Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 3, 2025

A SYSTEMATIC SURVEY OF IDENTIFYING DEPRESSION ON SOCIAL NETWORKING SITES

Muhammad Bilal Qureshi[*1], Zoama Afaq[2], Hina Gul[3], Zaib Zafar[4]

[*1,2,4]*Department of Computer Science & IT, Superior University, 10 KM Lahore- Sargodha Rd, Sargodha, Punjab 40100, Pakistan*
[3]*Kinnaird College for Women University Lahore*

[*1]bilalshah1728@gmail.com, [2]zoama.afaq.sgd@superior.edu.pk, [3]hina.gul@gmail.com, [4]zaib.zafar@superior.edu.pk

DOI: https://doi.org/10.5281/zenodo.15074485

Keywords
Social networking sites, Sentiment analysis, Machine learning, Support vector machine, Annotation, Demographics

Article History
Received on 15 February 2025
Accepted on 15 March 2025
Published on 24 March 2025

Corresponding Author: *

Abstract
Social Networking Sites (SNS) are being used for online communication more frequently for many years. People and groups openly discuss their opinions and share their experiences, feelings, and thoughts including their mental health. One of the most discussed and predominant mental health disorders today is Depression. Depression has become the leading cause of disability and premature mortality somewhat due to a lack of effective methods for early detection. However, the diagnosis rate of mental illness including depression has been improved in the last few years but still many cases remain undetected. Significant work can be done in academic research by getting text data from posts and comments of people on social networking sites. SNS can potentially provide inexpensive early detection of individuals who might require a specialist's evaluation, based on their naturally occurring linguistic behavior. Symptoms regarding mental illnesses more specifically Major Depressive Disorder are observable on Social Network Sites and web forums. Automated methods are increasingly able to detect depression and other mental illnesses. Depressed users have been identified from normal users by patterns in their language and online activity. Many approaches of text analysis have been employed to identify and predict depressed users through such websites. In this paper, a systematic review of the literature to predict Major Depressive Disorder (MDD) has been conducted. Large-scale monitoring of social media and automated detection methods could identify depressed or otherwise at-risk individuals, which may help to increase the well-being of an identity and complement existing screening procedures.

INTRODUCTION
One of the defining phenomena of the present times, reshaping the world, is the worldwide use of social media, which comes in many forms, including blogs, forums, business networks, photo-sharing platforms, social gaming, microblogs, chat apps, and finally social networks. All around the world, people are getting connected through SNS to share their interests, information, and experiences. There will be around 3.02 billion monthly active social media users by 2021 [1]. One significant and increasing trend in the use of social media platforms is that people are getting open to discuss their issues including the discussion related to their health problems.

According to Hussain et al. [2], over the web, communication between parties is facilitated via Social media for sharing information, ideas, and career interests. Similarly, Hussain et al. thoroughly investigated that various diseases including stress, anxiety, diabetes, arthritis and cancer, Major Depressive Disorder (MDD) can also be identified among populations via their behavioral attributes in the content posted over SNS.

The most commonly used SNS are Twitter, Facebook, and Instagram. Status identification of depressed people has become easier by observing the posts, text content, status updates, photo sharing, user's social engagement, and other related activities [3]. Data collected by social media profiles is proved to be the most commonly used approach to predict the mental state of users. This data could be used to correlate social media usage and behavioral patterns with depression, stress, anxiety, and other mental illnesses. Various studies have been conducted in this regard, majorly focusing on depression identification. Even though many cases of depression are still underdiagnosed, with roughly half the cases detected by primary care physicians, and only 13–49% receiving minimally adequate treatment [4].

Various methods have been used to identify mentally sick people such as self-disclosure, online membership forums, questionnaires, and surveys. Through their online activity and language patterns, they can be easily distinguished from healthy people [5]. However, according to Wegrzyn-Wolska et al. [6], there exist challenges for social network use and text mining in e-health care applications and medicine. But in the meanwhile, if early detection methods could be devised then specific individuals could be directed for in-depth assessment and could be targeted for further support and treatment.

Besides the selection of features, machine learning techniques, classification techniques, natural language processing, and data analysis approaches are used to serve this purpose. Studies to date have either examined how the use of SNS correlates with Major Depressive Disorder in users [7] or attempted to detect depression through analysis of the content created by users. This review focuses on the latter: studies aimed at predicting depression using SNS.

Research in this field was started in 2012. The earlier studies of similar nature focused on measuring the behavioral attributes relating to social activity, emotion, language and linguistic styles, ego network, and mentions of antidepressant medications. Initially, researchers used these cues to perform sentiment analysis on the language used by people suffering from depression and to build statistical classifiers that provide estimates of the risk of depression, before the reported onset. The research further progressed to develop probabilistic models trained on a corpus that is generated from social media posts to determine if these posts could indicate depression. Over the past, one and half year's research interests shifted towards building machine learning models based on messages on a social platform for the early detection of depression. Most recently deep learning models are being constructed to generate predictive systems for depression.

It is no doubt just a beginning phase to examine the ways SNS can help us to detect depression, but it is interesting to think that in the future, our SNS use would be an early detection tool for all kinds of mental illness.

II. METHODOLOGY FOR REVIEW

All the English language publications between January 2011 and October 2020 relevant to the detection of Major Depressive Disorder were extracted and reviewed. Research papers were searched by giving regular expressions like "depression identification through social media posts" and "tweets analysis for depression identification". In total one hundred studies were reviewed. After the research analysis of all these studies, it was evident (as mentioned in the introduction) that different linguistic and computational techniques were used by researchers to identify depression indicative text. Similarly, different features and demographics were used to highlight the behaviors and online activities of depressed users. In the same way, data was collected from different platforms like Twitter, Facebook, Reddit forums, etc. So, keeping in view the diversity presented in research instruments, methods, and algorithms, we compared the studies according to different perspectives such that eleven studies have been compared based on techniques of Sentiment Analysis, Expressions, and Emotions. Six studies have been analyzed for the perspective of annotation guidelines that different research studies

proposed for generating a tagged dataset of depression. A significant amount of work has been done by employing statistical analysis, machine learning, and deep learning techniques. A review of fourteen studies including the techniques employed, results, and strengths is presented. A detailed summary of features and demographics used by various research studies is also demonstrated. Ten most related research studies are summarized based on features and demographics. Lastly, there is a detailed discussion related to nine studies focusing on platforms used for data collection. After reviewing the literature, a comprehensive list of limitations and potential for future research are given.

III. RELATED WORK

Numerous existing approaches exist to symbolize text in social media such as Natural Language Processing, Sentiment Analysis/Opinion Mining, Text Analytics, and Machine learning. In this context, the current review has been done based on the perspectives including how the data was collected, what tools and techniques have been used for depression identification, what kind of data (labeled or unlabeled) has been used, and through which approach the psychological perspectives have been considered. Besides, based on language features, behavioral factors, and results have also been analyzed and interpreted.

A. STUDIES CONDUCTED WITH THE PERSPECTIVE OF SENTIMENT ANALYSIS/EXPRESSIONS/EMOTIONS

Using sentiment analysis Park et al. [8] proved that depressed people showed more negative sentiments than normal people. Similarly, based on the Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition) criteria for MDD, Moreno and colleagues identified depression in the language used in the Facebook posts of college students. Also, sentiments expressed in status updates and emotions expressed in status were examined by Settanni and Marengo [9] and they generated automated words using the Italian version of LIWC.

