Detecting Symptoms of Depression on Reddit

Tingting Liu
National Institute on Drug Abuse
Baltimore, MD, USA
tingting.liu@nih.gov

Devansh Jain
University of Pennsylvania
Philadelphia, PA, USA
devanshrjain7@gmail.com

Shivani Reddy Rapole
University of Pennsylvania
Philadelphia, PA, USA
srapole@seas.upenn.edu

Brenda Curtis
National Institute on Drug Abuse
Baltimore, MD, USA
brenda.curtis@nih.gov

Johannes C. Eichstaedt
Stanford University
Stanford, CA, USA
johannes.stanford@gmail.com

Lyle H. Ungar
University of Pennsylvania
Philadelphia, PA, USA
ungar@cis.upenn.edu

Sharath Chandra Guntuku
University of Pennsylvania
Philadelphia, PA, USA
sharathg@cis.upenn.edu

ABSTRACT

Depression is known to have heterogeneous symptom manifestations. Investigating various symptoms of depression is essential to understanding underlying mechanisms and personalizing treatments. Reddit, an online peer-to-peer social media platform, contains varied communities (subreddits) where individuals discuss their detailed mental health experiences and seek support. The current paper has two aims. The first is to identify psycho-linguistic and open-vocabulary language markers associated with different symptoms using 1,318,749 posts from 43 subreddit communities (e.g., r/bingeeating) clustered into 13 expert-validated depression symptoms (e.g., *disordered eating*). The second aim is to develop prediction models based on the above linguistic features and RoBERTa embeddings to detect specific symptom discourse in contrast to control subreddit posts contributed by the same Reddit users. These predictive models are then validated on a second sample of individuals (N = 2,986) who shared their Facebook posts and completed self-report depression (PHQ-9), anxiety (GAD-7), and loneliness (UCLA-3) surveys.

Based on the differential linguistic patterns that emerged across the various symptoms in our data, we identified three potential clusters, which could also be mapped to the Research Domain Criteria (RDoC) framework. RoBERTa embeddings demonstrated the highest accuracy at predicting most symptoms and were particularly robust at predicting the severity of *suicidal thoughts and attempts*, *self-loathing*, *loneliness*, and *disordered eating*. Our study demonstrates the potential of using large, pseudonymous online forums to train language-based symptom-estimation machine-learning models that can be applied to other text sources. Such technologies could be helpful in clinical psychology, population health, and other areas where early mental health monitoring could improve diagnosis, risk reduction, and treatment.

WebSci '23, April 30–May 01, 2023, Evanston, TX, USA

ACM ISBN 979-8-4007-0089-7/23/04…$15.00
https://doi.org/10.1145/3578503.3583621

CCS CONCEPTS

• **Applied computing** → **Health informatics**; **Psychology**; • **Computing methodologies** → **Natural language processing**.

KEYWORDS

Depression, Symptomology, Reddit, Psycholinguistics, Large Language Models, Heterogeneity

ACM Reference Format:
Tingting Liu, Devansh Jain, Shivani Reddy Rapole, Brenda Curtis, Johannes C. Eichstaedt, Lyle H. Ungar, and Sharath Chandra Guntuku. 2023. Detecting Symptoms of Depression on Reddit. In *15th ACM Web Science Conference 2023 (WebSci '23), April 30–May 01, 2023, Evanston, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3578503.3583621

1 INTRODUCTION

Depression is one of the most prevalent psychiatric disorders worldwide, ranking among the leading causes of disease burden [57]. At least 5% of adults worldwide suffer from depression at some point in their lives [45]. Natural language processing (NLP) of social media data has been utilized to characterize and predict depression [13, 19, 30]. However, prior research in this area has predominantly treated depression as a single syndrome, for instance, represented by a sum-score [23].

Depression is a complex condition with varied clinical presentations and manifestations that can be experienced as distinct individual symptoms [60]. For instance, one study identified 52 individual symptoms of depression, including sadness, suicidal ideation, and fatigue [21]. Understanding the heterogeneity across individual symptoms can provide valuable insights [24]. Profiling the symptoms of depression is essential for exploring causal mechanisms and developing personalized interventions. Specific symptoms of depression may be linked to different risk factors [28] and biomarkers [60], and may also respond differently to antidepressant treatments [8]. For example, a study on 7,500 pairs of twins identified three underlying genetic factors for the nine symptomatic criteria of DSM major depression [37]. Furthermore, particular life events (e.g., romantic breakup) can predict increases in specific symptoms of depression [24]. Such differences across specific symptoms of depression reflect the importance of characterizing each symptom for accurate prediction.

Recent studies of depression using social media language have investigated specific symptoms [64, 65]. Despite this progress, certain challenges remain - first, studies often lack solid validation for the classification models (e.g., clinical assessments) or rely solely on dimensions from one inventory, typically the Patient Health Questionnaire (PHQ-9 [38]). Although the PHQ-9 is based on DSM criteria [38], with anhedonia and depressed mood being core symptoms, depression often presents in various ways beyond the nine symptoms in PHQ-9 [22]. For example, many scales overlook somatic symptoms, which can lead to significant distress and influence depression diagnosis [24]. Thus, an additional challenge is to consider more relevant symptoms instead of translating directly from the DSM-5 criteria when characterizing depression. The third challenge is the high cost of collecting self-reported symptom-level depression data to build language-based prediction models. To achieve the necessary statistical power, large sample sizes are often required [18]. Given the complexity of depression symptoms and measurements [22], advancing precision prediction of depression at the symptom level at low costs is crucial.

In this paper, we aim to bridge the gaps in recent studies of depression using social media language by analyzing the heterogeneous symptoms of depression from Reddit and validating our predictions using additional Facebook language data from out-of-sample user data with self-reported depression, anxiety, and loneliness assessments. To provide clinically valuable insights for depression assessment at the symptom level [47], we summarize and validate the 13 symptoms from the seven most commonly used measurements of depression [21], with the help of clinical experts (see Table 1 for the list of symptoms). Reddit is an apt data source for this study due to its unique affordances, including anonymity and sub-communities where individuals discuss their mental health experiences and seek support. For an overview of our study, please refer to Figure 1.

