods for determining the ground truth into four main clusters: survey, interview, self-disclosure, and community membership (See Table 1).

- Survey: Survey is one of the dominant methods used by the related studies [23], [27], [30]–[39]. Questionnaires are used to collect information from participants regarding their social media activities as well as their depression history in order to assess the depression level of an individual. This self-reported survey demonstrates its efficacy in measuring the user's depression state with a high level of reliability, consistency, and certainty. Sample common survey instruments include Patient Health Questionnaire (PHQ) [40], Center for Epidemiologic Studies Depression scale (CES-D) [41], Beck Depression Inventory (BDI) [42], the 100- item International Personality Item Pool (IPIP) [43].
- Interview: A face-to-face interview would be able to elaborate on the participants' social media use and posted content, as well as uncover possible explanations for their behavior. Several studies [23], [30], [38] conducted semi-structured or structured interviews as an assistive and validation method for depression detection.
- Self-declared: Users may choose to disclose their depressive mental state publicly in online communication. Such a disclosure can be conveyed through the use of words/phrases related to depression or its symptoms in a post, comment, or on user profiles. For example, "I am diagnosed with depression", and "I am depressed today".[17]–[20], [22]–[28], [30], [33]–[37], [39], [44]–[58].
- Community membership: Online communities (e.g., Reddit) have become an important source of texts on mental health topics. They provide a venue where members can discuss their mental health status, seek support and advice, and receive emotional and informational support [24], [58]–[60].

Different sources were used to identify control data (non-depressed users or posts). For example, some studies treated the absence of signs of depression as non-depressed data by excluding those users who were not diagnosed with depression (e.g., questionnaire) [19], [23], [31], [33], [34], [36], or randomly selected data from the social media platform to construct a control group [56]–[58], [61], or selected a group of users who did not disclose that they were depressed [24], [37], [51], [57], [59].

4.4 Features in support of depression detection

Input features serve as clues or signals of depression. We classify the features used to predict depression in the related studies into the following categories: textual, visual, behavioral, demographic, temporal, and spatial features. Table 2 presents a summary of the different types of input features.

1. Textual Features. Text features are extracted from the textual content of social media posts. We further classify text features into the following sub-categories.

- Topics: Topics reflect the main themes of social media posts. Many studies [27], [51], [55], [57], [59], [60], [62] have leveraged topic modeling techniques (e.g., Latent Dirichlet allocation (LDA)) to extract topic features. LDA is a generative model that assumes documents as an outcome of combined topics, in which each topic has its word distributions. One study [45] employed Biterm Topic Modeling (BTM) to extract and identify depression-related topics.
- Linguistic and stylistic features: They refer to the choice of linguistic constituents such as words to express meaning. Many studies have leveraged Linguistic Inquiry and Word Count (LIWC) to extract linguistic style or psycholinguistic features [17], [19], [23], [24], [27], [33], [34], [36], [37], [46], [47], [49]–[51], [57], [59], [61]. For instance, LIWC has been sued to extract personal pronoun, conjunction, negation, and auxiliary verbs [23], [33], [37], [46], [47], [61]. Other studies [24], [55], [58], [63] focused on morphological features such as verb, pronoun, noun, adjective, and adverb by applying part of speech tagging. Further, one study [51] leveraged readability and coherence features. Moreover, one study [17] utilized Empath [64] that is comparable to LIWC, but neural embeddings are used to classify the words based on similarities.
- Word representation models: They concern the representation of word patterns and/or distributions. This is one of the most actively researched areas in natural language processing. Sample techniques for modeling word representations include TF-IDF [17], [19], [22], [58], [59], [61], [62], Bag of Word (BOW) enhanced with synonyms from the WordNet [61], n-gram (e.g., bigram) features [19], [27], [33], [51], [54], [58], [59], [61], [63], word embedding such as Word2Vec [48], a representation of words that enables words with similar meanings to have the same representations.
- Emotions: In the context of depression detection on social media, it is crucial to understand what the user feels and what emotions are expressed in a post or comment. Identifying words linked to basic emotions is essential in understanding a user's mental health. Thus, the rate of negative and positive words in a given text has been frequently used to determine the emotional status and measure the probability of having depression [19], [23], [26], [28], [33]–[37], [45]–[47], [50], [51], [53], [55], [57], [59], [62], [63]. It is becoming increasingly possible to extract emotions using techniques and tools such as LIWC [17], [19], [28], [33], [34], [37], [49]–[51], [57], [59], EMOTIVE system [46], [47], Valence Aware Dictionary and Entiment Reasoner (VADER) [45], ANEW [35], [57], LabMT [35], emoji sentiment [35], and polarity of emotions [51] obtained from AFINN [65], which depression detection studies have leveraged.
- Domain dictionaries: Studies have extracted domain-specific textual features from social media posts using domain-specific dictionaries [51], [56], [57]. This involves building dictionaries of depression-related terms. For example, a study constructed lexicons of antidepressant names and depression symptoms, which contain symptoms listed in DSM-IV criteria, to support depression detection [57].

2. Visual Features: In view that images could contain meaningful psychological information for the detection of mental disorders, some studies have analyzed the visual content of social media to extract major symptoms of depression [26], [31], [32], [39], [44], [57], [66]. We group visual features

explored in the above studies into four main categories: color properties, object features, aesthetics, and emotions.

- Color Properties: They include all attributes of colors regardless of whether they are subjective or objective. The subjective attributes measure hues, saturation, and value (HSV), while the objective attributes describe luminance, wavelength, and pureness. The most well-known color spaces are CIE, HSV, RGB, XYZ, and YUV. Extraction of these properties could be used to determine the presence of symptoms of depression. For instance, [26] characterized colors in social media profiles as dark versus light. Studies [31], [39], [57] have explored the color space HSV at the pixel level, which offers better color representation for humans [67]. Their results show that depressed users are more likely to post images that have lower saturation(grayer) and lower brightness(darker) [31], [39].
- Object Features: They to the information that is embedded in a posted image. Image features can come in various structures such as points, edges, and objects. Some researchers employed advanced computer vision techniques to detect objects in an image. For instance, face detection software was used to recognize human faces in both posted and profile images [31], [39], which indicates possible signs of depression. Guntuku et al. [39] noted that posted images from the depressed group did not contain faces.
- Aesthetics: This refers to an appreciation of the beauty of an image. Aesthetics can be measured in various aspects, such as balancing elements, repetition, motion blur effect, symmetry, and depth of field. Only one study [39] has employed aesthetic features, which shows that depressed users posted images with low aesthetics score.
- Emotions: Image is a natural way of expressing nonverbal behavior such as emotion. Emotions are closely linked to the object features, as discussed earlier. For instance, based on the facial expressions extracted from user profile images [39], depressed users were found to post images with more positive rather than negative emotions.

3. Behavioral Features: They refer to the frequency and types of behavioral activity and user interaction with others. Sample features are the number of followers and followees, number of inbound/outbound likes, number of posts, posted and received comments, length of post/comment, upvote and downvote (e.g., in Reddit), links, media items, status updates, and others. Behavioral features can be significant for identifying users who suffer from depression because this group of users tends to be less engaged in social interactions and have fewer social connections. A considerable number of studies [17], [19], [23]–[26], [30]–[35], [37], [47], [51], [53], [55], [57], [61], [63], [66] have employed Behavioral features in depression detection. For instance, Alsagri and Ykhlef, [61] employed the total number of posts, hashtags, mentions, emojis, retweets, replies, the number of following, followed, etc. In the same vein, authors examined the relationship status, group memberships, photo tags, and the number of user's likes [34]. Similarly, based on an analysis of multiple engagements and ego-centric measures such as volume, rate of replies, fraction of links, proportion of retweets, rate of question-centric posts, and number of followers/followees, De

Choudhury et al. [19] noted lower values for volume, replies, and number of followers for the depressed than for the non-depressed group.

4. Temporal Features: Time series is a record of an event that fluctuates over time [50]. The distribution of posts over time is also substantial to understanding mental health. Temporal features can contribute by indicating whether the depressed users are more likely to post during the daytime or nighttime and which day/days of the week have more frequent posts than others [19], [27], [33], [34], [37], [46], [50], [53], [55], [57], [58]. For instance, depressed users had different peak times and behaved differently compared to non-depressed users [50]. Specifically, depressed users were more active from midnight to mid-day, which was the opposite to the behavior of non-depressed users. Another study [19] also showed that depressed people were more active at night and inactive during the daytime. An analysis of the associations between the time of posting and the sleeping quality revealed that users who write a negative post late at night are more likely to experience sleeping issues and thus at a greater risk of depression [53]. These studies were consistent with the theory that users with depression signs are relatively more active during the evening and late at night [68].
5. Demographic Features: There is a strong association between demographic factors and the mental health status of individuals [69]. Demographic information provides an accurate picture of a patient to the mental health professional and enables a clear inference of who is greatly affected by a mental health problem. It has been widely used in related studies [18]–[20], [23], [32]–[36], [38], [39], [57], [70]. For instance, based on age, gender, and race inferred from content analysis of tweets, users who wrote tweets about their feelings of depression were more likely (94%) to be under 25 and be female (77%) [18]. Similarly, another investigation of gender differences in depression rates showed that women were exposed (1.5 times) to depression more than men .
6. Spatial Features: They refer to geographic locations and spatial information. A study shows that the actual state-level depression rates and the ratio of depression tweets in the United States were geographically highly correlated [33]. Several studies employed geolocation features to understand how the behavior of depressed users may differ from non-depressed users [30], [33], [52], [53]. For instance, depressed users shared fewer geotags and moved around less than non-depressed users [30]. Husain [53] also observed that depressive users' location remains constant for a long time (i.e., at home) when they post on social media.