In his research, Newman et al. [10] referred to a fact that has been proved by various psychologists that depressed users used more negative words and less positive emotions. Polarity can be checked based on three measures (i.e. microblog content, user behaviors, and interactions with others). Yang et al. [11] considered ten features of depressed persons described by psychological researchers. Each feature was analyzed via a regression analysis technique with Binary logistic. The most powerful features in identifying depressed users were a time of being forwarded and time of mentioning others. For measuring facial expressions, manual FACS coding, pitch extraction, and active appearance modeling were used. Results to identify depressed and non-depressed were consistent with DSM-IV criteria. These findings suggest that the sentiments of the users can be analyzed using facial and vocal expressions analysis and there is a difference between the words or language used by depressed users as compared to the non-depressed users.

Park et al. [12] described depressive moods of users from the language used and real-time moods captured in Twitter. The aim was to create two groups to perform cross-sectional analysis and to compare the language of depressed users with non-depressed users. The authors have determined research feasibility by conducting pilot tests. Annotators labeled tweets manually. The tweets were collected using Twitter APIs and filtered manually by making sure that tweets were not written by the social worker who talked about depression. The authors have used the Center for Epidemiological Studies Depression (CES-D) and demographics as features and LIWC (Linguistic Inquiry and Word Count) as a sentiment tool. The results indicated that there is a correlation between the depressive state of a user and the tweet sentiment of that user. The depressed users have used negative emotions more than the non-depressed users but they have the same usage ratio of positive words.

In the year 2013, sentiment analysis was further used for depression detection using a micro-blog online network. Wang [13] used the Chinese Social network Sina Microblog to identify depressed people. The author applied subject-dependent sentiment analysis along with vocabulary and man-made rules to calculate the depression inclination of each micro-blog. Secondly, a depression detection model was constructed based on ten features of depression

derived from psychological research. Some of the psychologists were involved in this study to verify the depression-related features. For the predictive analysis and to verify the model, they have used three kinds of classifiers titled Bayes, Trees, and Rules, whose precisions are all-around 80%. As per the model, the developed application was able to detect depression to monitor mental health online. In the meanwhile, it was detected that one of the features ignored the most while identifying depression is user interaction which was given minimal attention as relationships between depressed people are difficult to analyze. But for a deeper understanding influence of ties between the users could be studied further.

De Choudhury [14] researched depression detection and predict postpartum in terms of postpartum changes like effect, language and behavior. Again, they used statistical models to predict whether the mother has PPD or not. Logistic regression model (demographics model), stepwise regression models, and various other models were developed for both aspects that whether a mother is suffering from PPD and detection of postnatal time horizon based on Facebook activity. Online surveys and interviews were conducted among new mothers (who use Facebook) to share their post-natal experiences. The survey-driven and self-reported data, based on the behavioral activities, linguistic and emotional expressions, indicated differences between mothers with PPD and without PPD. The results indicate that experience of PPD can be best predicted by increased social isolation; due to the stigma associated with mental illness. The less effective predictors were emotional measures.

The participants of the CLPsych 2015 worked to see if PTSD and depression could be predicted by using the data of self-declared patients on Twitter. To figure out words that are most closely associated with PTSD and depression, topic models of language were built by the participants e.g. anxiety topic models have words like stress, feel, worry, hard, time, etc [15]. Sequences of characters were considered as features. A relative count of n-grams was built by applying a rule-based approach. The latter resulted in the highest prediction performance. It was concluded from all the approaches that PTSD and depression (either condition) were hard to detect due to the overlapping of language words.

According to Tandoc et al. [16], symptoms of depressed user's diagnostics and Major depressive episodes (MDE) are concluded by Facebook through disclosures of feelings via 'Status Update' features available on Facebook. Ensemble learning techniques are used to classify depression among non-depressed individuals. The study was about to analyze the depression associated with Facebook envy among college students. This was done through sentiment analysis; the Facebook status updates were analyzed, depression scale was used to identify depression among them and later to identify whether this depression is due to Facebook envy or not. The results simply indicated that Facebook envy is associated with depression among college students.

Chen [17] incorporated measures of basic eight Ekman's emotions (Anger, Disgust, Fear, Happiness, Sadness and Surprise, shame) as features from Twitter posts. We first extracted emotions with their expression intensities as strength scores to create emotion features. A time-series analysis was applied to the emotion strength scores over time and produced a selection of descriptive statistics as temporal features. They showed that emotions expressed in tweets possessed predictive power to indicate the depressive state of a user. The measurements of changes in an individual's emotions over time demonstrated the effectiveness of emotions as features and improve the performance of the proposed model. After learning the traces and patterns of depressed users from these features, the trained classifiers can be easily applied for detecting Twitter users with depression who did not post about their conditions and users who are at risk of depression.

Resnik [18] worked on topic modeling to see how depressed and non-depressed individuals use language differently. They further explored the use of supervised topic models in the analysis of linguistic signals for detecting depression, providing promising results using several models. Qualitative examples have confirmed that LDA, and now additional LDA-like models, can uncover meaningful and potentially useful latent structures.

Depressive disorders may happen in combinations with stress and anxiety. This co-occurring relation was a focus of Budhaditya Saha [19] who focused to classify online communities based on co-occurrence.

A joint modeling framework was constructed by using psycholinguistics features and topics present in the language. A few good examples were found indicating the significance of language features in predicting co-occurring communities interested in depression. Their empirically validated model outperforms the state-of-the-art approaches on the crawled dataset.

Vanhalst et al. [20] determined the sentiments of patients through their facial expressions and identified whether they are associated with loneliness or not. For facial expression recognition, moods such as sad, happy, fear, etc. were also considered. All the participants were from the low-income community and results shown significant associations between loneliness and depression more in women than men.

Investigating online communities for mental health conditions, Dao [21] identified latent Meta groups for depression and autism using various features from blog posts. The researcher has used the HDP algorithm to conclude latent topics from the corpus, which was constructed by gathering information related to mood, affective words, language styles, and generic words in the posts of community members. To discover Meta communities, they applied a nonparametric clustering algorithm. Moreover, while clustered into the same Meta community, shared and common sentiment between depression and autism-related were also analyzed. In their use of latent topics, the separation between groups is illustrated while visualizing discovered online Meta communities. The findings indicated that the sentiment-bearing difference in mental health communities, signifying a possible direction to structure interventions so that support and help can be provided in mental healthcare for vulnerable online communities.

For the summary of the research analysis of all these papers, see table 1.

Table 1 Summary Of Studies Presented In Section Of Sentiment Analysis/Expressions/Emotions

Author	Platform/ Size of Dataset	Research	Domain	Approach to get Features	Features	Algorithms	Annotation	Performance Measure
Park et al. (2012)	Twitter/1018 tweets from 69 participants	Semantic Analysis	No	LIWC as a semantic tool	CESD and demographics (gender, age, occupation, education)	Regression analysis, content analysis	Manual annotation	P-score
Wang et al. (2013)	Sina Microblog / 180 users	Subject dependent semantic analysis	Yes	sentiment analysis utilizing vocabulary and man-made rules	Microblog Content: 1st person singular, +ve and -ve emoticons Interactions: Mentioning, being forwarded	Bayes, Tree, and Rules	Manual annotation	ROC, F measure, MAE, precisions of all of them > 80%

Author	Platform/ Size of Dataset	Research	Domain	Approach to get Features	Features	Algorithms	Annotation	Performance Measure
					, being commented Behaviors: posting time			
De Choudhury (2014)	Facebook/ 600k postings of 165 mothers	Statistical Analysis	No	LIWC as semantic tool	Postpartum changes like affect, language and behaviour , user activity, social capital	Logistic regression model	Not mentioned	Pseudo-R2b 0.32
Tandoc et al. (2015)	Facebook/ 736 college students	Statistical Analysis	Not mentioned	Not mentioned	Facebook use, Envy, gender, age, friends list	Regression analysis	Not mentioned	P Score, T Score
Chen (2018)	Twitter/ 585 users with 2000 tweets each	Sentiment analysis	Not mentioned	Emotive + LIWC	Ekman Emotions: Anger, Disgust, Fear, shame, Happiness, Sadness and surprise	LR, SVM, NB, DT and RF	Not mentioned	Prediction accuracy > 85%
Resnick (2015)	Twitter / 3M tweets from 2,000 Twitter users	Topic modeling	Yes	Unigrams, LIWC	LDA and sLDA features	sLDA , supervised anchor topic models, supervised nested	Manual	precision at R=0.5 to 74% and precision at R=0.75 to 62%.