The major contributions of this paper are threefold:

- We identify the differential linguistic markers for 13 expert-validated symptoms of depression using Reddit posts.
- We showcase the ability of Reddit posts to predict specific symptoms of depression across various linguistic models.
- We validate the predictive models trained on Reddit on out-of-sample individuals who consented to share self-report surveys and Facebook posts.

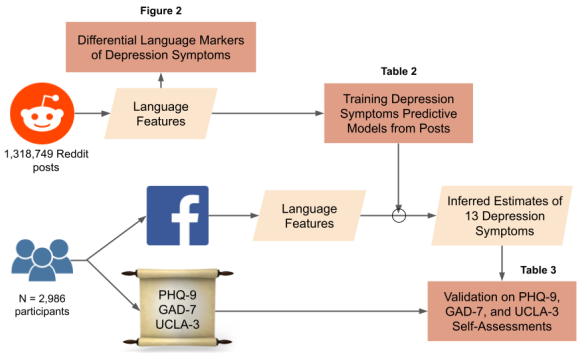

Figure 1: Overview of the Approach Taken to Study Language Markers of Depression Symptoms.

2 RELATED WORK

2.1 Depression and Social Media Language

Social media language has been found to reflect individuals' daily lives and mental status, providing valuable insights into mental health predictions [40]. Among all mental health conditions, depression is most commonly researched using social media [10]. Past studies have successfully used social media language to predict depression (see reviews in [26, 30, 34, 55]). For example, one of the most cited papers extracted linguistic features from Twitter to analyze depression [13] and found signals for characterizing the onset of depression (e.g., increased negative emotions). The most commonly used linguistic features or variables in predictive analyses include the distribution of words and phrases, the syntactic composition of posts (e.g., length of posts), psycholinguistic categories from the Linguistic Inquiry and Word Count (LIWC) dictionaries [48], Latent Dirichlet Allocation (LDA) topics [7], and domain-specific lexicons (see reviews in [10, 53]). These features are typically used to correlate and classify depression status obtained through self-assessments, self-disclosure, or forum membership [30]. Although some studies have evaluated the predictive ability of language posted on Reddit about multiple mental health conditions [12, 36], specific symptoms of depression have not been examined. Common self-assessments in these analyses include the PHQ-9 [38], the Center for Epidemiologic Studies Depression Scale Revised (CES-D [17]), and Beck's Depression Inventory (BDI, [6]). Furthermore, most of these predictions rely either on a single sum-score from self-assessed surveys or one self-reported depression status [30], with Pearson r ranging between 0.15 and 0.35, or often fall into a binary classification (depression versus not depression with Area Under the Receiver Operating Characteristic Curve (AUC) values around 0.7 [1]).

2.2 Depression Symptoms Prediction

Few studies analyzed depression at the symptom level using social media language. For example, [64] used a semi-supervised statistical model to profile depression using texts (i.e., tweets) from Twitter to emulate the nine symptoms from the PHQ-9 questionnaire (e.g., lack of interest). Their results also found overlapping LDA topics (sacrificing specificity) across symptoms of depression due to inherent comorbidity. A similar approach was also proposed to match tweets to their corresponding PHQ-9 categories in [39]. However, validation of machine learning models in most symptom-level studies relies on expert annotations rather than self-assessments of lived experiences. Though the classic survey like PHQ-9 could reflect the DSM classifications of depression to some extent, using only one scale may be insufficient for profiling symptoms of depression due to well-known differences across scales [22, 24]. Therefore, in this study, we identify differential language markers of various depression symptoms on Reddit and validate them on an independent sample of Facebook users who provided a diverse array of self-assessments on their symptoms.

Table 1: Descriptives of the depression symptoms data collected in this study

Symptom	#posts	#users	Subreddits
Control	338,059	102,884	subreddits in control data from Gkotsis (2017), e.g., AskReddit, trees
Anger	5,465	4,204	Anger
Anhedonia	6,167	4,942	anhedonia, DeadBedrooms
Anxiety	223,952	114,060	Anxiety, AnxietyDepression, HealthAnxiety, PanicAttack
Concentration deficit	11,756	9,545	DecisionMaking, shouldi
Disordered eating	51,398	15,251	bingeeating, BingeEatingDisorder, EatingDisorders, eating_disorders, EDAnonymous
Fatigue	5,396	4,360	chronicfatigue, Fatigue
Loneliness	96,400	35,737	ForeverAlone, lonely
Sad mood	75,886	56,873	cry, grief, sad, Sadness
Self-loathing	60,526	32,582	AvPD, SelfHate, selfhelp, socialanxiety, whatsbotheringyou
Sleep problem	31,766	22,908	insomnia, sleep
Somatic complaint	55,399	26,082	cfs, ChronicPain, Constipation, EssentialTremor, headaches, ibs, tinnitus
Suicidal thoughts and attempts	288,443	152,809	AdultSelfHarm, selfharm, SuicideWatch
Worthlessness	68,136	52,077	Guilt, Pessimism, selfhelp, whatsbotheringyou

2.3 Mental Health and Reddit

Reddit, a popular semi-anonymous online discussion forum, contains different communities (i.e., subreddits). Reddit is widely used to discuss mental health concerns and seek community support. Previous studies have characterized the language of health conditions using Reddit including mental health [27] and physical health [61]. For example, [25] mapped subreddits to the best-matching DSM-5 category using a multiclass classifier. In the 54 papers reviewed by [9] on depression and anxiety investigation using Reddit, approximately two-thirds of studies used language extracted from Reddit as the basis for predictive mental health classifications. Apart from identifying linguistic markers of specific conditions, Reddit has also been used to predict changes in mental health status. For instance, [14] used posts from the same users on Reddit to predict future suicidal ideation. They studied the shifts from mental health-related discourse to suicide-related posts and it was found that linguistic cues about esteem and network support were linked to reducing such a shift.