Table 2
A summary of features indicative of depression in social media

Features		Description (Example)
Text	Topical	Determine the common themes and meaningful links between the topic in the datasets (e.g., LDA, BTM).
	Linguistic and stylistic	Describe how the language is used by the users, and give more details about their attitude (e.g., LIWC, SLIWC,Empath).
	Word representation models	The pattern of words, characters distributions and frequency of words (e.g., N-gram, TF-IDF, BOW, word embedding)
	Emotion	Determine the polarity and emotion that are embedded in the text (e.g., LIWC, ANEW, LabMT)
	Domain	Textual features specific to domain of depression
Visual	Color	Characteristics of color whether it is subjective or objective (e.g., HSV (hue, saturation, value), RGB)
	Object	Information embedded in the content of posted image (point, edge, object)
	Aesthetics	Appreciation of beauty of an image (balancing, motion blur, repetition)
	Emotion	All related emotional data that can be identified in an image
Behavioral	Record the behavioral activity of the users (number of followers and following, number of likes, number of posts)	
Temporal	The time series and distribution of posts over time	
Demographic	Identify what personal demographic information that lies within posts or user profiles or through surveys/interviews (e.g., age, gender, education level)	
Spatial	Determine the effect of geographic location (geotags, GPS)	

It is worth noting that the effects of some input features are inconsistent across different studies. For instance, some studies [19], [50], [53] showed that depressed users were more active after midnight and had decreased levels of activity throughout the day. On the other hand, one study [58] did not find any difference in posting time between depressed and non-depressed users. In addition, [26] found that users who suffer from severe depression wrote more posts, whereas [33] observed that depression-indicative posts had a lower volume because of users' weakness of social connections. Further, one study [31] showed that sadness is a significant predictor of depression in Instagram images. Another study [39] concluded that depressed users tend to share profile images that suppress positive emotions instead of expressing negative emotions and appear to be less expressive and more neutral. The latter study attributed the finding to social media self-presentation biases.

4.5 Machine learning techniques

Machine learning and/or statistical learning techniques are the key to building models for detecting and predicting depression in social media users. There are three major types of machine learning techniques: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is focused on learning the mapping between inputs and outputs and is only applicable when all the data is labeled. Unsupervised learning determines the pattern by using unlabeled data or when there is a lack of labeled training data. Like unsupervised learning, reinforcement learning does not require labeled data. Reinforcement learning allows an agent to learn to accomplish a goal in an interactive environment, aiming to increase the possibility of reward by taking a specific action. Among them, supervised techniques dominate depression detection studies.

Supervised techniques have been widely used to build models for depression detection by treating it as a classification or prediction problem [17], [19], [21], [22], [24], [28], [31], [33], [34], [46]–[51], [54], [55], [57]–[62]. Sample techniques include Support Vector Machines [17], [19], [21], [22], [28], [33], [34], [46]–[48], [50], [51], [54], [56], [58], [59], [61], [63], Logistic Regression [17], [21], [24], [34], [46], [59], [62], Naïve Bayes [21], [34], [46], [47], [57], [61]–[63], Decision Tree [21], [34], [46], [47], [50], [61], Random Forest [17], [28], [31], [46], [47], [55], [58], [59], linear regression [27], [35], [39], Adaboost [59], Perceptron [22],Support Vector Regression [55], k-nearest neighbor [49], [50], Maximum Entropy [63], log-linear [47], Passive Aggressive [22], linear discriminant analysis [62] and ensemble classifiers [50]. In addition, several studies applied regression analyses such as linear regression [27], [30], [35], [36], [39], Poisson regression [66], and binomial regression [20]. Furthermore, deep learning techniques such as CNN [28], [48], [71], RNN [48], [71], [72], Gated Recurrent Units (GRU) [56] and Multilayer Perceptron [59], [60] have demonstrated their efficacy in detecting the signals of depression.

In addition, a number of studies combined supervised learning with unsupervised learning techniques. For example, [57] used Multimodal Dictionary Learning Model, Wasserstein Dictionary Learning, Multiple Social Networking Learning, and Naïve Bayes and found that the Multimodal Dictionary Learning model outperformed other models.

Reinforcement learning techniques have also been used for detecting depression on social media. In particular, [71] showed that the proposed reinforcement method could perform significantly better than existing state-of-the-art techniques. In addition, [72] propose an emotion-based reinforcement attention network (ERAN) to detect depression. By using ERAN, depression can be detected more accurately by extracting deep emotional features, and obtaining accurate and complete information via sentence-level attention.

Building a supervised classification model requires determining the outputs as well as input features. We already summarized the input features in Table 2. The number of possible output values determines whether depression detection is treated as a binary classification (e.g., depressed vs. non-depressed) or a multi-class classification problem. All the previous studies built binary classification models except for one [28], and the latter classified depression into three levels of severity: low, medium, and high. Furthermore, among other methods, SVM seems to be the most commonly used ML algorithm because it works very well in a high dimensional space and is efficient when the number of features is greater than the sample size. From our observation, there is no single classifier that outperforms others, as the performance accuracy depends on different factors such as data type, size of sample, features, or nature of data (category, quantity) (see Table 3).

4.5.1 Evaluation of algorithms and performance reporting

Most studies using supervised techniques describe how they validated their models for depression detection, with k-fold cross validation being the most widely used method. There was a variation of k used in those papers, ranging from 5 [33], [37], [48], [54], [56]–[58], [62], [71],6 [21],10 [19], [24], [34], [39], [49]–[51], [59], [61], 20 [35] to a leave-one-out range [46], [47]. Another common evaluation technique is holding out data as a test set, and reports the performance on that data [24], [27], [35], [51], [55], [56], [61]. The size of the held-out dataset ranged from 10–20% to 30%.

Lastly, several papers reported performance in a way that could be used as a benchmark for other research. For this purpose, several metrics for classification models have been used, including accuracy, precision, recall, and F-measure [21], [22], [34], [47], [48], [58], [59], [61], specificity [31], [35], [37], sensitivity [35], [49], and area under the curve (AUC) [24], [33]–[35], [48], [61]. For regression models, common performance metrics have been used like adjusted R squared [36] and mean squared error [39].

Table 3
Classification Performances of Depression Detection Related literature.

ML Technique	Study	Performance			
		Accuracy	Precision	Recall	F-measure
SVM	[63]	91%	0.83	0.79	N/A
	[33]	73%	0.82	0.67	N/A
	[46]	89%	0.91	0.87	0.88
	[59]	90%	0.89	0.93	0.91
	[50]	71%	0.58	1	0.73
	[61]	82.5%	0.73	0.85	0.79
	[21]	82%	0.83	0.83	0.83
	[58]	90.77%	0.65	0.61	0.63
	[19]	70%	0.74	0.629	N/A
	[28]	N/A	0.87	0.84	0.85
	[34]	81.67%	0.83	0.82	0.82
	[54]	N/A	N/A	N/A	0.91
	[51]	61%	0.12	0.80	0.21
	[48]	73.4%	0.73	0.74	0.73
	[47]	90%	N/A	N/A	N/A
	[22]	80%	0.95	0.43	0.6
	[56]	N/A	N/A	N/A	0.513
	[17]	86.5%	N/A	N/A	0.853
LR	[59]	89%	0.89	0.92	0.89
	[21]	82%	0.86	0.82	0.84
	[24]	80%	0.81	0.81	0.8
	[46]	N/A	N/A	N/A	N/A
	[62]	90%	0.90	0.89	0.90
	[34]	N/A	N/A	N/A	N/A
	[17]	86.6%	N/A	N/A	0.849

ML Technique	Study	Performance			
		Accuracy	Precision	Recall	F-measure
DT	[50]	71%	0.58	0.98	0.73
	[61]	77.5%	0.65	0.59	0.619
	[21]	67%	0.67	0.68	0.75
	[46]	N/A	N/A	N/A	N/A
	[34]	N/A	N/A	N/A	N/A
	[47]	89%	N/A	N/A	N/A
NB	[61]	80%	0.65	0.8	0.72
	[21]	86%	0.81	0.82	0.81
	[46]	N/A	N/A	N/A	N/A
	[34]	N/A	N/A	N/A	N/A
	[63]	83%	0.88	0.82	N/A
	[62]	87%	0.89	0.87	0.87
	[47]	N/A	N/A	N/A	N/A
	[57]	73%	0.83	0.73	0.72
RF	[46]	92%	0.94	0.90	0.92
	[31]	70%	0.60	0.7	0.64
	[59]	85%	0.83	0.87	0.85
	[58]	92%	0.79	0.519	0.62
	[28]	N/A	0.86	0.66	0.74
	[47]	90%	N/A	N/A	N/A
	[55]	N/A	0.63	N/A	N/A
	[37]	N/A	0.86	0.52	0.651
	[17]	86.9%	N/A	N/A	0.863
Linear regression	[35]	AUC: 0.73	N/A	N/A	N/A
	[27]	Pearson correlation: 0.386	N/A	N/A	N/A