Author	Platform/ Size of Dataset	Research	Domain	Approach to get Features	Features	Algorithms	Annotation	Performance Measure
						LDA model		
Saha et al. (2016)	Live Journal /620 000 posts of 80000 users in 247 online communities	Sentiment analysis, topic modeling	Not mentioned	LIWC, LDA for topics	Language and topics	STL regression and MTL framework	Manual	Auc, sensitivity, and specificity
Dao (2017)	Live Journal/24 communities with at least 200 posts	Topic modeling, sentiments analysis	Not mentioned	ANEW, LIWC	ANEW features, LIWC features, Generic word-based topic features, mood tags	LDA, affinity propagation algorithm, Bayesian nonparametric (BNP) topic modeling	Not mentioned	Not mentioned

B. STUDIES CONDUCTED WITH THE PERSPECTIVE OF DATA ANNOTATION

Annotation guidelines are required to label the data based on defined entities to make the contextual analysis of the sentence accordingly [22]. Few studies have been conducted to provide annotation guidelines for depression disorders since 2012, before that annotation was limited to the field of linguistics.

For performing sentiment analysis to detect mental illness from public posts, Ji et al. [23] and Amir et al. [24] addressed the issue of contextual analysis, which indicated that the context of the sentence should be known while analyzing what this sentence is about. Based on PHQ9 symptoms, Saxena [25] provided manual guidelines for annotating the Twitter data. Similarly, based on DSM-IV and DSM-V, Mowery et al. [26, 27] based on depression symptoms and Psycho-Social stressors provided manual annotation schemes.

Mowery et al. [28] present a new annotation scheme representing depressive symptoms (derived from DSM-5) and psycho-social stressors (derived from DSM-4) associated with major depressive disorder (derived from DSM-5 manual). With the help of annotators, the researchers have applied the scheme to Twitter data. They concluded that there are considerable challenges including the selection of target entity and attributes in attempting to reliably annotate Twitter data for mental health symptoms. However, with the help of domain experts, the tweets were labeled; but this annotation scheme could be used to generate a tagged dataset, and ML techniques could be applied for better insights.

Cavazos-Rehg [29] coded tweets using symptoms from the DSM-5 manual for Major Depressive

Disorder (MDD). Supportive or helpful tweets about depression were the most common theme (n=787, 40%), closely followed by disclosing feelings of depression (n=625; 32%). Two-thirds of tweets revealed one or more symptoms for the diagnosis of MDD and/or communicated thoughts or ideas that were consistent with struggles with depression after accounting for tweets that mentioned depression trivially.

Saxena [25] proposed an enhanced approach that considers explicit as well as implicit depression-indicative symptoms. This research aimed to examine whether the post is depression indicative or not; that was done by three annotators following the labeling guidelines for assistance, using the ground truth dataset of self-reported users. It should be noted that the annotation scheme was developed based only on the symptoms mentioned in the PHQ-9 scale. The second stage was to determine the severity of depression indicated in those specific posts; for that purpose, PHQ-9 was used to define the symptoms and compare them with the depression indication mentioned in the post. The researcher has mainly labeled the data as 0 (non-depressed) or 1 (depressed) and measured the F-measure over the chosen baseline.

A regression model to predict users' degree of depression was built by Andrew Schwartz [30] by using their status updates and survey responses. Moreover, it was found that the degree of depression increases from summer to winter, results showed consistency with the literature. Seven depression facet items from the larger Neuroticism item pool were used to estimate the degree of depression. Topics, n-grams, lexica, and several words were considered as Language features excluding friend networks and other online activities of users. Table 2 summarized the schemes used for annotation in previous studies. According to which, the authors have used different questionnaires including DSM-IV, DSM-V, and PHQ-9 for depression concerns i.e. to match symptoms from these clinically approved questionnaires. The scheme survey responses show that the author has made a list of questions and gave it to annotators to answer those questions reviewing the data and label that accordingly. See table 2.

Table 2 Summary Of Studies Presented In Section Of Dataset Annotation

Author/Year	Annotation Scheme (Based on)
Mowery et al. (2015)	DSM-IV and DSM-V
Saxena (2018)	PHQ9 symptoms
Schwartz (2014)	Survey responses
Cavazos-Rehg (2016)	DSM-V

C. STUDIES CONDUCTED WITH THE PERSPECTIVE OF STATISTICAL ANALYSIS AND MACHINE LEARNING TECHNIQUES

While choosing the algorithms there is a gradual shift from qualitative analysis towards statistical techniques like Correlation and Regression Park et al. [31], and then finally in most recent research studies different machine learning algorithms have been employed for model construction [32,33]. See tab. 3 for detailed research analysis of studies with the perspective of algorithms used.

Mourao et al. [34] differentiated depressed individuals from healthy ones based on many brain properties by using SVM. This pattern recognition approach classified depressed and non-depressed groups by taking into account brain activities and structure. SVM can be used to diagnose neurological and psychiatric disease, prediction for treatment onset and response.

De Choudhury et al. [35] explored the possibilities for weighing social media posts for a better understanding of depression in populations. They gathered around 69K Twitter posts from clinical depression sufferers by using the crowdsourcing

technique. They developed a probabilistic model (SVM classifier) on this corpus to predict if Twitter posts could be depression-indicative or not. The SVM-based model includes signs of social activity, emotion, and language demonstrated on Twitter, as features. Based on this trained model, the authors developed a social media depression index to illustrate levels of depression in populations.

Coppersmith et al. [36] analyzed the language of users in Twitter data for depression, post-traumatic stress disorder (PTSD), seasonal affective disorder (SAD), and bipolar disorder. They demonstrated the effectiveness of NLP techniques is yielding insights for specific disorders along with evidence. They employed two Language models, 1-gram, and 5-gram and examined sequences up-to 5 characters. Further, they built classifiers to separate each group from the control group. They provided some validations with LIWC.

Tsugawa et al. [37] have demonstrated the effectiveness of using social media platforms among Japanese Twitter users to recognize depression. In addition to using Twitter User's activity history, a web-based questionnaire was used to collect ground truth data in predicting the existence of depression for Twitter users. In addition to the features used by De Choudhury et al. [35], Tsugawa et al. [37] used bag-of-words and word frequencies to identify the ratio of tweet topics. Even though subtle changes in behavioral features were identified between their research and the one done by De Choudry et al. [35] which could be due to cultural aspects, the authors have identified similar analytical patterns for the use of negative words, posting frequency, re-tweet rate, and the tweets containing URL. Feature engineering using Twitter user activity positively contributed towards a classification accuracy of 69%, with 0.64 precision, and 0.43 recall using support vector machine (SVM) classifiers. Topics identified using topic modeling also added positive contributions to the predictive model compared to the use of the bag-of-words model, which could result in overfitting. Regarding the amount of Twitter data required in identifying depression, the authors have highlighted that at least two months of data is sufficient and data over a longer period could lead to lower accuracy.

Coppersmith et al. [38] created a large, varied, language-centric dataset and provided significant grist for the field of ten mental illnesses (ADHD, Anxiety, Bipolar, Borderline, Depression, Eating, OCD, PTSD, Schizophrenia, and Seasonal Affective). By identified self-declared statements from Twitter data, they performed various statistical analyses to examine a broad range of mental health conditions. Language differences were explored systematically amongst these ten diseases concerning the general population, and to each other by using LIWC.