3 METHODS

3.1 Data

3.1.1 Depression Symptoms. In this study, 13 symptoms of depression (see in Table 1) were derived from a larger set of 52 subsymptoms compiled from the seven most frequently used depression assessments in psychology and psychiatry literature, such as the Beck Depression Inventory (BDI-II; [5]), the Hamilton Rating Scale for Depression (HRSD; [31]), and Center of Epidemiological Scales (CES-D; [17]). A trained psychologist categorized the 52 symptoms based on their similarities and comorbidities, resulting in the 13 symptoms used in this study, which two clinical experts further validated.

3.1.2 Reddit. We used the PushShift Reddit dataset [4], which includes posts from January 2010 to December 2019, to identify subreddits discussing different symptoms. Based on prior works [15, 27], we categorized the depression-related subreddits to our 13 symptoms labels and expanded the list to include subreddits that used the names of the 13 symptoms we selected (e.g., r/Anger for the *anger* symptom). We then assessed a random set of 50 posts to confirm the relevance of the discussions to the symptom.

Following prior work [27], we created a control dataset of the same users, consisting of posts from non-depression-related subreddits. We collected 1,318,749 Reddit posts across the 13 depression symptoms and the control dataset. We excluded comments on posts to limit the scope of the data to describe the experience of symptoms, as comments often contain a mix of offering support while discussing symptoms. The study was deemed exempt by the University of Pennsylvania Institutional Review Board.

Descriptive statistics on the dataset at the symptom level are provided in Table 1, including the subreddits used to collect posts for each symptom. Additionally, *anhedonia*, *concentration deficit*, *fatigue*, *sad mood*, and *worthlessness* had relatively few posts from the listed subreddits. We gathered extra posts pertaining to these symptoms by conducting a keyword search using the name and core descriptions of each symptom within the subreddits, including r/AskReddit, r/lonely, r/CasualConversation, r/SeriousConversation, r/depression, r/self, r/ADHD, r/Advice, and r/NoStupidQuestions. Finally, we excluded posts with fewer than 10 words from our analyses.

3.1.3 Facebook: We obtained Facebook status updates from a sample of 2,986 individuals who consented to share their Facebook data and responded to depression, anxiety, and loneliness self-assessment survey instruments. Depression was assessed using the PHQ-9 [38], anxiety was assessed by General Anxiety Disorder-7 (GAD-7) [58], and loneliness was assessed by the UCLA-3 item [32]. We calculated sum scores for each scale as the user's depression, anxiety, and loneliness scores. Individuals were enrolled through the Qualtrics Panel, an online crowdsourcing platform for research participants' recruitment. Of these individuals, 69.7% identified as female and the mean age was 43 yrs (SD: 12), and 63.8% had a Bachelor's degree or higher. This data analysis was exempted by the University of Pennsylvania Institutional Review Board, see more details about this data in [41]. We used this Facebook data for validating the models built on Reddit data.

3.2 Language Features

We tokenized every post using *happierfuntokenizer* from the DLATK Python library [56] and masked specific tokens, such as URLs into meta tags. We then extracted four different features that were shown in prior literature [15, 30, 43] to have significant associations with mental health: a) psycho-linguistic lexicon - Linguistic Inquiry and Word Count 2015 (LIWC) [48], b) open-vocabulary topics generated from latent Dirichlet allocation (LDA) [7], and c) RoBERTa embeddings [42].

3.2.1 LIWC:. We utilized the expert-curated LIWC lexicon which consists of different psycholinguistic categories and associated words for each category. We counted the tokens in each post that match the tokens in the LIWC dictionary [49] and then normalized the counts by the number of words in the post to obtain the relative frequency of each LIWC category.

3.2.2 Reddit Topics: After removing the top 100 most frequent words in the Reddit dataset as stop words, we applied LDA [7] to identify topics in the posts. LDA is a probabilistic generative model that assumes posts are generated by a combination of topics, and topics are distributions of words. Given that the words within a post are known, it is possible to estimate topics as latent variables using Gibbs sampling. We used the MALLET implementation of LDA [44] to generate 200 topics, with an alpha level of 5. We obtained the distribution of each topic for all posts in the Reddit dataset.

3.2.3 RoBERTa embedding. We used RoBERTA [42], a pre-trained contextual word embedding model, to generate numeric vector representations of posts' language. RoBERTa embedding of each word depends on the other words used near it, so these models capture contextual semantic information in a language unlike other co-occurrence or count-based methods. To calculate post embeddings, we averaged representations for each word in the post, where a word is represented by its 10^{th} layer in the RoBERTa model. We used *roberta-base* from transformers [63] to compute embeddings.

3.2.4 Facebook Topics: Using a similar process for generating Reddit topics, we generated 500 topics on the Facebook dataset. We then obtained the probability distributions of the 500 topics for every user in the Facebook dataset. We also obtained the distribution of Facebook topics for every Reddit post to build a symptom-level predictive model. Prior work has found that LDA topics are sensitive to the domain in which they are trained and applied [3, 11]; hence, we also considered the Facebook topics as a candidate feature set in validating the Reddit models on PHQ-9, GAD-7, and UCLA-3 self-assessments from the Facebook dataset.

3.3 Differential Language Markers of Depression Symptoms

After we obtained the language features in the above section, we performed a differential language analysis to identify statistically significant correlations between the features and the symptom labels. We designed this as a post-level analysis and used two sets of language features (LIWC and Reddit topics, respectively) as independent variables in the logistic regression model to predict the symptom labels of each post. This was set up as a one-vs-all classification task where, for instance, when obtaining language markers for *anger*, posts in the subreddits belonging to *anger* were given the outcome of 1, and posts in all other subreddits (Control + other depression symptoms) were given the dummy outcome of 0. These dummy outcomes were the dependent variables when training the logistic regression models. In this analysis, we wanted to study the specific language markers of each symptom compared to the control + all other symptoms. Based on conventional linguistic analysis, we used a p-value of < 0.01 to identify significant linguistic markers, and all p-values were corrected for the false discovery rate during multiple hypothesis testing using the Bonferroni correction.