ML Technique	Study	Performance			
		Accuracy	Precision	Recall	F-measure
	[39]	Pearson correlation: 0.619	N/A	N/A	N/A
Ada boost	[59]	79%	0.72	0.93	0.81
KNN	[50]	60%	0.59	0.64	0.61
	[49]	N/A	0.59	0.88	0.71
Support Vector Regression	[55]	N/A	0.74	N/A	N/A
Perceptron	[22]	89%	0.83	0.87	0.85
Passive Aggressive	[22]	93%	0.9	0.89	0.9
Maximum Entropy	[63]	80%	0.84	0.79	N/A
Log linear	[47]	91%	0.9	0.89	0.9
Linear discriminant analysis	[62]	90%	0.91	0.89	0.90
CNN	[48]	87.9%	0.87	0.87	0.869
	[28]	91%	0.91	0.72	0.8
	[71]	78%	0.78	0.78	0.78
RNN BiLSTM	[48]	80.5%	0.8	0.83	0.8
	[72]	81%	0.82	0.82	0.82
CNN + Reinforcement Language	[71]	87%	0.87	0.87	0.87
RNN + Reinforcement Language	[71]	87%	0.87	0.87	0.87
GRU	[56]	N/A	N/A	N/A	0.33
MLP	[59]	91%	0.90	0.92	0.93
	[60]	N/A	0.32	0.62	0.42
Emotion-based Reinforcement Attention Network (ERAN)	[72]	90.6%	0.91	0.89	0.90
Ensemble	[50]	64%	0.58	0.96	0.72

5 Research Challenges And Opportunities

Based on our proposed multi-dimensional classification of the literature and discussion of the previous findings, we discuss research challenges and potential research opportunities in detecting depression on social media.

5.1 Feature fusion

Different features indicative of depression (see Table 2) can be complementary. Thus, feature fusion has the potential to not only improve the performance in detecting depression but also gain a more comprehensive understanding of the behavior of depressed users on social media. Many studies have hinted at the potential benefits of feature fusion [20], [26], [28], [30], [31], [34], [39], [44], [47], [49], [51], [52], [56], [58]. For instance, the inclusion of emojis, pictures, and gifs within tweets may give a more accurate perception of user behavior [28]. Integrating visual and textual features might outperform using either modality alone [31], [39].

5.2 Generalization

Previous studies typically limited their participants to a specific population with respect to nationality, language spoken, academic institution, specific platform, ethnicity, and age, centered around specific hashtags or online communities [20], [23], [33], [34], [38], [52], [56], [66], [70], used the data of a small number of users or a non-large dataset [24], [25], [35], [36], [55], [56], [70], had limited inclusion of demographic and socioeconomic information [24], [25], [30], [45], [51], [52] or collected data within a limited time period [18]. Ideally, they would need to validate their findings with a more general or diverse population.

5.3 Diagnosis verification

Performing a predictive analysis based solely on social media data without the cooperation from clinical psychologists has its inherent limitations [26]. A natural follow-up question is how to integrate social media data with medical data in patient medical records from sources, such as hospitals and mental health organizations, to improve the quality of data analysis. Having a database that combines social media data with patient-level data would potentially be an asset for caregivers/ clinical psychologists to assess the depression severity level, and predict unanticipated changes in health status and behavior [73]. In addition, most of the publications had limited theoretical and clinical grounding of the term “depression” [31]. Another ambiguity arises when researchers identify positive depressive samples using depression-related search terms. Even though a post may contain the key terms, it remains uncertain whether the user who generated the post was actually suffering from depression or whether the user might have different interests in this topic. These cases cast doubts on the credibility and validity of verifying the mental health status. The discussion opens the door to better capturing positive and negative samples of mental illness. Validation of the detection results should be based on clinical and/or theoretical principles. This can be accomplished in several ways. For instance, a research method can be grounded in the related literature from domains such as psychology and medicine. To this end, screening questionnaires (e.g., BDI) are highly recommended to determine the presence of depression. In addition, collaboration with clinical psychologists would bring expertise for gaining a deep understanding of the social media content of depressed users.

5.4 Absence of causes

Multiple psychological studies have emphasized the importance of identifying the cause(s) of depression in order to determine treatment compliance and the best strategy to cope with the disorder [74], [75]. Such factors may include genetics, personality, a traumatic event, social, biological, economic, environmental, religious, and so on [75], [76]. Therefore, identifying the causes of depression through the lens of social media data is a significant step forward.

5.5 Motivation of disclosing

Although various social media platforms have been used to disclose self-expressions of depression, it remains unclear what motivates users to share their self-expressions on social media platforms. For instance, does the depressed user intend to seek support, vent negative feelings (catharsis), avoid feeling lonely, or cope with his/her condition on social media? There have been some related speculations regarding the motivation for disclosing [26], [31], [32], [59]; however, theories remain lacking when it comes to explaining users' platform choice behavior and providing guidance on selecting social media platforms for collecting data from users who are experiencing depression.

5.6 Integration with offline behavior

Some users might display different behavior or personality online from an offline environment. Social media data alone lacks knowledge regarding the behavioral, cultural and social status of the user in "the offline world" [18], [23], [33], [66]. Thus, integrating online social media and offline behaviors provides a perfect opportunity to search, study, and find valuable insights into depression conditions.

5.7 Privacy setting and ethical concerns

The social media data that is used to support depression detection research is typically obtained from the public profiles and user posts. Given that social media platforms vary in privacy settings [20], [24], [66], it remains unclear whether the privacy setting might play a role in user disclosure of information about their mental status. Ethics and privacy are ongoing concerns and might arise with analyzing social media data, particularly when the data is considered sensitive. Very few users recognize that their mental health information could be extracted from their online activities [11]. It is crucial to ensure that participants clearly understand the nature of their participation and the types of social media data that would be gathered [77].

5.8 Cultural differences

Culture impacts how depression is expressed and shapes thoughts about mental health and illness, as well as the treatment of mental illness [78]. It is advantageous to use social media to research and study people from many different cultures. An exploratory study found cross-cultural differences in depression expression in online data, especially with regard to cognition, expressed emotions, and functioning [79].

One potential area for future study is how people who experience depression from different cultures behave when they are online and whether there are cross-cultural differences in terms of textual, visual, user behavior, and other features. To this end, we reviewed a few studies that used data from non-English speaking users (e.g., Japanese, Chinese, Arabic). Our preliminary observations suggest some differences, particularly in the user posting behavior. For example, for Japanese speakers, post time and the number of mentions, followers, and followees did not reveal significant differences between the depressed group and the non-depressed group [80]. However, English speakers have consistently demonstrated the difference [19].

5.9 Early detection of depression

Early risk detection is an emerging and promising research area that has significant implications for individuals with mental disorders. In view of the potentially serious consequences of depression, it is crucial to identify early signs of depression. Human interaction with social media effectively contributes to constructing the "digital phenotype" and identifying early signals of different condition [81]. This may be achieved through analyzing and tracking user behavior in term of user engagement, language style, time span, etc.

5.10 Advanced machine learning techniques

The performance of depression detection models depends not only on the input features but on machine learning techniques as well. Deep learning has demonstrated its superior performance to conventional counterparts in computer vision and healthcare [82], among many other real-world applications that involve text and/or image data. In particular, deep learning has generated promising results in classifying depression-related information [48]. In addition, using machine learning techniques on large data sets is incredibly beneficial for detecting depression [82]. However, advanced machine learning techniques remain underexplored in depression detection research.

5.11 The role of Conversational Agent

A Conversational Agent (CA) is a dialog system based on NLP that is able to respond to a user query by using human language (e.g., Alexa, Siri). It is valuable to leverage CA in the healthcare field, and mainly, there is a room for improvement with regard to the responses of CA to matters related to the mental health [83]. CA provides an excellent opportunity to improve medical care service referrals, particularly for mental health. In the U. S., over 250 million adults own a smartphone [84], and around 21% of American adults have used a chatbot for shopping, banking services, or information sharing [85]. Thus, it would be promising to explore the technology of CA and its responses to cope with questions and statements about depression and other mental disorders in future research.

5.12 Intervention of depression

Despite the stream of research on social media-based depression detection, very few studies have discussed the role of social media in depression intervention and investigated how commenters, community members, and moderators can effectively deal with users who post about depression. Depression intervention is essential as it aims to prevent a relapse, which might lead to multiple severe consequences, such as drug and alcohol excessive use [86], incapacity for work [87], and even the risk of suicidal attempts [88]. Different kinds of interventions can be adopted to mitigate the impact of the disorder, such as peer support, computerized cognitive behavioral therapy, physical activity, and counseling [89]. Having an appropriate level of social support would effectively reduce the risk of depression [90]. The Internet and computer-based technologies actively contributed to the intervention of depression, especially for college students [91]. For instance, mobile apps, virtual reality systems, and websites have shown to be practical and feasible for depression intervention and mental well-being enhancement across college students [92].