Even before the onset of MDD, Nadeem [39] explored the predictive potential of social media. Data was collected by crowdsourcing from self-declared users. Tweets were examined by using the bag-of-words approach and the risk of depression was estimated via various statistical classifiers. They addressed the issue as a text classification problem instead of using behavioral features and achieved 81% accuracy.

Amir [40] researched to explore the extent to which user embedding captures information relevant to mental health analysis. The aim was to explore that whether user embedding (words embedding to rank the document and know the context of the sentence) is flexible enough to discriminate between users diagnosed with mental illness and demographically matched controls. The authors wanted to know that the user embedding in Twitter will help them to develop a public health system or not. If it is correlated with mental health, they can characterize and classify the risk groups on social media. Considering the modest size of the dataset, the authors adopted the Non-Linear Subspace Embedding approach (NLSE) that is based on small, labeled data used for generic representations. Various techniques are there for word embedding learning; however, the conditional probability-based method was used for the specific study. Besides, a comparison of user embedding models was done; for instance, if the user is suffering from depression, the bag-of-words model can easily be pick-up on such clues. To measure homophily, induced rankings were evaluated concerning average Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC) curves. The results imply that they carry information about the stated condition. The findings also indicate that users with PTSD and depression along with demographically matched controls can indeed capture mental health-related signals. These captured

user embedding would allow clinical psychologists to get the benefit of moment-by-moment quantification using data from smartphones.

Considering the statistical techniques like regression, random forest trees, sliding window approach, and temporal modeling, Stepanov et al. [41] were aimed to develop an automatic way (technique) to detect the extent and presence of depression. The team has taken a corpus of transcribed speeches and audio/video recordings from the AVEC workshop 2017. Extracting features from different modalities including audio, video, and language, PHQ-8 scores were automatically predicted. The authors have performed experiments on features like speech, language, and behavior to predict depression severity. The results achieved for behavioral characteristics were on the lowest error, and the best was achieved for audio features. The model gave surprising results in the case of visual features; it failed to generalize enough the unseen data.

Reece [42] and his team predicted the occurrence of PTSD and depression in Twitter data. They built computational models on considering linguistic style, measuring effect, and context of tweets as features. Supervised learning-based models successfully differentiated between depressed and healthy content, albeit in a separate population. The state-space temporal analysis suggested that depression can be detected from tweet data even before the onset. Similarly, they revealed the indication of PTSD after trauma through the state-space time series model even before the diagnosis. This research forms the basis of data-driven predictive techniques for early screening and detection.

Based on the semi-supervised statistical model, Yazdavar et al. [43] explored the clinical depression from tweets. This research highlighted the potential for detecting depression by analyzing Twitter data considering PHQ-9 symptoms. The alignment of medical findings based on PHQ-9 diagnosis was to be evaluated with the expression and duration of these symptoms on Twitter. The selected model identified clinical depressive symptoms in the tweets with an accuracy of 68%.

Benton [44] introduced the initial groundwork for estimating suicide risk and mental health in a deep learning framework. By modeling multiple conditions, the system learns to make predictions about suicide risk and mental health at a low false-positive rate. Conditions are modeled as tasks in multitask learning (MTL) framework, with gender prediction as an additional auxiliary task. We demonstrate the effectiveness of multi-task learning by comparison to a well-tuned single-task baseline with the same number of parameters. Our best MTL model predicts potential suicide attempts, as well as the presence of atypical mental health, with AUC > 0.8. We also find additional large improvements using multi-task learning on mental health tasks with limited training data.

By utilizing the public social media activity on Twitter, Jamil [45] proposed an SVM-based user-level classifier to identify users at-risk. Similarly, for tweets containing disease symptoms, a tweet-level classifier was implemented. Their proposed platform can be used for raising an alarm so that help could be provided to users at risk. It was a manual method of tagging and perhaps community features (such as user's location, their network) can provide further insights.

Wang and Singh [46] worked to predict depression using tweets and to generate a labeled dataset. Textblob's python package was used to label the text with the help of a polarity score. Deep learning models like CNN, RNN, and GRU were constructed for prediction along with the use of logistic regression and SVM. Dataset was collected from the pages with expressions "depression quotes", "damn depression" and "depression notes". The tweets were labeled manually, after that a script was written in python to calculate the polarity score and label each tweet accordingly (0 for non-depressed and 1 for depressed). The authors later cross-checked the manual and automated predictions; the effect of words-based and characters-based models learned embedding's and pertained embeddings were examined qualitatively. As per the comparison between different models, GRU captures long-term dependencies, along with better accuracy.

Weerasinghe et al. [47] analyzed different machine learning algorithms and various features like word clusters, the bag-of-words, topic models, and part of speech n-grams to predict mental illness through text posted on Twitter. In their research, they not only confirmed some old patterns but also identified some new patterns in the posts of mentally sick

people.

Tadesse [48] examined posts of Reddit users to identify depression. They used NLP and Machine learning techniques. The best single feature was bigram with SVM that gives 80% accuracy and 0.8 F1 score. They also measured the combining effect of LIWC + LDA + bigram and achieved 91% accuracy and 0.93 F1 score with Multilayer Perceptron.

Trotzek [49] addressed the early detection of depression using machine learning models based on messages on a social platform. In particular, a convolutional neural network based on different word embedding was evaluated and compared to a classification based on user-level linguistic metadata. An ensemble of both approaches was shown to achieve state-of-the-art results in a current early detection task. Furthermore, the popular ERDE score as a metric for early detection systems was examined in detail and its drawbacks in the context of shared tasks were illustrated. A slightly modified metric was proposed and compared to the original score. Finally, a new word embedding was trained on a large corpus of the same domain as the described task and was evaluated as well.

Zogan [50] proposed a hybrid model based on BiGRU and CNN models. They assigned the multi-modalities attribute which represents the user behavior into the BiGRU and user timeline posts into CNN to extract the semantic features. Their hybrid model showed 85% accuracy and 0.81 F1 score. Hence, it improves classification performance and identifies depressed users outperforming other strong methods. See table 3.

Table 3 Summary Of Studies Presented In Section Of Dataset Annotation

Author	Platform/Data Size	Mental Illness Criteria	Features (Predictors)	Model	Results
De Choudhury (2013)	Twitter/ 476	Survey (CESD+BDI)	LIWC, Sentiments Metadata, social networks N-grams, LIWC	PCA,SVM w/RBF Kernel	Accuracy 0.72
Coppersmith (2015)	Twitter/ 21866	Self-declared	Sentiments, Metadata, user activity	Log Linear Classifier	Precision Depression=.48 Bipolar =.64 PTSD=.67 SAD=.42
Tsugawa et al. (2015)	Twitter/209	Survey (CESD)	N-grams, LIWC, Sentiments Topics, Metadata, user activity	SVM	Accuracy 0.69
Coppersmith (2015)	Twitter/4026	Self-declared	N-grams, LIWC	not mentioned	Precision Depression=.48 Bipolar =.63 Anxiety=.85 Eating Dis=.76
Nadeem (2016)	Twitter/ 2.5M tweets from 900 Corpus from the AVEC workshop 2017	Self-declared	N-grams/gender	Decision Trees, Linear Support Vector Classifier, Logistic Regression,	ROC AUC score of 0.94, a precision score of 0.82, and an 81% accuracy; for naïve