3.4 Prediction Models for Depression Symptoms

Considering the dimensionality of our feature set, we trained Random Forest classifiers on Reddit language to predict depression symptoms. We treated each linguistic feature set (LIWC, Reddit Topics, RoBERTa embeddings, and Facebook Topics) as an independent variable and treated the exact symptom labels as dependent variables in predictive models. Each feature set was considered independently to enable a comparative analysis of their predictive performance. In the predictive analyses, we trained two models: 1) symptom vs. control, and 2) symptom vs control + all other symptoms. We provide the results from the second set of models (aimed at higher specificity) in the Appendix. The models were evaluated using 5-fold cross-validation, and we report AUC on the test set estimates from the 5-folds.

3.5 Validation on PHQ-9, GAD-7, and UCLA-3 self-assessments

We applied the Reddit models to language features extracted from a different sample of Facebook users who took the depression (PHQ-9), anxiety (GAD-7), and loneliness (UCLA-3) survey assessments to see if they could generalize to out-of-sample, out-of-platform, and a different source-label dataset. We included anxiety and loneliness survey assessments here for two reasons. First, anxiety and loneliness are often considered sub-symptoms of depression [20, 59], and we can fully capture the variance in the manifestations of depression symptoms by including them in validation. Second, general anxiety and loneliness are also comorbid with depression [16, 20], which could generate substantial impairment in functioning for the depression diagnosis.

As discussed in the next section, Reddit topics and RoBERTa obtained the best performance in the within-sample cross-validation analyses. Therefore, we extracted these feature sets at the user level from Facebook data. We then trained a classifier using each linguistic feature set (LIWC, Reddit Topics, Facebook Topics, and RoBERTa embeddings) as an independent variable and distantly supervised symptom labels (i.e., the exact symptom label from Reddit data) as a dependent variable.

After the Reddit classifiers were trained, we applied them to the Facebook dataset to generate user-level predictions of depression symptoms, which provided inferred estimates (probabilities from the classification models) of 13 depression symptoms. To validate the predictive utility of these estimates, we assessed their performance using Spearman correlations. We correlate the probabilities

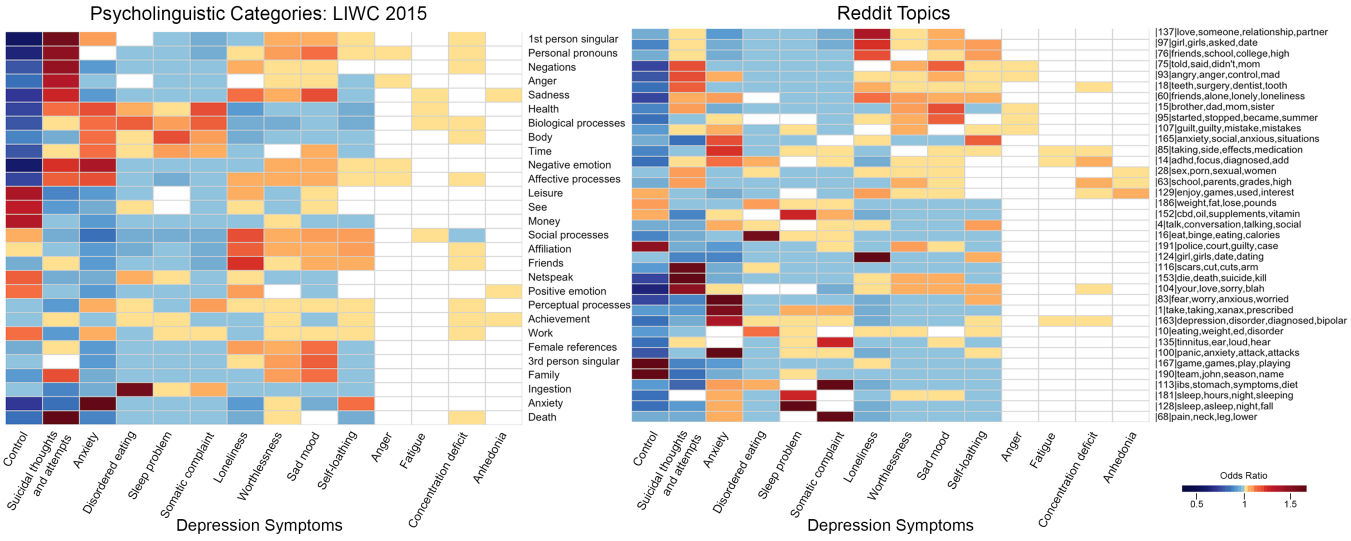

Figure 2: Differential Language Markers Associated with Depression Symptoms. The top 3 LIWC 2015 categories and the top 3 Reddit LDA topics that are positively correlated with each depression symptom in the differential language analysis are shown in the heatmaps. For each topic, the top 4 examples of words are shown after the topic ID. The darker the color, the larger the odds. Red: odds ratio > 1, more likely to use this linguistic feature; blue: odds ratio < 1, less likelihood of linguistic feature. Only odds ratios that are significant at $p < 0.01$ with Bonferroni correction are presented.

from the predictive models with users' self-report depression, anxiety, and loneliness scores at the user level. We used a p-value of < 0.05 as the significance level for this analysis.