6 Conclusion

Depression has become a significant issue for individuals and society at large. This mental disorder may lead to severe consequences, affecting all aspects of a person's life. In this paper, we explore depression detection on social media from multiple dimensions. The paper categorizes social media features that signal depression into different types such as textual, visual, user behavior, demographic, temporal, and spatial features. Among the various machine learning techniques that have been used to build depression detection models, SVM is most commonly used. Despite the various types of depressive disorders, existing studies have focused on general depression. Among the various social media platforms, Facebook, Twitter, Reddit, and Instagram are the most common platforms for data collection. In view of the challenges and limitations of depression detection on social media, this paper presents a number of future research opportunities, such as fusing multimodal features, improving the generalizability of depression models, depression validation, integrating both online and offline behavior, identifying the causes, and addressing ethical concerns. Finally, we not only highlight the importance but also suggest ways for researchers across multiple disciplines to collaborate in addressing the depression problem.

Declarations

Statements and Declarations:

Competing Interests: The authors declare no competing interests.

Author contributions: A.A. designed and conducted the survey, analyzed the data, designed the framework, and wrote the initial draft of the manuscript. L.Z. provided critical feedback on the manuscript, contributed to the data interpretation, drafted, revised, and approved the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding: No funding was received for this study.

References

1. U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health., “Depression (NIH Publication No. 21-MH-8079).” Bethesda, MD: U.S. Government Printing Office., 2021.
2. M. Marcus, M. T. Yasamy, M. van van Ommeren, D. Chisholm, and S. Saxena, “Depression: A Global Public Health Concern: (517532013-004).” American Psychological Association, 2012. doi: 10.1037/e517532013-004.
3. “CDC Centers for Disease Control and Prevention: Depression - Mental Illness - Mental Health Basics - Mental Health.” 2016.
4. “The Australian Psychological Society (APS).” [Online]. Available: https://www.psychology.org.au/for-the-public/Psychology-topics/Depression
5. S.-R. Khalsa, K. S. McCarthy, B. A. Sharpless, M. S. Barrett, and J. P. Barber, “Beliefs about the causes of depression and treatment preferences,” *J. Clin. Psychol.*, vol. 67, no. 6, pp. 539–549, Jun. 2011, doi: 10.1002/jclp.20785.
6. N. Salari *et al.*, “Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis,” Global Health, vol. 16, no. 1, p. 57, Dec. 2020, doi: 10.1186/s12992-020-00589-w.
7. S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, “The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users,” *IJERPH*, vol. 17, no. 6, p. 2032, Mar. 2020, doi: 10.3390/ijerph17062032.
8. A. Halfin, “Depression: the benefits of early and appropriate treatment,” *Am J Manag Care*, vol. 13, no. 4 Suppl, pp. S92-97, Nov. 2007.
9. G. Schomerus and M. C. Angermeyer, “Stigma and its impact on help-seeking for mental disorders: what do we know?,” *Epidemiol Psichiatr Soc*, vol. 17, no. 1, pp. 31–37, Mar. 2008, doi: 10.1017/S1121189X00002669.
10. S. DEY, I. SARKAR, S. CHAKRABORTY, and S. ROY, “DEPRESSION DETECTION USING INTELLIGENT ALGORITHMS FROM SOCIAL MEDIA CONTEXT - STATE OF THE ART, TRENDS AND FUTURE ROADMAP,” *jxu*, vol. 14, no. 8, Aug. 2020, doi: 10.37896/jxu14.8/007.
11. S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, “Detecting depression and mental illness on social media: an integrative review,” *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, Dec. 2017, doi: 10.1016/j.cobeha.2017.07.005.
12. J. Murrieta, C. C. Frye, L. Sun, L. G. Ly, C. S. Cochancela, and E. V. Eikey, “#Depression: Findings from a Literature Review of 10 Years of Social Media and Depression Research,” in *Transforming Digital Worlds*, G. Chowdhury, J. McLeod, V. Gillet, and P. Willett, Eds., in Lecture Notes in Computer Science, vol. 10766. Cham: Springer International Publishing, 2018, pp. 47–56. doi: 10.1007/978-3-319-78105-1_6.
13. R. Salas-Zárate, G. Alor-Hernández, M. del P. Salas-Zárate, M. A. Paredes-Valverde, M. Bustos-López, and J. L. Sánchez-Cervantes, “Detecting Depression Signs on Social Media: A Systematic Literature Review,” *Healthcare*, vol. 10, no. 2, p. 291, Feb. 2022, doi: 10.3390/healthcare10020291.

14. A. Zafar and S. Chitnis, “Survey of Depression Detection using Social Networking Sites via Data Mining,” in 2020 *10th International Conference on Cloud Computing, Data Science &* Engineering *(Confluence)*, Noida, India: IEEE, Jan. 2020, pp. 88–93. doi: 10.1109/Confluence47617.2020.9058189.
15. A. H. Yazdavar *et al.*, “Multimodal mental health analysis in social media,” *PLoS ONE*, vol. 15, no. 4, p. e0226248, Apr. 2020, doi: 10.1371/journal.pone.0226248.
16. J. Willers, “Methods for extracting data from the Internet,” Master of Science, Iowa State University, Digital Repository, Ames, 2017. doi: 10.31274/etd-180810-5256.
17. I. Sekulić, M. Gjurković, and J. Šnajder, “Not Just Depressed: Bipolar Disorder Prediction on Reddit.” arXiv, Mar. 27, 2019. Accessed: Sep. 27, 2022. [Online]. Available: http://arxiv.org/abs/1811.04655
18. P. A. Cavazos-Rehg *et al.*, “A content analysis of depression-related tweets,” Computers in Human Behavior, vol. 54, pp. 351–357, Jan. 2016, doi: 10.1016/j.chb.2015.08.023.
19. M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, “Predicting Depression via Social Media,” AAAI, Jul. 2013. [Online]. Available: https://www.microsoft.com/en-us/research/publication/predicting-depression-via-social-media/
20. M. A. Moreno *et al.*, “Feeling bad on Facebook: depression disclosures by college students on a social networking site,” *Depress. Anxiety*, vol. 28, no. 6, pp. 447–455, Jun. 2011, doi: 10.1002/da.20805.
21. M. Nadeem, “Identifying Depression on Twitter,” *arXiv:1607.07384 [cs, stat]*, Jul. 2016, Accessed: Jul. 08, 2021. [Online]. Available: http://arxiv.org/abs/1607.07384
22. A. Trifan, D. Semeraro, J. Drake, R. Bukowski, and J. L. Oliveira, “Social Media Mining for Postpartum Depression Prediction,” *Stud Health Technol Inform*, vol. 270, pp. 1391–1392, Jun. 2020, doi: 10.3233/SHTI200457.
23. M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, “Characterizing and predicting postpartum depression from shared facebook data,” in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, Baltimore Maryland USA: ACM, Feb. 2014, pp. 626–638. doi: 10.1145/2531602.2531675.
24. M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, “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*, San Jose California USA: ACM, May 2016, pp. 2098–2110. doi: 10.1145/2858036.2858207.
25. P. A. Cavazos-Rehg *et al.*, “An Analysis of Depression, Self-Harm, and Suicidal Ideation Content on Tumblr,” Crisis, vol. 38, no. 1, pp. 44–52, Jan. 2017, doi: 10.1027/0227-5910/a000409.
26. N. L. Tefera and L. Zhou, “A Scorecard Method for Detecting Depression in Social Media Users,” presented at the Hawaii International Conference on System Sciences, 2018. doi: 10.24251/HICSS.2018.071.
27. H. A. Schwartz *et al.*, “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, Baltimore, Maryland, USA: Association for Computational Linguistics, 2014, pp. 118–125. doi: 10.3115/v1/W14-3214.

28. R. U. Mustafa, N. Ashraf, F. S. Ahmed, J. Ferzund, B. Shahzad, and A. Gelbukh, “A Multiclass Depression Detection in Social Media Based on Sentiment Analysis,” in *17th International Conference on Information Technology–New Generations (ITNG 2020)*, S. Latifi, Ed., in Advances in Intelligent Systems and Computing, vol. 1134. Cham: Springer International Publishing, 2020, pp. 659–662. doi: 10.1007/978-3-030-43020-7_89.
29. Hootsuite & We Are Social, “Digital 2021 Global Digital Overview.” 2021. [Online]. Available: https://datareportal.com/reports/digital-2021-global-overview-report.
30. S. Park, I. Kim, S. W. Lee, J. Yoo, B. Jeong, and M. Cha, “Manifestation of Depression and Loneliness on Social Networks: A Case Study of Young Adults on Facebook,” in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, Vancouver BC Canada: ACM, Feb. 2015, pp. 557–570. doi: 10.1145/2675133.2675139.
31. A. G. Reece and C. M. Danforth, “Instagram photos reveal predictive markers of depression,” *arXiv:1608.03282 [physics]*, Aug. 2016, Accessed: Jul. 07, 2021. [Online]. Available: http://arxiv.org/abs/1608.03282
32. K. Lup, L. Trub, and L. Rosenthal, “Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed,” *Cyberpsychology, Behavior, and Social Networking*, vol. 18, no. 5, pp. 247–252, May 2015, doi: 10.1089/cyber.2014.0560.
33. M. De Choudhury, S. Counts, and E. Horvitz, “Social media as a measurement tool of depression in populations,” in *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci ’13*, Paris, France: ACM Press, 2013, pp. 47–56. doi: 10.1145/2464464.2464480.
34. A. Wongkoblap, M. A. Vadillo, and V. Curcin, “A Multilevel Predictive Model for Detecting Social Network Users with Depression,” in 2018 *IEEE International Conference on Healthcare* Informatics *(ICHI)*, New York, NY: IEEE, Jun. 2018, pp. 130–135. doi: 10.1109/ICHI.2018.00022.
35. B. J. Ricard, L. A. Marsch, B. Crosier, and S. Hassanpour, “Exploring the Utility of Community-Generated Social Media Content for Detecting Depression: An Analytical Study on Instagram,” *J Med Internet Res*, vol. 20, no. 12, p. e11817, Dec. 2018, doi: 10.2196/11817.
36. M. Park, C. Cha, and M. Cha, “Depressive Moods of Users Portrayed in Twitter,” in *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)*, 2012.
37. A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with Twitter data,” *Sci Rep*, vol. 7, no. 1, p. 13006, Dec. 2017, doi: 10.1038/s41598-017-12961-9.
38. J. L. Carey *et al.*, “SoMe and Self Harm: The Use of Social Media in Depressed and Suicidal Youth,” presented at the Hawaii International Conference on System Sciences, 2018. doi: 10.24251/HICSS.2018.420.