Author	Platform/Data Size	Mental Illness Criteria	Features (Predictors)	Model	Results
				Ridge Classifier, Naïve Bayes	Bayes development set
Stepanov (2017)	Twitter/ 279951 tweets from 378 users	PHQ-8	speech, language and visual features extracted from face	Regression, random forests trees, sliding window approach and temporal modeling	MAE 4.66 Behavioral set MAE 4.73 Language features MAE 5.17.
Reece et al. (2017)	Twitter/ 10400 tweets	Survey (CESD)	LIWC, Sentiments, Metadata Time series, LabMT	Random Forest	AUC depression=.87, PTSD=.89
Yazdavar et al. (2017)	Twitter/ 156,612 tweets	PHQ-9	PHQ-9 symptoms as features	LDA, K-means, LSA, BTM, Partially Labeled LDA	accuracy of 68% and precision of 72%
Jamil (2017)	Twitter/ 279951 tweets from 378 users were obtained from 25,362 users	Self-declared	Text driven features: Location, Metro regions, ID, Tweet's ID, Node , Name , Screen name , Description Image, Geo coordinates of a user's location, Text, URL Media: image, clip Message time, Time, Retweet count, Favorite count, Created at Retrieval time of tweet being replied to Favorites count	LDA, SVM	Recall 0.8750 precision 0.7778
Benton (2017)	Twitter/ 9611 tweets	Self-declared	N-grams, Gender	Neural Network	AUC Depression=.79 Bipolar=.75

Author	Platform/Data Size	Mental Illness Criteria	Features (Predictors)	Model	Results
Wang & Singh (2018)	Twitter/ 13385 tweets	Not mentioned	Textual features	CNN, RNN and GRU, SVM, Logistic Regression	Depression=.76 Suicide Attempt= .83 Accuracy above 97%
Weerasinghe (2019)	Twitter/ 3000 tweets from 327 users	Self-declared	bag-of-words, word clusters, part of speech n-gram features, and topic models	Linear SVM	Precision, Recall, F1
Tadesse (2019)	Around 2000 Reddit posts	Self-declared	bigrams LIWC + LDA + bigram	SVM Multilayer Perceptron	Accuracy=80%, F1=.80 Accuracy=91%, F1=.93
Zogan (2020)	Twitter posts of 1402 users	Self-declared	Multi-Modalities + Word Embedding	Bidirectional Gated Recurrent Units (BiGRUs) and Convolutional Neural Networks	Accuracy=85%, F1=.81
Trotzek (2020)	Reddit/ 135 users	Self-declared	Linguistic metadata features	Neural Network	ERDEo Scores, F1 Score, Precision, and Recall

D. STUDIES CONDUCTED WITH THE PERSPECTIVE OF DEMOGRAPHICS

It has been studied that a key idea to process textual information over SNS profiles was proposed by Hu and Liu [51]. But the findings indicate that gap is in the demographics factor; there is a need to identify unique demographical factors. For instance, usage patterns of OSNs (Online social networks) should be considered. In this regard, various studies have been reviewed that were conducted with the major aim of covering and studying the demographics factors while identifying depression using social media.

Dao, Nguyen, Phung and Venkatesh [52] analyzed the impact of age, and social connectivity on the online messages of members of an online depression community by using statistical techniques and ML methods. Feature extraction was done using LDA (probabilistic modeling tool) while language style was captured using LIWC. Live journal online community was used to collect the data. Two features were considered while selecting the post i.e., the writing style (language) and the topic of the post. Using ML and statistical techniques, the posts were discriminated against based on low vs. high valence mood, different degrees of social connectivity, and different age categories. Three sub-corpuses were formed based on the features such as age, mood, and social connectivity. These features were further

classified for age 22-26 and less than 51; the mood was measured using Affective Norms for English Words and social connectivity was measured on basis of several friends, several community memberships, and many followers. The statistical findings indicate that the language styles of the people with high and low moods are different, along with that the people with different degrees of social connectivity have different topics. The authors concluded that people have the potential of using social media in case of depression screening, specifically in an online setting. According to Benton [53], anxiety prediction was improved on a shared dataset by considering gender in addition to 10 comorbid conditions. Coppersmith [38] has used psychological dictionaries Linguistic Inquiry and Word Count [54] to characterize differences between mental illness conditions, with some success. Preotiuc-Pietro et al. [55] observed that age estimation of users
successfully identified users having PTSD diagnosis, and depression and PTSD language predictors were largely overlapping with the language predictive of personality. This suggests that users with a particular personality or demographic profiles chose to share their mental health diagnosis on Twitter, and thus that the results of these studies (mostly, prediction accuracies) may not generalize to other sources of autobiographical text.

Settani et al. [56] included age and gender to extend demographics factors and to investigate the relationship between textual content present on SNS and self-reported measures of emotional well-being. Emotional text content along with positive and negative emoticons are considered to highlight the differences between age groups. They correlated the emotional textual content with depression, stress, and anxiety via statistical analysis. Participants were then self-reported Facebook users from North Italy. Textual data of four trimesters contained posts and comments including emoticons. Automated textual analyses were conducted with LIWC. In short, this study significance and feasibility of examining emotional well-being through studying profiles of individuals.

Yang [57] used GIS technology to analyses spatial patterns that automatically detected depressed users on Twitter. DSM-IV criteria were used to detect depressive users. They examined the risk factors at a country level and identified the impact of education, income, and race on the depression rate. While performing statistical analysis and stepwise regression, they found that the relationship between seasonality, climate, and depression was localized and geographically different. Many seasonal factors (Relative humidity, temperature, sea level pressure, precipitation, snowfall, weed speed, global solar radiation, and length of day) contributed to the geographic variations of depression rate. A semi-automated three-stage framework was proposed to analyze geographically distributed health issues.

Cavazos-Rehg [29] and his team performed identification of depression through tweets. They aimed to explore the common themes of depression-related conversation on Twitter and to examine how well do some tweets correspond with clinical symptoms of depression. Besides, the demographic characteristics of Twitter users such as age, gender, race/ethnicity, marital status, income, occupation, and location of the tweeters were also analyzed. The aim was to identify what kinds of posts have been posted related to depression (symptoms or treatments); after running the analysis on data, it was found that inferred characteristics of Tweeters vary along with expressed feelings of depression versus the typical Twitter user. It was found that out of 2000 tweets, 787 were supportive or helpful tweets about depression, while 625 tweets were closely followed by disclosing feelings of depression.

Tian [58] analyzed Sina Weibo postings to learn themes and built a text classifier to identify the postings indicating depression. The depressed population was compared on demographic characteristics, diurnal patterns, and patterns of emoticon usage. Disclosure of depression was found out as the most popular theme. Their findings indicated that depressed people were more engaged on social media and more active during sleep time, and the usage frequency of negative emoticons was also high.

D. STUDIES CONDUCTED WITH THE PERSPECTIVE OF FEATURE SELECTION

Analyzing social media data to identify depressed people various features have been considered for tweet level and user level analysis. Behaviors of depressed and non-depressed users are different and

observable in terms of language usage and online activities that they use personal pronouns and negative words more frequently and their frequencies of interaction with others are also low. Besides language features, posting time, no. of posts, likes, re-tweets, night-time activity, comments, no. of followers and followings, and personal profile features such as gender, and age can also be used in describing depressed and non-depressed individuals. Variation patterns can also be observed over multiple days.

De Choudhury [59] used post-centric features (emotions, time, linguistic style, n-grams) and user-centric features such as engagement (volume, reply, re-tweets, links) and ego-network (in-links, out-links) to understand the depression in populations. De Choudhury selected 49 different features of mothers' activities including age, ethnicity, income, occupation, linguistic style, emotion, and social capital. On the other hand, Park et al. [31] proved an increase in the posting frequency in depressed people for 6 months. Shaw and colleagues also showed similar findings that depressed people show more frequent user engagement on Facebook. Similarly, identity items like relationship status 'single' has been more closely linked with depression and anxiety. Although some of the specific findings are mixed studies generally suggest that social anxiety may be visible on SNSs through compensatory behaviors (increases in information disclosure) or relative inactivity or social withdrawal.