4 RESULTS

4.1 Differential Language Markers of Depression Symptoms

As illustrated in Figure 2, we identified different linguistic markers across symptoms from LIWC and Reddit Topics. The symptom of *suicidal thoughts and attempts* (hereinafter referred to as suicide) is strongly associated with LIWC Death (OR: 3.602), and topics about self-harm ("scars," "cuts," OR: 1.662) and death ("die," "death," "suicide"; OR: 1.762). *Anxiety* is marked by LIWC Anxiety (OR: 3.022) and topics about panic and anxiety ("panic," "anxiety," "attack"; OR: 1.9461). *Disordered eating* is positively linked to LIWC Ingestion (OR: 1.617) and binge eating-related topics ("eat," "binge," "calories"; OR: 1.617). *Loneliness* is positively linked to LIWC Friends (OR: 1.230) and dating-related topics. Unlike the above symptoms, *anger, anhedonia, concentration deficit,* and *fatigue* do not show clear profiles of language associations.

Our analysis reveals that it is possible to identify clusters of symptoms based on their language profiles. Interestingly, most of these symptom clusters could be mapped to the five domains of the Research Domain Criteria (RDoc) framework [33]. This framework aims to organize biological and behavioral manifestations of mental health across different domains, as shown in the Appendix Table A1.

The clearest cluster contained ***disordered eating, sleep problem, and somatic complaints***. These three symptoms exhibit similar language profiles, as they are all associated with a higher likelihood of using LIWC categories such as ingestion, body, health, and time. As might be expected, each of these three symptoms is also marked by direct symptom mentions. For example, *sleep problem* is linked to topics about sleep ("sleep," "night"; OR: 2.359), and sleep-related ingredients and supplements ("cbd," "oil"; OR: 1.291). *Somatic complaint* is associated with LIWC perceptual processes (OR: 1.077) and topics such as stomach-related illness ("ibs," "stomach"; OR: 3.371), other physical symptoms related to hearing ("tinnitus," "ear"; OR: 1.289) and body pain ("pain," "neck," "leg"; OR: 1.739). It is worth noting that some hearing-related physical diseases are linked to affective disorders in the literature (e.g., Tinnitus [54]). The language profiles of these three symptoms also exhibit some differences. While *somatic complaint* is positively associated with sleep-related language (e.g., "sleep," "asleep," "night"; OR: 1.036), *disordered eating* shows a negative association with such language. Additionally, *disordered eating* is less likely to discuss topics related to panic and anxiety medications (e.g., alprazolam or "xanax"; OR: 0.978).

The symptoms of ***suicide*** and ***anxiety*** exhibit similar language patterns. Both are associated with increased use of language reflecting negative emotions, health, and friends. Additionally, they are linked to decreased use of language related to social processes, affiliation, ingestion, leisure activities, and personal interests (e.g., games). Apart from the language related to death and harm mentioned earlier, the language associated with *suicide* also includes anger-related words (LIWC Anger, OR: 1.342; "angry," "control," "mad," OR: 1.191), negations (OR: 1.481), 1st person singular pronouns (OR: 1.640), and sadness (OR: 1.264). Furthermore, individuals exhibiting suicidal tendencies are more likely to discuss feeling sorry about their loved ones ("you," "love," "sorry"; OR: 1.515).

Table 2: Symptoms vs. Control: 5-fold Cross Validation Prediction Performance (AUC) for Different Reddit Models

Symptom	LIWC	Reddit Topics	RoBERTa	FB Topics
Anger	0.934	0.949	**0.971**	0.951
Anhedonia	0.864	0.932	**0.971**	0.929
Anxiety	0.912	0.932	**0.940**	0.928
Concentration deficit	0.874	0.937	**0.969**	0.954
Disordered eating	0.903	0.949	**0.958**	0.937
Fatigue	0.911	0.927	**0.963**	0.946
Loneliness	0.843	0.901	**0.918**	0.885
Sad mood	0.902	0.903	**0.941**	0.933
Self-loathing	0.886	0.923	**0.940**	0.922
Sleep problem	0.928	**0.982**	0.980	0.979
Somatic complaint	0.886	**0.960**	0.953	0.946
Suicidal thoughts and attempts	0.901	0.941	**0.957**	0.935
Worthlessness	0.863	0.895	**0.949**	0.942

Note: FB = Facebook. Model with best AUC for each symptom in **bold**.

When comparing these two symptoms, it is important to note that *anxiety* is associated with more language about anxiety (as mentioned earlier), body (OR: 1.137), work (OR: 1.035), and topics related to fear ("fear," "worry"; OR: 2.686), while *suicide* is negatively associated with them. On the other hand, *anxiety* is associated with less language about negations (OR: 0.905), death (OR: 0.863), and sadness (OR: 0.901), whereas *suicide* is more likely to use or mention these language markers.

Loneliness, worthlessness, sad mood, and ***self-loathing*** are four symptoms share similar language patterns. They share a tendency to discuss topics related to negative emotions and social aspects (e.g., affiliation), while being less likely to mention death, sleep, anxiety, pain, and medication.

It is important to note that although the four symptoms mentioned above include discussions about eating, they tend to use language from the LIWC ingestion category less frequently (OR: [0.955, 0.971]), which covers a wider range of vocabulary. *Loneliness* is more closely linked to discussions about dating and relationships ("girls," "date"; OR: 1.937), which may be related to the demographics of Reddit users (e.g., young men). Unlike the other three symptoms, *sad mood* is less associated with the language of achievement (OR: 0.988), and *worthlessness* is uniquely more linked to language about money (OR: 1.021).

4.2 Prediction Models for Depression Symptoms

Table 2 and Appendix Table A2 show the within-sample cross-validation results of Random Forest models trained using three feature sets: LIWC, Reddit Topics, and RoBERTa for identifying the symptoms of Reddit posts. Table 2 shows the results of each symptom against control, and Table A2 shows the results of each symptom against all other symptoms + control.