39. S. C. Guntuku, D. Preotiuc-Pietro, J. C. Eichstaedt, and L. H. Ungar, "What Twitter Profile and Posted Images Reveal About Depression and Anxiety," *arXiv:1904.02670 [cs]*, Apr. 2019, Accessed: Jul. 09, 2021. [Online]. Available: http://arxiv.org/abs/1904.02670
40. K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9: Validity of a brief depression severity measure," J Gen Intern Med, vol. 16, no. 9, pp. 606–613, Sep. 2001, doi: 10.1046/j.1525-1497.2001.016009606.x.
41. L. S. Radloff, "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," Applied Psychological Measurement, vol. 1, no. 3, Jun. 1977, doi: 10.1177/014662167700100306.
42. A.Beck and R. Steer, *Manual for the Beck Hopelessness Scale. The Psychological Corporation*. 1988.
43. L. Goldberg, *A broad-bandwidth, public domain, personality inventory measuring the lower level facets of several five-factor models*. Tilburg: Tilburg Univ. Press, 1999.
44. N. Andalibi, P. Ozturk, and A. Forte, "Depression-related Imagery on Instagram," in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, Vancouver BC Canada: ACM, Feb. 2015, pp. 231–234. doi: 10.1145/2685553.2699014.
45. N. Armin, "Understanding Depression During the COVID-19 Pandemic Through Social Media Data," The University of Mississippi, 2021.
46. X. Chen, M. D. Sykora, T. W. Jackson, and S. Elayan, "What about Mood Swings: Identifying Depression on Twitter with Temporal Measures of Emotions," in *Companion of the The Web Conference* 2018 *on The Web Conference 2018 - WWW '18*, Lyon, France: ACM Press, 2018, pp. 1653–1660. doi: 10.1145/3184558.3191624.
47. X. Chen, M. Sykora, T. Jackson, S. Elayan, and F. Munir, "Tweeting Your Mental Health: an Exploration of Different Classifiers and Features with Emotional Signals in Identifying Mental Health Conditions," presented at the Hawaii International Conference on System Sciences, 2018. doi: 10.24251/HICSS.2018.421.
48. A. Husseini Orabi, P. Buddhitha, M. Husseini Orabi, and D. Inkpen, "Deep Learning for Depression Detection of Twitter Users," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, New Orleans, LA: Association for Computational Linguistics, 2018, pp. 88–97. doi: 10.18653/v1/W18-0609.
49. M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. ulhaq, "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique," in 2018 *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Rajshahi: IEEE, Feb. 2018, pp. 1–4. doi: 10.1109/IC4ME2.2018.8465641.
50. Md. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," Health Inf Sci Syst, vol. 6, no. 1, p. 8, Dec. 2018, doi: 10.1007/s13755-018-0046-0.
51. Z. Jamil, D. Inkpen, P. Buddhitha, and K. White, "Monitoring Tweets for Depression to Detect At-risk Users," in *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology*

–- *From Linguistic Signal to Clinical Reality*, Vancouver, BC: Association for Computational Linguistics, 2017, pp. 32–40. doi: 10.18653/v1/W17-3104.

52. E. M. Lachmar, A. K. Wittenborn, K. W. Bogen, and H. L. McCauley, "#MyDepressionLooksLike: Examining Public Discourse About Depression on Twitter," *JMIR Ment Health*, vol. 4, no. 4, p. e43, Oct. 2017, doi: 10.2196/mental.8141.
53. M. Husain, "Social Media Analytics to Predict Depression Level in the Users," 2019, doi: 10.4018/978-1-5225-8567-1.
54. I. Pirina and Ç. Çöltekin, "Identifying Depression on Reddit: The Effect of Training Data," in *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task*, Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 9–12. doi: 10.18653/v1/W18-5903.
55. P. Resnik, W. Armstrong, L. Claudino, T. Nguyen, V.-A. Nguyen, and J. Boyd-Graber, "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*, Denver, Colorado: Association for Computational Linguistics, 2015, pp. 99–107. doi: 10.3115/v1/W15-1212.
56. F. Sadeque, D. Xu, and S. Bethard, "Measuring the Latency of Depression Detection in Social Media," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, Marina Del Rey CA USA: ACM, Feb. 2018, pp. 495–503. doi: 10.1145/3159652.3159725.
57. G. Shen *et al.*, "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 3838–3844. doi: 10.24963/ijcai.2017/536.
58. M. Stankevich, V. Isakov, D. Devyatkin, and I. Smirnov, "Feature Engineering for Depression Detection in Social Media:," in *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, Funchal, Madeira, Portugal: SCITEPRESS - Science and Technology Publications, 2018, pp. 426–431. doi: 10.5220/0006598604260431.
59. M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," IEEE Access, vol. 7, pp. 44883–44893, 2019, doi: 10.1109/ACCESS.2019.2909180.
60. D. Maupomés and M. Meurs, "Using Topic Extraction on Social Media Content for the early detection of depression," in *CLEF (Working Notes)*, Sep. 2018. [Online]. Available: https://CEUR-WS.org
61. H. S. Alsagri and M. Ykhlef, "Machine Learning-Based Approach for Depression Detection in Twitter Using Content and Activity Features," *IEICE Trans. Inf. & Syst.*, vol. E103.D, no. 8, pp. 1825–1832, Aug. 2020, doi: 10.1587/transinf.2020EDP7023.
62. J. Zhou, H. Zogan, S. Yang, S. Jameel, G. Xu, and F. Chen, "Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia," *arXiv:2007.02325 [cs]*, Jul. 2020, Accessed: Jul. 09, 2021. [Online]. Available: http://arxiv.org/abs/2007.02325

63. A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, “Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression,” in 2017 *International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju: IEEE, Oct. 2017, pp. 138–140. doi: 10.1109/ICTC.2017.8190959.
64. E. Fast, B. Chen, and M. Bernstein, “Empath: Understanding Topic Signals in Large-Scale Text,” in *Proceedings of the* 2016 *CHI Conference on Human Factors in Computing Systems*, May 2016, pp. 4647–4657. doi: 10.1145/2858036.2858535.
65. F. Å. Nielsen, “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs,” *arXiv:1103.2903 [cs]*, Mar. 2011, Accessed: Jul. 09, 2021. [Online]. Available: http://arxiv.org/abs/1103.2903
66. N. Andalibi, P. Ozturk, and A. Forte, “Sensitive Self-disclosures, Responses, and Social Support on Instagram: The Case of #Depression,” in *Proceedings of the* 2017 *ACM Conference on Computer Supported Cooperative Work and Social Computing*, Portland Oregon USA: ACM, Feb. 2017, pp. 1485–1500. doi: 10.1145/2998181.2998243.
67. J. Bigun, *Vision with Direction: a Systematic Introduction to Image Processing and Computer Vision*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2006.
68. L. Lustberg and C. F. Reynolds, “Depression and insomnia: questions of cause and effect,” *Sleep Medicine Reviews*, vol. 4, no. 3, pp. 253–262, Jun. 2000, doi: 10.1053/smrv.1999.0075.
69. J. Kim and H. Kim, “Demographic and Environmental Factors Associated with Mental Health: A Cross-Sectional Study,” IJERPH, vol. 14, no. 4, p. 431, Apr. 2017, doi: 10.3390/ijerph14040431.
70. S. Cash, L. M. Schwab-Reese, E. Zipfel, M. Wilt, and M. Moreno, “What College Students Post About Depression on Facebook and the Support They Perceive: Content Analysis,” *JMIR Form Res*, vol. 4, no. 7, p. e13650, Jul. 2020, doi: 10.2196/13650.
71. T. Gui, Q. Zhang, L. Zhu, X. Zhou, M. Peng, and X. Huang, “Depression detection on social media with reinforcement learning,” in *China National Conference on Chinese Computational Linguistics*, Springer, 2019, pp. 613–624.
72. B. Cui, J. Wang, H. Lin, Y. Zhang, L. Yang, and B. Xu, “Emotion-Based Reinforcement Attention Network for Depression Detection on Social Media: Algorithm Development and Validation,” JMIR Med Inform, vol. 10, no. 8, p. e37818, Aug. 2022, doi: 10.2196/37818.
73. K. A. Padrez *et al.*, “Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department,” BMJ Qual Saf, vol. 25, no. 6, pp. 414–423, Jun. 2016, doi: 10.1136/bmjqs-2015-004489.
74. J. Srinivasan, N. L. Cohen, and S. V. Parikh, “Patient Attitudes regarding Causes of Depression: Implications for Psychoeducation,” *Can J Psychiatry*, vol. 48, no. 7, pp. 493–495, Aug. 2003, doi: 10.1177/070674370304800711.
75. M. Hansson, J. Chotai, and O. Bodlund, “Patients’ beliefs about the cause of their depression,” *Journal of Affective Disorders*, vol. 124, no. 1–2, pp. 54–59, Jul. 2010, doi: 10.1016/j.jad.2009.10.032.