Reece [42] extracted various statistical features from Instagram photos, using color analysis, metadata components, and algorithmic face detection to identify depression disorder. Phone sensor data can be used to detect Depression. It includes GPS sensing, proximity sensor, sleeps markers i.e., bedtime/ wake time, etc.), social media usage (status updates, posts, last account activity, etc.). Schwartz et al. [60] translated raw sensor data into knowledge extracted features to identify stress, moods, and behaviors. Fu et al. [61] perform feature selection over biomarkers associated with diagnosing depression and its treatment response. Feasibility for various challenges of clinical development was shown in this study by using sensor data from personal sensing.

Mowery et al. [28] aimed to identify the role of features; for instance, lexical features are critical for identifying depressive symptoms. Also, to determine top-ranked features that produced the optimal classification performance. Two experiments were performed by using Supervised ML classifiers on Twitter Data Set. However, there is no consistent count of features for predicting depressive-related tweets. But still, identification of most discriminating feature sets and natural language processing classifiers for each depression symptom and classification of rarer depressive symptoms which can lead to the major depressive disorder should be considered next.

For a detailed summary of the attributes/features, see table 4.

Table 4 Summary Of The Attributes

Tweet level Attributes		User-Level Attributes	
Linguistic	Positive & Negative Emotion Words	Posting Behavior	Social Engagement (Length of time)
	Positive & Negative Emoticons		Tweeting time (Insomnia index:)
	Punctuation Marks & Associated		Tweeting type
	Emotion Words		Tweeting linguistic style
	Degree Adverbs & Associated		Post per day
	Five-color theme	Social	Social Attention (Number of comments, re-tweets,

Tweet level Attributes		User-Level Attributes	
			and likes)
	Saturation		Post per user (volume)
	Brightness		Content Style (Words, Emoticons)
	Warm/Cool color		Social Influence (Stressed Neighbor
		Social Interaction	Count, Strong-tie Count, Weak-tie
Visual			Count, Follower Count, Fans Count)
	Clear/Dull color		Social Structure
		Metadata	average word count per tweet
		Demographic characteristics	Age, gender, race/ethnicity, marital status, income, occupation, and location of the Tweeter
	Pixel Analysis		

E. STUDIES CONDUCTED WITH THE PERSPECTIVE OF DATA COLLECTION SOURCES

The most common form to get data from users was to ask them to fill the survey forms based on psychologically approved questionnaires e.g., CES-D, PHQ, BDI, and DSM-V manual. But on the other hand, various approaches have also been used such as to get data of self-declared users from depression-related forums or search data through Twitter API or public profiles on #depressed or #depression with the search string "I am diagnosed with depression".

It is worth mentioning that many studies used publicly accessible data. Regular expressions like "I am diagnosed with depression" were used to identify self-declared depressed users on social media platforms like Twitter, Live Journal, etc. Another source is online forums and discussion groups. They offer a space in which users can ask for advice, receive and provide emotional support, and generally discuss stigmatized mental health problems openly.

Leiva [62] analyzed online text messages on Reddit and worked to detect depression before it gets severe. He employed many machine learning techniques i.e. logistic regression, k-nearest neighbors, random forest, SVM, genetic algorithms, and various sentiment analysis techniques. Tsugawa et al. [63] gathered a Japanese sample from Twitter data and predicted depression by using the CES-D criteria. The most recent 6-16 weeks tweets were enough for identifying depression.

Shen et al. [64] detected depression among the users on large scale through harvesting social media data. Also, the online behavior of depressed and non-depressed users was analyzed. Comparison methods include Naïve Bayesian, Multiple Social Networking Learning (MSNL), and Wasserstein Dictionary Learning (WDL). Multi-modal depressive dictionary learning Data was collected using Twitter APIs, based on the tweets between 2009 and 2016. Three datasets were created i.e. depressed using the pattern I am diagnosed with depression and non-depressed if the users have never used the word "depress" and third included unlabeled dataset.

Schwartz et al. [60] used a personality survey across Facebook users' to determine continuous depression. This study observed depression fluctuations in individuals over different seasons. A list of phrases, words, topics that are more closely associated with depression was also provided in this study. Authors believe that the most reliable way of data gathering in making predictive systems is based on surveys. But it

incurred high costs thus motivated users to get publicly accessible assessment criteria.

Bagroy [65] used Reddit forum posts and studied the mental well-being of university students. Posts of university subreddits were collected and estimated distress level. Their findings suggest that the ratio of mental health posts was high during the academic year in universities with quarter-based schedules rather than semester-based schedules. To find an association between depression and social media users, Aldarwish et al. [66] collected a dataset from Facebook, LiveJournal, and Twitter. By using this user-generated content their artificial intelligence-based proposed system classifies users. For classification, they used SVM and Naïve Bayes. Further, they compared manual and confusion matrix-based predictions.

A neural network-based architecture for depression detection and self-harm risk classification was proposed by Yates [67]. Their described Reddit Self-reported Depression Diagnosis (RSDD) dataset contained posts of over 9000 self-reported users. Further, they applied the classification to identify depressed users and the proposed model outperforms in terms of Recall and F1. They also applied a classification approach to the task of estimating the self-harm risk posed by posts on the ReachOut.com mental health support forum.

De Choudhury et al. [68] gathered data of Reddit users who talked about concerns regarding mental health and then moved towards suicidal ideation. Various features were studied to predict this move: poor linguistic style matching with the community, expressions of hopelessness, heightened self-focus, reduced social engagement, and anxiety, impulsiveness, and loneliness.

Reviewing the research studies, it is evident that various studies have been conducted from different perspectives; the summary of all the research studies is given in figure 1 below. It shows the timeline of research studies from 2012 to 2018, conducted with the previously discussed perspectives.