The highest AUC across three models for the same symptom is presented in bold. In Table 2, we see that the model utilizing RoBERTa embeddings demonstrates the highest accuracy for most symptoms, except for *sleep problem* and *somatic complaint*. Additionally, it also achieves the highest accuracy for the two core symptoms in DSM-5–*anhedonia* and *sad mood*, indicating the superior predictive power compared to all three models. The model using Reddit topics has the largest accuracy (.982 for *sleep problem*).

Table 3: Symptoms vs. Control: Validation Correlations with Self-Reported Assessments of Depression, Anxiety, Loneliness

Symptom	Depression	Anxiety	Loneliness
Anger	0.152	0.105	0.12
Anhedonia	0.072	0.052	0.062
Anxiety	0.197	0.156	0.136
Concentration deficit	0.053	0.043	0.056
Disordered eating	0.214	0.206	0.103
Fatigue	0.029	0.003	0.045
Loneliness	0.206	0.174	0.134
Sad mood	0.135	0.112	0.089
Self-loathing	0.236	0.204	0.154
Sleep problem	0.194	0.157	0.129
Somatic complaint	0.092	0.043	0.059
Suicidal thoughts and attempts	0.234	0.188	0.153
Worthlessness	0.062	0.043	0.077

Note: All Spearman correlations are significant at $p < 0.05$, except grayed ones.

4.3 Validation on PHQ-9, GAD-7, and UCLA-3 self-assessments

As shown in table 3, most of the language-predicted symptoms were significantly associated with self-reported mental health surveys–except *fatigue* (the non-significant coefficients at p-value of 0.05 are grayed out). Overall, the symptom estimates from linguistic features extracted using RoBERTa embeddings have the strongest correlations with the survey scores. Because most of the symptom estimates reached the highest accuracy when estimated using RoBERTa embeddings, we only report the RoBERTa results in Table 3.

Language-estimated *suicide* ($\rho = 0.234$), *self-loathing* ($\rho = 0.236$), *disordered eating* ($\rho = 0.214$), and *loneliness* ($\rho = 0.206$) show the strongest correlations with self-reported depression (all $p < 0.001$). This is consistent with these symptoms being among the strongest language markers on the Reddit data. The correlations between language-estimated *fatigue* and self-reported depression and anxiety scores are insignificant.

Similar to our findings with language markers, language-based *anhedonia*, *worthlessness*, and *concentration deficit* have weaker correlations with all three self-reported scales. The correlations between *somatic complaint* and all three self-reported scales are weaker, particularly when correlating with anxiety, although they remain significant. In contrast to the challenges in identifying markers of *anger* on Reddit, we found robust positive correlations between the estimated *anger* using RoBERTa and self-reported depression, anxiety, and loneliness on this Facebook dataset.

To evaluate the robustness of the language-based symptom predictions compared to all the self-reported symptoms measured by the PHQ, we conducted a correlation analysis between the RoBERTa-estimated symptoms and the nine individual items of the PHQ-9, as shown in Table 4. We found that *suicide*, *self-loathing*, *loneliness*, and *disordered eating* have robust, significant correlations with all nine items (all $\rho > 0.08$). Moreover, consistent with findings in the above sections, *anhedonia, concentration deficit, fatigue*, and *worthlessness* have weaker or non-significant correlations across all items.

Table 4: Correlations between Symptoms Predicted by the Reddit RoBERTa Embeddings model and Individual PHQ-9 Items from Facebook dataset

Symptom	PHQ 1 (little interest)	PHQ 2 (feel depressed)	PHQ 3 (sleep issues)	PHQ 4 (feel tired)	PHQ 5 (appetite)	PHQ 6 (self)	PHQ 7 (concen--tration)	PHQ 8 (slow movement)	PHQ 9 (death)
Anger	0.11	0.138	0.121	0.136	0.136	0.113	0.085	0.064	0.102
Anhedonia	0.023	0.067	0.082	0.096	0.068	0.051	0.028	-0.047	-0.012
Anxiety	0.137	0.159	0.164	0.189	0.179	0.137	0.13	0.043	0.046
Concentration deficit	0.007	0.045	0.065	0.081	0.072	0.024	0.016	-0.069	-0.036
Disordered eating	0.15	0.169	0.162	0.186	0.194	0.151	0.175	0.081	0.065
Fatigue	-0.008	0.038	0.056	0.062	0.047	0.009	-0.012	-0.081	-0.042
Loneliness	0.166	0.158	0.142	0.153	0.173	0.17	0.156	0.147	0.13
Sad mood	0.083	0.097	0.132	0.154	0.122	0.096	0.075	0.022	0.01
Self-loathing	0.167	0.185	0.17	0.209	0.207	0.173	0.177	0.103	0.095
Sleep problem	0.138	0.171	0.144	0.168	0.164	0.153	0.146	0.067	0.088
Somatic complaint	0.056	0.072	0.098	0.098	0.112	0.048	0.046	-0.026	-0.01
Suicidal thoughts and attempts	0.196	0.189	0.178	0.185	0.186	0.183	0.158	0.158	0.146
Worthlessness	0.023	0.073	0.079	0.078	0.071	0.037	0.015	-0.052	0.002

Note: All correlations are Spearman correlations. Non-significant correlations are grayed out. All other correlations are significant at $p < 0.05$.

Even though we did not include a specific language-estimated symptom to reflect slow movement (PHQ item 8), we observed that this item correlated with many symptoms but in opposite directions (e.g., positively correlated with *suicide*, ρ = 0.158, but negatively correlated with *fatigue*, ρ = -0.081). The four negative correlations found with PHQ 8 are associated with symptoms with less distinct language markers, namely *anhedonia, concentration deficit, fatigue*, and *worthlessness*. In addition, we observe that the largest correlations for many PHQ items do not stem from their corresponding symptom language estimates. For example, PHQ 3 (sleep issues) has the strongest correlations with language-estimated *suicide* (ρ = 0.178), instead of the language-estimated *sleep problem* (ρ = 0.144). This indicates that language-based depression symptom prediction from Reddit may generalize across symptoms, but at the cost of specificity.