76. M. E. Addis, P. Truax, and N. S. Jacobson, “Why do people think they are depressed?: The Reasons For Depression Questionnaire.,” *Psychotherapy: Theory, Research, Practice, Training*, vol. 32, no. 3, pp. 476–483, 1995, doi: 10.1037/0033-3204.32.3.476.
77. B. Inkster, D. Stillwell, M. Kosinski, and P. Jones, “A decade into Facebook: where is psychiatry in the digital age?,” The Lancet Psychiatry, vol. 3, no. 11, pp. 1087–1090, Nov. 2016, doi: 10.1016/S2215-0366(16)30041-4.
78. Y. E. Chentsova-Dutton and J. L. Tsai, “Understanding depression across cultures,” in *I. H. Gotlib & C. L. Hammen (Eds.), Handbook of depression*, 2nd ed.Guilford Press, 2009, pp. 363–385.
79. K. Loveys, J. Torrez, A. Fine, G. Moriarty, and G. Coppersmith, “Cross-cultural differences in language markers of depression online,” in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, New Orleans, LA: Association for Computational Linguistics, 2018, pp. 78–87. doi: 10.18653/v1/W18-0608.
80. S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, “Recognizing Depression from Twitter Activity,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul Republic of Korea: ACM, Apr. 2015, pp. 3187–3196. doi: 10.1145/2702123.2702280.
81. S. H. Jain, B. W. Powers, J. B. Hawkins, and J. S. Brownstein, “The digital phenotype,” Nat Biotechnol, vol. 33, no. 5, pp. 462–463, May 2015, doi: 10.1038/nbt.3223.
82. H. Sanyal, S. Shukla, and R. Agrawal, “Study of Depression Detection using Deep Learning,” in 2021 *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA: IEEE, Jan. 2021, pp. 1–5. doi: 10.1109/ICCE50685.2021.9427624.
83. A. S. Miner, A. Milstein, S. Schueller, R. Hegde, C. Mangurian, and E. Linos, “Smartphone-Based Conversational Agents and Responses to Questions About Mental Health, Interpersonal Violence, and Physical Health,” JAMA Intern Med, vol. 176, no. 5, p. 619, May 2016, doi: 10.1001/jamainternmed.2016.0400.
84. Statista, “Number of smartphone users in the United States from 2018 to 2025.” 2021. [Online]. Available: https://www.sta¬tista.com/statistics/201182/forecast-of-smartphone-users-in-the-us/.
85. K. Del Valle, “Conversational commerce: A new opportunity for card payments. MasterCard.” Jan. 2018. [Online]. Available: https://newsroom.mastercard.com/documents/conversational-commerce-a-new-opportunity-for-card-payments.
86. B. Birmaher *et al.*, “Childhood and Adolescent Depression: A Review of the Past 10 Years. Part I,” *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 35, no. 11, pp. 1427–1439, Nov. 1996, doi: 10.1097/00004583-199611000-00011.
87. L. Judd, M. Paulus, K. Wells, and M. Rapaport, “Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population,” *AJP*, vol. 153, no. 11, pp. 1411–1417, Nov. 1996, doi: 10.1176/ajp.153.11.1411.
88. M. M. Weissman, “Depressed Adolescents Grown Up,” JAMA, vol. 281, no. 18, p. 1707, May 1999, doi: 10.1001/jama.281.18.1707.

89. National Collaborating Centre for Mental Health (Great Britain), Ed., *Depression in adults with a chronic physical health problem: treatment and management: National clinical practice guideline 91*. London: British Psychological Society and the Royal College of Psychiatrists, 2010.
90. F. Hao, G. Pang, Y. Wu, Z. Pi, L. Xia, and G. Min, “Providing Appropriate Social Support to Prevention of Depression for Highly Anxious Sufferers,” *IEEE Trans. Comput. Soc. Syst.*, vol. 6, no. 5, pp. 879–887, Oct. 2019, doi: 10.1109/TCSS.2019.2894144.
91. E. B. Davies, R. Morriss, and C. Glazebrook, “Computer-Delivered and Web-Based Interventions to Improve Depression, Anxiety, and Psychological Well-Being of University Students: A Systematic Review and Meta-Analysis,” J Med Internet Res, vol. 16, no. 5, p. e130, May 2014, doi: 10.2196/jmir.3142.
92. E. G. Lattie, E. C. Adkins, N. Winquist, C. Stiles-Shields, Q. E. Wafford, and A. K. Graham, “Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review,” J Med Internet Res, vol. 21, no. 7, p. e12869, Jul. 2019, doi: 10.2196/12869.

Tables

Table 1 is available in the Supplementary Files section.