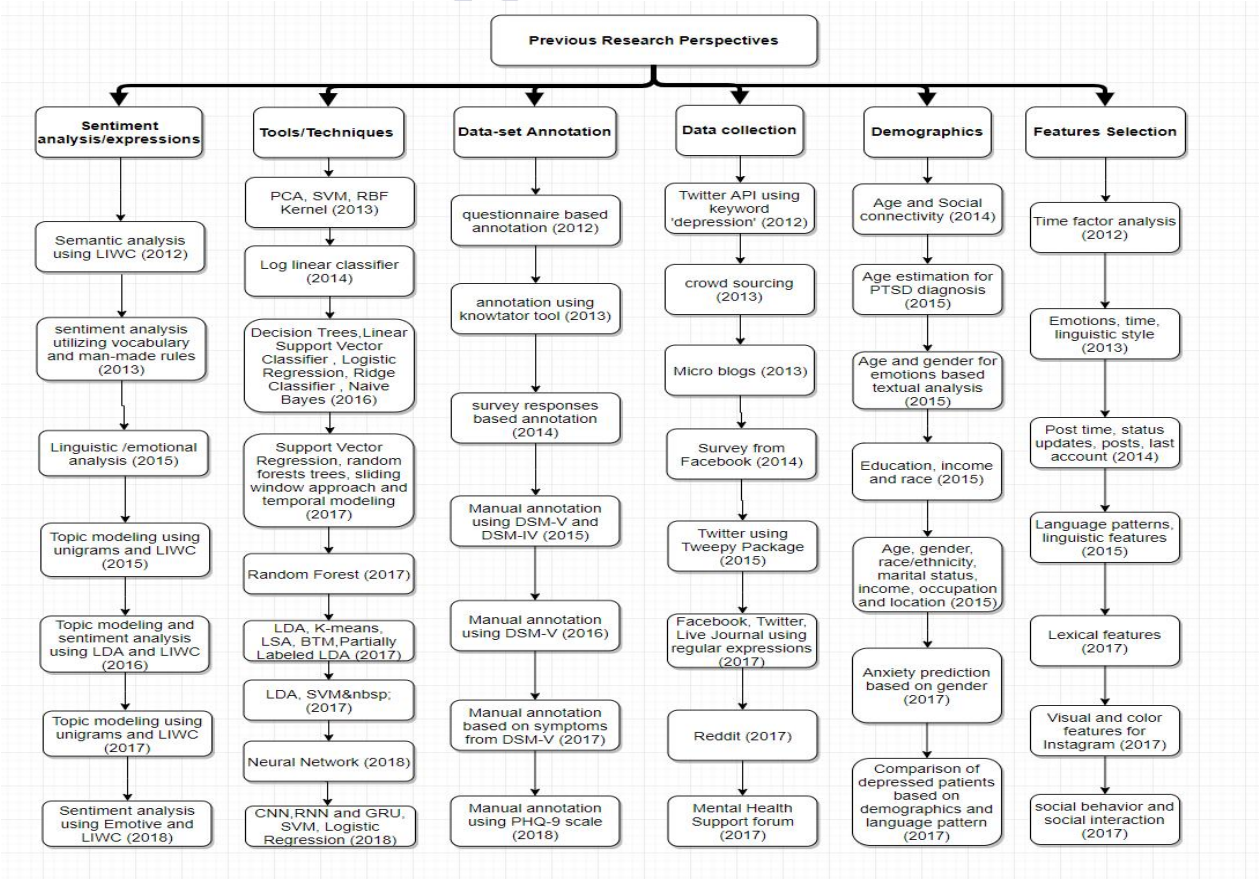

FIGURE 1. Summary of previously conducted studies.

IV. LIMITATIONS AND FUTURE DIRECTIONS

Detection of undiagnosed cases of depression might be the greatest potential of social media. But the studies presented so far did not explicitly focus on identifying people who are unaware of their depression state. Previous research studies did not consider peoples' status behind the scenes e.g., their social and cultural status or socio-economic status De Choudhury et al. [35]. Additional sources of behavioral data such as web surfing history, Facebook feeds, search query logs, and antidepressant purchases could be helpful for future research. Similarly, the viral impact of depression based on people's socio graph should also be studied to measure depression rate at the community level.

User interaction is given minimal attention as relationships between depressed people are difficult to analyze. But for the deeper understanding influence of ties between the users could be studied further [32]. Moreover, studies focusing on schemes of annotation and guidelines were purely manual and were not applied to a sample dataset. However, these annotation schemes could be used to generate a tagged dataset so that ML techniques could be applied for better insights [26]. Only symptoms given in PHQ-9 were considered for annotation. Treatments, medicine names, and negative words used by people can further be the focus of annotation schemes [25]. The extent to which a post content can lead to harm or suicidal intent should be investigated [26].

Informed features should be known and studied before conducting the research. For example, previous diagnosed information and response time of the participants are the most informed features to predict depression using PHQ-8. [41]. In the future, more multiple modalities can be combined to improve the performance and avoid features in any of these modalities. Datasets in most studies were not labeled with the help of domain experts. In the future, there is a need to cover the gaps by using improved encoding-decoding architecture, which is appropriate for sentiment classification. The stability of accuracy and loss graphs could be improved with the model's improvement. There should be a dataset labeled by domain experts rather than prescribing labels of "depressed" and "non-depressed" [46].

Emoticons and newly coined words (e.g., LOL) were ignored. A limited LIWC dictionary was used. In the future, the dictionary should be extended. Offline clinical studies would be incorporated with the online data. Social relationships for sentiments should be considered while predicting depression in terms of change in behavioral activities [51].

V. CONCLUSION

In this paper, various studies that have been reviewed aimed to suggest that depressive disorders are identifiable from the data obtained through social networking sites such as Facebook, Twitter, and depression-related forums. Advancements in machine learning and natural language processing methods support the screening of social media data at a larger scale. The analysis of previous research studies including the comparisons of techniques, datasets, features, a summary of results, strengths, weaknesses obtained by various researchers available as existing literature is also presented. No doubt self-monitoring of individuals' mental health could help to increase the well-being of an identity. Several techniques such as gaming apps and principles of applied behavior analysis could be advantageous if introduced over smartphones or social media profiles.

REFERENCES

[1] Gs.statcounter.com. Social Media Stats Pakistan | StatCounter Global Stats. 2018. [Online]. Available: http://gs.statcounter.com/social-media-stats/all/pakistan

[2] J. Hussain, M. Ali, H. S. M. Bilal, M. Afzal, H. F. Ahmad, O. Banos, and S. Lee, "SNS based predictive model for depression," In International Conference on Smart Homes and Health Telematics, pp. 349-354, 2015.

[3] T. Libert, D. Grande and D. Asch, "What web browsing reveals about your health", BMJ, vol. 351, no. 165, pp. h5974-h5974, 2015.

[4] P. Wang, M. Lane, M. Olfson, H. Pincus, K. Wells and R. Kessler, "Twelve-Month Use of Mental Health Services in the United States", Archives of General Psychiatry, vol. 62, no. 6, p. 629, 2005.

[5] M. Couper, "The Future of Modes of Data Collection", Public Opinion Quarterly, vol. 75, no. 5, pp. 889-908, 2011.

[6] K. Wegrzyn-Wolska, L. Bougueroua, and G. Dziczkowski, "Social media analysis for e-health and medical purposes," In 2011 International Conference on Computational Aspects of Social Networks (CASoN), pp. 278-283, 2011. IEEE.

[7] E. Seabrook, M. Kern and N. Rickard, "Social Networking Sites, Depression, and Anxiety: A Systematic Review", JMIR Mental Health, vol. 3, no. 4, p. e50, 2016. Available: 10.2196/mental.5842.

[8] 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, 1-8. New York, NY: ACM, 2012.

[9] M. Settanni and D. Marengo, "Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts", Frontiers in Psychology, vol. 6, 2015. Available: 10.3389/fpsyg.2015.01045.

[10] M. Newman, C. Groom, L. Handelman and J. Pennebaker, "Gender Differences in Language Use: An Analysis of 14,000 Text Samples", Discourse Processes, vol. 45, no. 3, pp. 211-236, 2008. Available: 10.1080/01638530802073712.

[11] Y. Yang, C. Fairbairn and J. Cohn, "Detecting Depression Severity from Vocal Prosody", IEEE Transactions on Affective Computing, vol. 4, no. 2, pp. 142-150, 2013. Available: 10.1109/t-affc.2012.38.

[12] 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, 1-8. New York, NY: ACM, 2012.

[13] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network." In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 201-213). Springer, Berlin, Heidelberg, 2013.

[14] 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 (pp. 626-638). ACM, 2014.

[15] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "CLPsych 2015 shared task: Depression and PTSD on Twitter." In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 31-39), 2015.

[16] E. Tandoc, P. Ferrucci and M. Duffy, "Facebook use, envy, and depression among college students: Is facebooking depressing?", Computers in Human Behavior, vol. 43, pp. 139-146, 2015. Available: 10.1016/j.chb.2014.10.053.

[17] 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(pp. 1653-1660). International World Wide Web Conferences Steering Committee, 2018.

[18] 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 (pp. 99-107), 2015.

[19] B. Saha, T. Nguyen, D. Phung and S. Venkatesh, "A Framework for Classifying Online Mental Health-Related Communities With an Interest in Depression", IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 4, pp. 1008-1015, 2016. Available: 10.1109/jbhi.2016.2543741.

[20] Vanhalst, B. Gibb and M. Prinstein, "Lonely adolescents exhibit heightened sensitivity for facial cues of emotion", Cognition and Emotion, vol. 31, no. 2, pp. 377-383, 2015. Available: 10.1080/02699931.2015.1092420.

[21] B. Dao, T. Nguyen, S. Venkatesh and D. Phung, "Latent sentiment topic modelling and

nonparametric discovery of online mental health-related communities", International Journal of Data Science and Analytics, vol. 4, no. 3, pp. 209-231, 2017. Available: 10.1007/s41060-017-0073-y.