5 DISCUSSION

The current paper (1) determined the language profiles of 13 expert-validated symptoms of depression on Reddit, (2) evaluated the predictive performance of four different linguistic feature sets (LIWC, Reddit Topics, RoBERTa embeddings, and Facebook Topics), and (3) validated the generalizability of Reddit language in predicting depression (PHQ-9), anxiety (GAD-7), and loneliness (UCLA-3) survey assessments on an out-of-sample dataset of 2,986 Facebook users. Additionally, we conducted a comprehensive assessment of the symptom prediction models against the PHQ symptoms.

One significant finding from our study is that *suicidal ideation, self-loathing, disordered eating*, and *loneliness* have more robust language markers on Reddit than other symptoms. This pattern is consistent in the subsequent predictions and validations using Facebook data. This pattern can be explained from two perspectives.

First, from a theoretical perspective, these symptoms have well-documented individual manifestations, and the links between these symptoms and the severity of depression are strong. Depression is often associated with a high suicide risk rate (e.g., around 15%; [46]) and is highly correlated with eating disorders and loneliness [2, 62]. Depression has also been linked to increased self-focus and self-criticism [51]. Anxiety, which has its own unique manifestations and is highly correlated with depression, also demonstrated clear language patterns in our findings.

Second, from a measurement perspective, these symptoms may be more easily detected through language use on social media. Past investigations of depression-related language features have identified top linguistic features associated with these symptoms [26, 30, 34, 55]. For example, depression has been linked to higher use of the first-person singular [13]. Additionally, eating-related topics are among the largest topics derived from depression-related Twitter language [52].

Another key finding is that Reddit provides a cost-effective method for detecting depression symptoms accurately. Language features extracted from Reddit show high accuracy in our prediction and validation analyses. Such high accuracy can be attributed to the use of language features extracted from Reddit posts, which are analyzed at the post level to predict the forum membership of each post, and are in line with [27]. Our study is unique compared to past research analyzing Reddit language, as clinical experts validated our depression symptoms. Moreover, while not all symptoms were easily detected in language, over half of the symptoms in our study demonstrated satisfactory validity on a different dataset with self-reported item-level and instrument-level measures.

5.1 Limitations

One of the limitations of this study, which can also be considered a strength in terms of rigor, is that we applied validation to a different sample across platforms (i.e., from Reddit to Facebook). This approach reduces the transfer learning capability but increases the external validity of our results. We did not explore domain adaptation techniques that could yield higher correlation coefficients on the target domain, as seen in prior works [29, 35]. Second, in the validation study, we only correlated with individual items from

the PHQ-9 which measures 9 symptoms of depression. To better capture the multidimensional nature of depression, future studies should consider additional scales that capture a wider range of symptom dimensions. Third, due to the limited number of posts on Reddit for some symptoms (e.g., *anger*, *worthlessness*), we added posts generated from keyword searches to increase the data size. This may have introduced noise into our language-based estimates, even though the AUCs for prediction are high and comparable to the other symptoms. Especially for four symptoms, *fatigue*, *concentration deficit*, *anhedonia*, and *worthlessness*, it was challenging to collect sufficient language data from Reddit to identify their clear language markers. Future research could investigate the language manifestations of these symptoms further.

5.2 Ethical Considerations

The current study is a secondary data analysis of de-identified Reddit and Facebook data. The Facebook language data were collected from human subjects and approved by the Institutional Review Board (IRB). Analysis of social media language and mental health conditions needs careful ethical considerations. First, individuals' social media posts are highly sensitive and can contain Personal Identifiable Information (PII). This data set was carefully collected and stored in a secure location and could only be accessed by researchers on relevant projects after undergoing human subjects training and being added to the IRB protocols. On the one hand, it can inform the development of interventions and reduce the risks associated with mental health disorders. On the other hand, it can create biases and propagate stigma toward individuals with these disorders if used improperly. Further, while the participants on Reddit accepted the platform ToS that the data is publicly available did not explicitly consent to their data being used for research purposes [50]. Therefore, using social media language-based mental health indicators should involve inputs from multi-disciplinary stakeholders, including participants, computer scientists, psychologists, medical professionals, ethicists, and others. This would ensure the approach is ethical and respectful of individual privacy and mental health.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institute On Minority Health And Health Disparities of the National Institutes of Health under Award Number R01MD018340, and was funded by the Intramural Research Program of the National Institutes of Health (NIH), National Institute on Drug Abuse (NIDA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] Norah Saleh Alghamdi, Hanan A Hosni Mahmoud, Ajith Abraham, Samar Awadh Alanazi, and Laura García-Hernández. 2020. Predicting depression symptoms in an Arabic psychological forum. *IEEE Access* 8 (2020), 57317–57334.

[2] Daniele Marano Rocha Araujo, Giovana Fonseca da Silva Santos, and Antonio Egídio Nardi. 2010. Binge eating disorder and depression: a systematic review. *The world journal of biological psychiatry* 11, 2-2 (2010), 199–207.

[3] Yang Bao, Nigel Collier, and Anindya Datta. 2013. A Partially Supervised Cross-Collection Topic Model for Cross-Domain Text Classification. In *Proceedings of the 22nd ACM International Conference on Information Knowledge Management* (San Francisco, California, USA) *(CIKM '13)*. Association for Computing Machinery, New York, NY, USA, 239–248. https://doi.org/10.1145/2505515.2505556

[4] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 830–839.

[5] Aaron T Beck, Robert A Steer, and Gregory Brown. 1996. Beck depression inventory–II. *Psychological assessment* (1996).

[6] Aaron T Beck, Robert A Steer, and Margery G Carbin. 1988. Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical psychology review* 8, 1 (1988), 77–100.

[7] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[8] Sarah R Blumenthal, Victor M Castro, Caitlin C Clements, Hannah R Rosenfield, Shawn N Murphy, Maurizio Fava, Jeffrey B Weilburg, Jane L Erb, Susanne E Churchill, Isaac S Kohane, et al. 2014. An electronic health records study of long-term weight gain following antidepressant use. *JAMA psychiatry* 71, 8 (2014), 889–896.