Figures

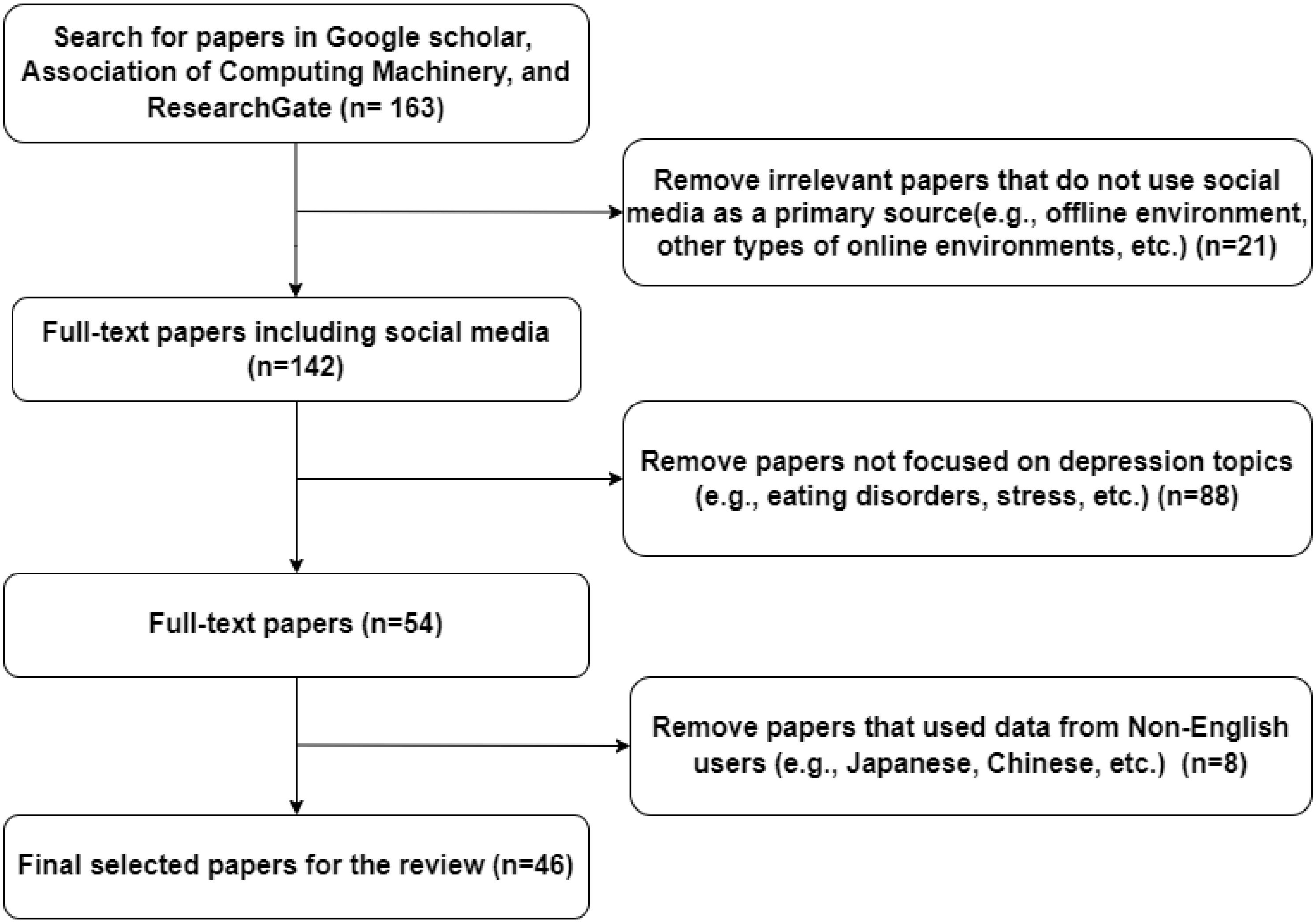

Figure 1

Paper selection process

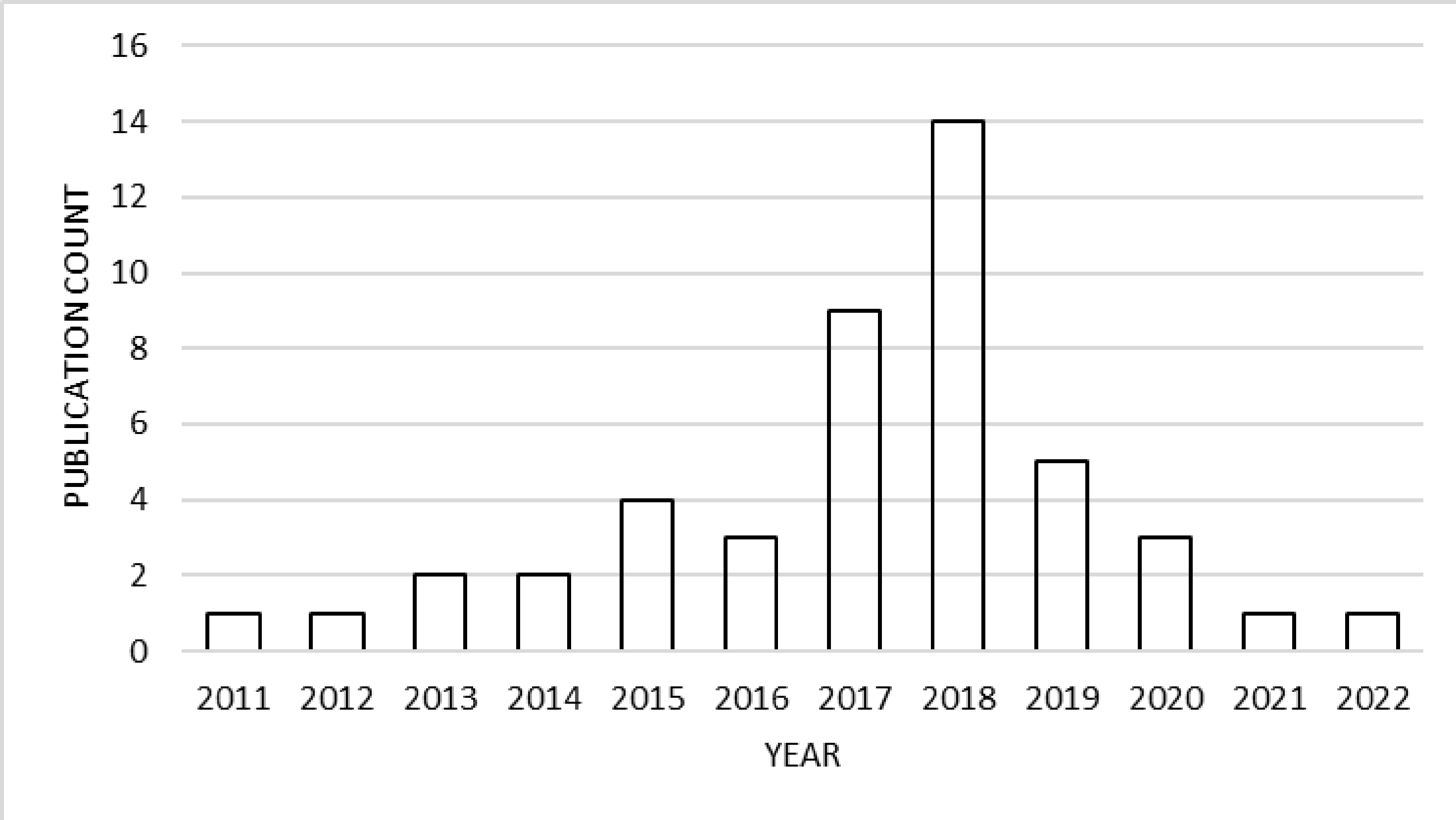

Figure 2

Publication count by year

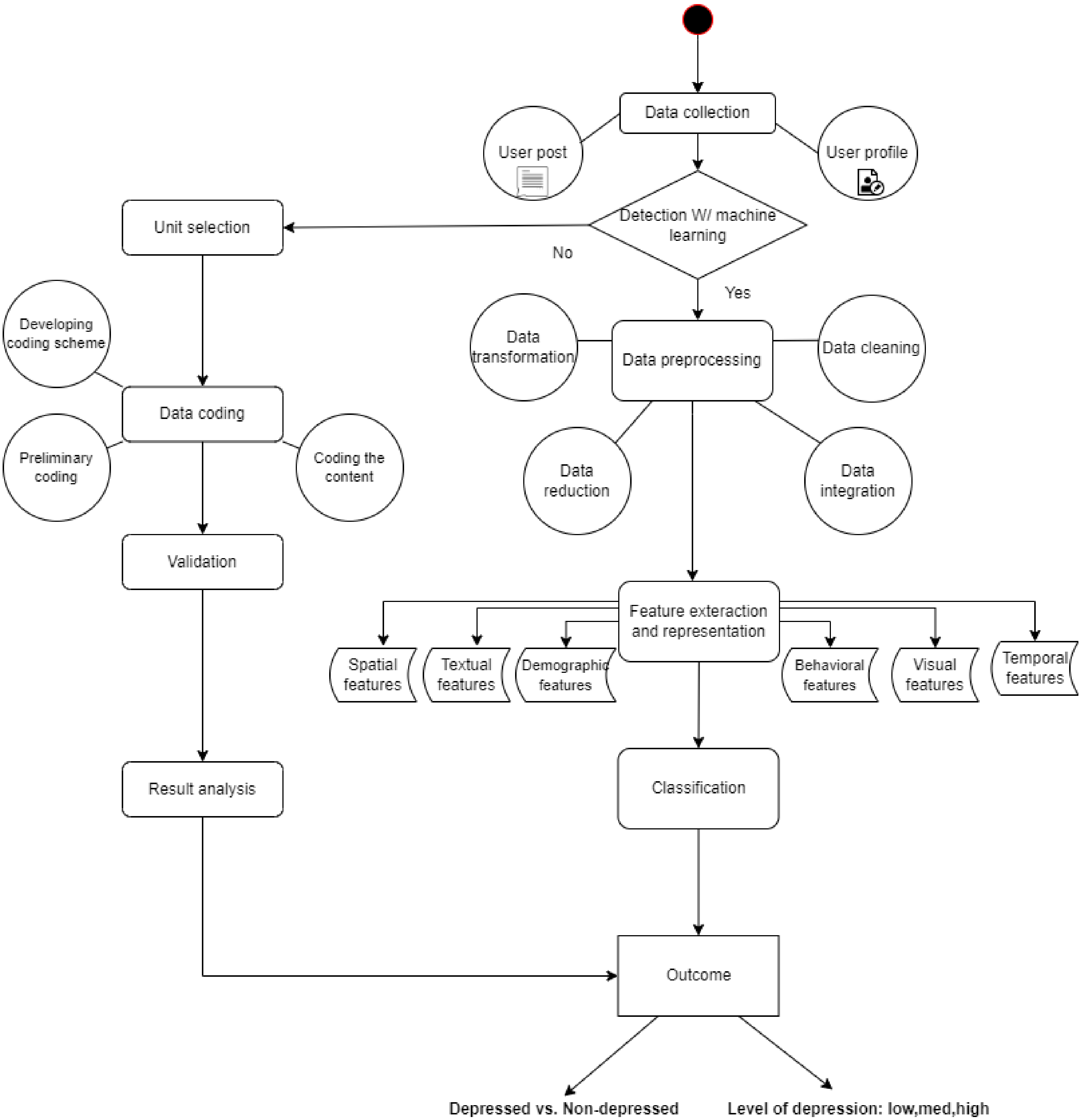

Figure 3

The depression detection process on social media.