[22] S. Mohammad, "A practical guide to sentiment annotation: Challenges and solutions." In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 174-179), 2016.

[23] X. Ji, S. A. Chun, and J. Geller, "Knowledge-based tweet classification for disease sentiment monitoring." In Sentiment Analysis and Ontology Engineering (pp. 425-454). Springer, Cham, 2016.

[24] S. Amir, G. Coppersmith, P. Carvalho, M. J. Silva, and B. C. Wallace, "Quantifying Mental Health from Social Media with Neural User Embeddings.", 2017. Available: arXiv preprint arXiv:1705.00335.

[25] A. Saxena, "A Semantically Enhanced Approach to Identify Depression-Indicative Symptoms Using Twitter Data", (Doctoral dissertation, Wright State University), 2018.

[26] D. Mowery, C. Bryan, and M. Conway, "Towards developing an annotation scheme for depressive disorder symptoms: A preliminary study using Twitter data." In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 89-98), 2015.

[27] D. Mowery, C. Bryan, and M. Conway, "Feature studies to inform the classification of depressive symptoms from Twitter data for population health." arXiv preprint arXiv:1701.08229, 2017.

[28] D. Mowery, C. Bryan, and M. Conway, "Towards developing an annotation scheme for depressive disorder symptoms: A preliminary study using Twitter data." In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 89-98), 2015.

[29] P. Cavazos-Rehg et al., "A content analysis of depression-related tweets", Computers in Human Behavior, vol. 54, pp. 351-357, 2016. Available: 10.1016/j.chb.2015.08.023.

[30] H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, and L. Ungar, "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, pp. 118-125, 2014.

[31] 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, 1-8. New York, NY: ACM, 2012.

[32] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network." In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 201-213). Springer, Berlin, Heidelberg, 2013.

[33] W. P. Murphy, "Using supervised learning to identify descriptions of personal experiences related to chronic disease on social media.", 2014.

[34] Mourao-Miranda et al., "Patient classification as an outlier detection problem: An application of the One-Class Support Vector Machine", NeuroImage, vol. 58, no. 3, pp. 793-804, 2011. Available: 10.1016/j.neuroimage.2011.06.042.

[35] 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 (pp. 47-56). ACM, 2013.

[36] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, "From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses." In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 1-10), 2015.

[37] 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 (pp. 3187-3196). ACM, 2015.

[38] G. Coppersmith, M. Dredze, C. Harman, and K. Hollingshead, "From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses." In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 1-10), 2015.

[39] M. Nadeem, "Identifying depression on Twitter.", 2016. Available: arXiv preprint arXiv:1607.07384

[40] S. Amir, G. Coppersmith, P. Carvalho, M. J. Silva, and B. C. Wallace, "Quantifying Mental Health from Social Media with Neural User Embeddings.", 2017. Available: arXiv preprint arXiv:1705.00335.

[41] E. Stepanov, S. Lathuiliere, S. A. Chowdhury, A. Ghosh, R. L. Vieriu, N. Sebe, and G. Riccardi, "Depression Severity Estimation from Multiple Modalities.", 2017. Available: arXiv preprint arXiv:1711.06095.

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

[43] A. H. Yazdavar, H. S. Al-Olimat, M. Ebrahimi, G. Bajaj, T. Banerjee, K. Thirunarayan and A. Sheth, "Semi-supervised approach to monitoring clinical depressive symptoms in social media." In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 (pp. 1191-1198), 2017.

[44] A. Benton, M. Mitchell and D. Hovy, "Multi-task learning for mental health using social media text." 2017. Available: arXiv preprint arXiv:1712.03538.

[45] Z. Jamil, "Monitoring Tweets for Depression to Detect At-risk Users" (Doctoral dissertation, Université d'Ottawa/University of Ottawa), 2017.

[46] D. Singh and A. Wang, "Detecting Depression Through Tweets" Standford University CA 9430, pp.1-9, 2018.

[47] J. Weerasinghe, K. Morales and R, Greenstadt, "Because... I was told... so much": Linguistic Indicators of Mental Health Status on Twitter. Proceedings on Privacy Enhancing Technologies, 2019(4), pp. 152-171, 2019.

[48] M. M. Tadesse, H. Lin, B. Xu and L. Yang, "Detection of depression-related posts in reddit social media forum." IEEE Access, 7, 44883-44893, 2019.

[49] M. Trotzek, S. Koitka and C. Friedrich, "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences", IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 3, pp. 588-601, 2020. Available: 10.1109/tkde.2018.2885515.

[50] H. Zogan, X. Wang, S. Jameel and G. Xu. "Depression detection with multi-modalities using a hybrid deep learning model on social media.", 2020. Available: arXiv preprint arXiv:2007.02847.

[51] X. Hu and H. Liu, "Text analytics in social media." In Mining text data (pp. 385-414). Springer, Boston, MA., 2012.

[52] B. Dao, T. Nguyen, S. Venkatesh and D. Phung, "Latent sentiment topic modelling and nonparametric discovery of online mental health-related communities", International Journal of Data Science and Analytics, vol. 4, no. 3, pp. 209-231, 2017. Available: 10.1007/s41060-017-0073-y.

[53] A. Benton, M. Mitchell and D. Hovy, "Multi-task learning for mental health using social media text." 2017. Available: arXiv preprint arXiv:1712.03538.

[54] J. W. Pennebaker, M. E. Francis and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001." Mahway: Lawrence Erlbaum Associates, 71(2001), 2001.

[55] D. Preoţiuc-Pietro, J. Eichstaedt, G. Park, M. Sap, L. Smith, V. Tobolsky and L. Ungar, "The role of personality, age, and gender in tweeting about mental illness." In Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From

linguistic signal to clinical reality(pp. 21-30), 2015.

[56] M. Settanni and D. Marengo, "Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts", Frontiers in Psychology, vol. 6, 2015. Available:10.3389/fpsyg.2015.01045.

[57] W. Yang and L. Mu, "GIS analysis of depression among Twitter users." Applied Geography, 60, 217-223, 2015.

[58] X. Tian, P. Batterham, S. Song, X. Yao and G. Yu, "Characterizing depression issues on sina weibo." International journal of environmental research and public health, 15(4), 764, 2018.

[59] 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 (pp. 47-56). ACM, 2013.

[60] H. A. Schwartz, J. Eichstaedt, M. L. Kern, G. Park, M. Sap, D. Stillwell, and L. Ungar, "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, pp. 118-125, 2014.

[61] C. Fu et al., "Pattern Classification of Sad Facial Processing: Toward the Development of Neurobiological Markers in Depression", Biological Psychiatry, vol. 63, no. 7, pp. 656-662, 2008. Available: 10.1016/j.biopsych.2007.08.020.

[62] V. Leiva and A. Freire, "Towards Suicide Prevention: Early Detection of Depression on Social Media." In International Conference on Internet Science (pp. 428-436). Springer, Cham, 2017.

[63] 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 (pp. 3187-3196). ACM, 2015.

[64] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu and W. Zhu, "Depression detection via harvesting social media: A multimodal dictionary learning solution." In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). pp. 3838-3844, 2017.

[65] S. Bagroy, P. Kumaraguru and M. De Choudhury, "A social media based index of mental well-being in college campuses." In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 1634-1646). ACM, 2017.

[66] M. M. Aldarwish and H. F. Ahmad, "Predicting depression levels using social media posts." In 2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS) (pp. 277-280). IEEE, 2017.

[67] A. Yates, A. Cohan and N. Goharian, "Depression and self-harm risk assessment in online forums.", 2017. Available: arXiv preprint arXiv:1709.01848.

[68] 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 (pp. 2098-2110). ACM, 2016.