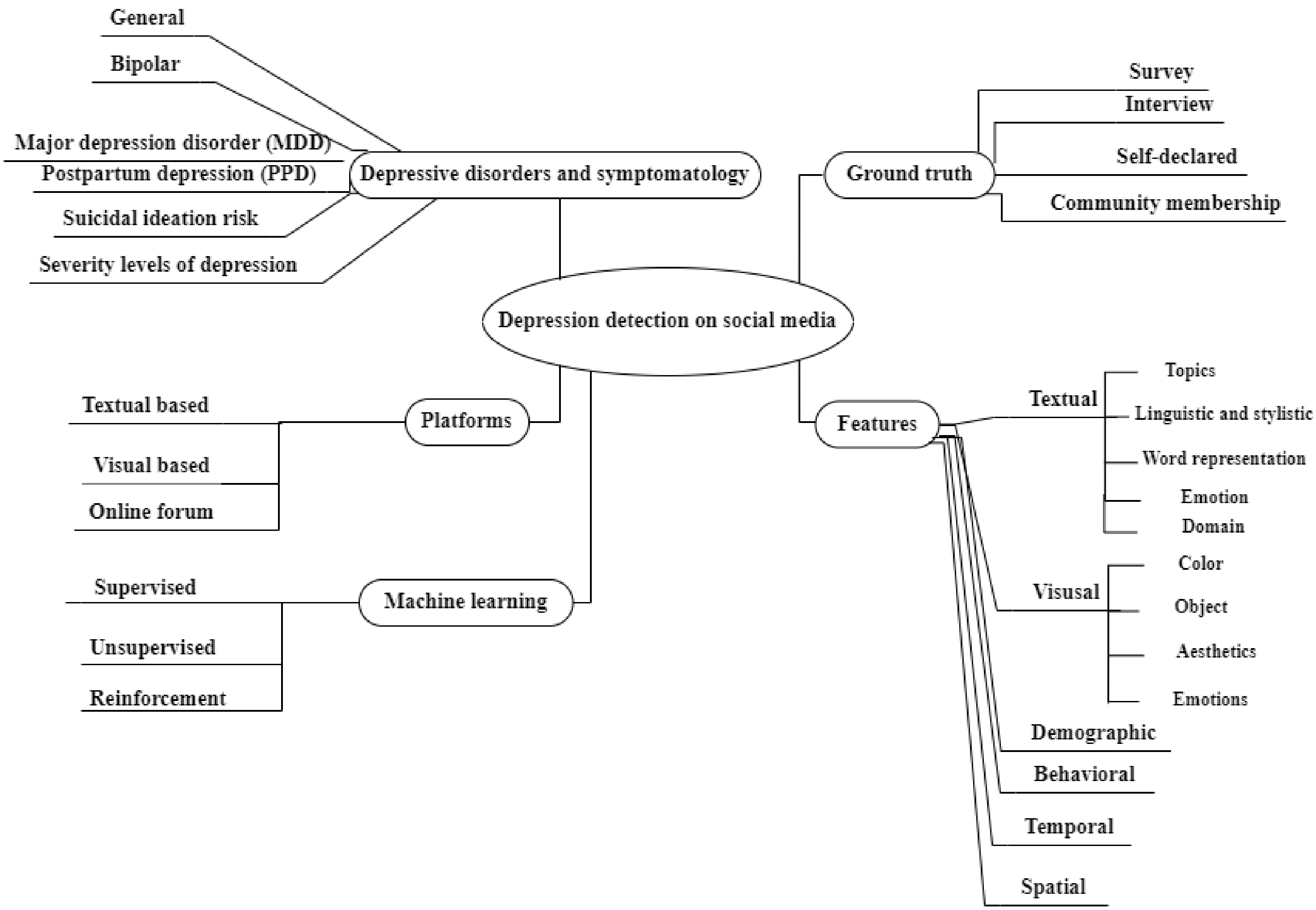

Figure 4

A multi-dimensional classification of studies on depression detection on social media

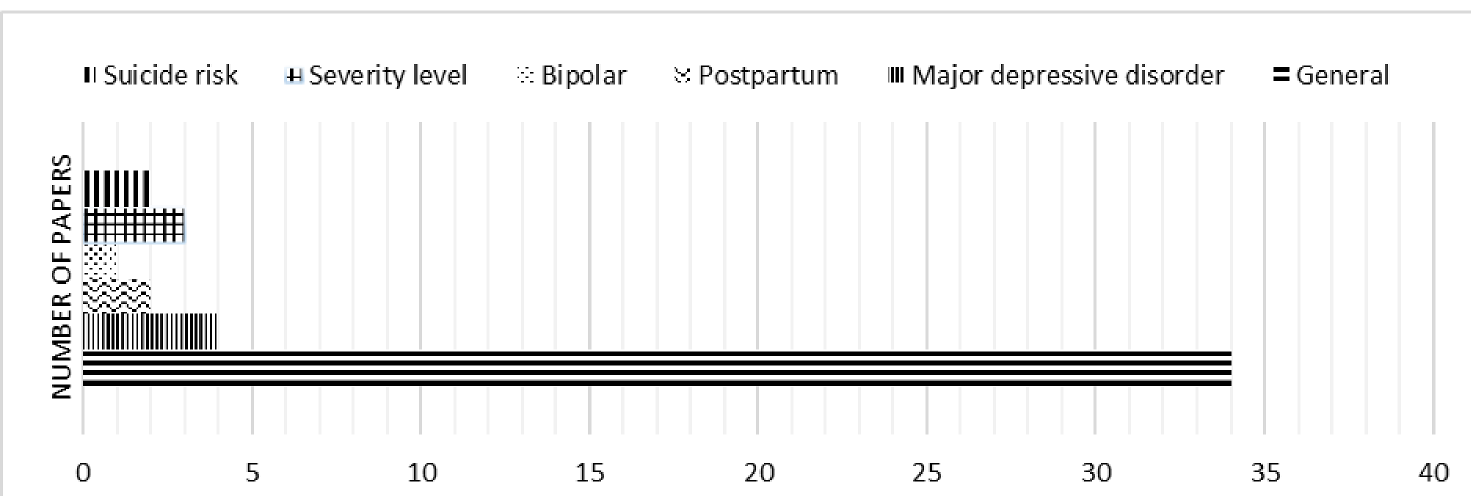

Figure 5

Types of depressive disorders

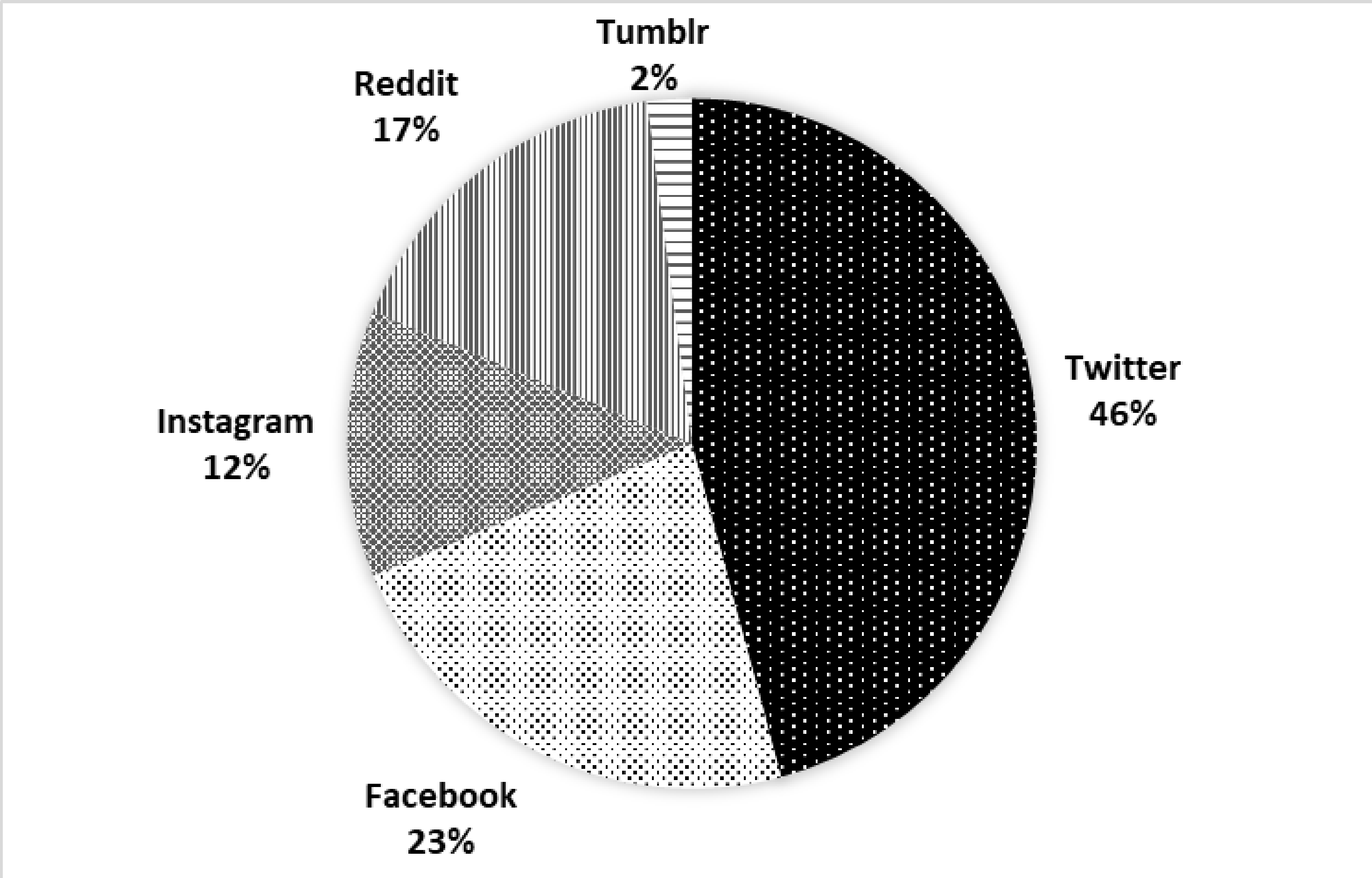

Figure 6

Social media platform as data sources

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- floatimage7.png