[9] Nick Boettcher et al. 2021. Studies of depression and anxiety using reddit as a data source: Scoping review. *JMIR mental health* 8, 11 (2021), e29487.

[10] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 1–11.

[11] Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Leveraging Multi-Domain Prior Knowledge in Topic Models. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence* (Beijing, China) *(IJCAI '13)*. AAAI Press, 2071–2077.

[12] Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. 2018. SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. *arXiv preprint arXiv:1806.05258* (2018).

[13] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

[14] Munmun De Choudhury and Emre Kiciman. 2017. The Language of Social Support in Social Media and Its Effect on Suicidal Ideation Risk. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017), 32–41. https://doi.org/10.1609/icwsm.v11i1.14891

[15] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2098–2110.

[16] Keith S Dobson. 1985. The relationship between anxiety and depression. *Clinical Psychology Review* 5, 4 (1985), 307–324.

[17] William W Eaton, Corey Smith, Michele Ybarra, Carles Muntaner, and Allen Tien. 2004. Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R). (2004).

[18] Johannes C Eichstaedt, Margaret L Kern, David B Yaden, HA Schwartz, Salvatore Giorgi, Gregory Park, Courtney A Hagan, Victoria A Tobolsky, Laura K Smith, Anneke Buffone, et al. 2021. Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods* 26, 4 (2021), 398.

[19] Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences* 115, 44 (2018), 11203–11208.

[20] Evren Erzen and Özkan Çikrikci. 2018. The effect of loneliness on depression: A meta-analysis. *International Journal of Social Psychiatry* 64, 5 (2018), 427–435.

[21] Eiko I Fried. 2017. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of affective disorders* 208 (2017), 191–197.

[22] Eiko I Fried, Jessica K Flake, and Donald J Robinaugh. 2022. Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology* (2022), 1–11.

[23] Eiko I Fried and Randolph M Nesse. 2015. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. *Journal of affective disorders* 172 (2015), 96–102.

[24] Eiko I Fried and Randolph M Nesse. 2015. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC medicine* 13, 1 (2015), 1–11.

[25] Manas Gaur, Ugur Kursuncu, Amanuel Alambo, A. Sheth, Raminta Daniulaityte, Krishnaprasad Thirunarayan, and Jyotishman Pathak. 2018. "Let Me Tell You About Your Mental Health!": Contextualized Classification of Reddit Posts to DSM-5 for Web-based Intervention. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (2018).

[26] Felipe T Giuntini, Mirela T Cazzolato, Maria de Jesus Dutra dos Reis, Andrew T Campbell, Agma JM Traina, and Jo Ueyama. 2020. A review on recognizing depression in social networks: challenges and opportunities. *Journal of Ambient Intelligence and Humanized Computing* 11, 11 (2020), 4713–4729.

[27] George Gkotsis, Anika Oellrich, Sumithra Velupillai, Maria Liakata, Tim JP Hubbard, Richard JB Dobson, and Rina Dutta. 2017. Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific reports* 7, 1 (2017), 1–11.

[28] Robin N Groen, Evelien Snippe, Laura F Bringmann, Claudia JP Simons, Jessica A Hartmann, Elisabeth H Bos, and Marieke Wichers. 2019. Capturing the risk of persisting depressive symptoms: A dynamic network investigation of patients' daily symptom experiences. *Psychiatry Research* 271 (2019), 640–648.

[29] Sharath Chandra Guntuku, Anneke Buffone, Kokil Jaidka, Johannes C Eichstaedt, and Lyle H Ungar. 2019. Understanding and measuring psychological stress using social media. In *Proceedings of the international AAAI conference on web and social media*, Vol. 13. 214–225.

[30] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* 18 (2017), 43–49.

[31] Max Hamilton. 1960. A RATING SCALE FOR DEPRESSION. *Journal of Neurology, Neurosurgery & Psychiatry* 23, 1 (1960), 56–62. https://doi.org/10.1136/jnnp.23.1.56 arXiv:https://jnnp.bmj.com/content/23/1/56.full.pdf

[32] Mary Elizabeth Hughes, Linda J Waite, Louise C Hawkley, and John T Cacioppo. 2004. A short scale for measuring loneliness in large surveys: Results from two population-based studies. *Research on aging* 26, 6 (2004), 655–672.

[33] Thomas Insel, Bruce Cuthbert, Marjorie Garvey, Robert Heinssen, Daniel S Pine, Kevin Quinn, Charles Sanislow, and Philip Wang. 2010. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. , 748–751 pages.

[34] Md Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M Kamal, Hua Wang, Anwaar Ulhaq, et al. 2018. Depression detection from social network data using machine learning techniques. *Health information science and systems* 6, 1 (2018), 1–12.

[35] Kokil Jaidka, Sharath Guntuku, and Lyle Ungar. 2018. Facebook versus Twitter: Differences in self-disclosure and trait prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.

[36] Zheng Ping Jiang, Sarah Ita Levitan, Jonathan Zomick, and Julia Hirschberg. 2020. Detection of mental health from reddit via deep contextualized representations. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*. 147–156.

[37] Kenneth S Kendler, Steven H Aggen, and Michael C Neale. 2013. Evidence for multiple genetic factors underlying DSM-IV criteria for major depression. *JAMA psychiatry* 70, 6 (2013), 599–607.

[38] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (2001), 606–613.

[39] Diya Li, Harshita Chaudhary, and Zhe Zhang. 2020. Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining. *International Journal of Environmental Research and Public Health* 17, 14 (2020), 4988.

[40] Tingting Liu, Salvatore Giorgi, Kenna Yadeta, H Andrew Schwartz, Lyle H Ungar, and Brenda Curtis. 2022. Linguistic predictors from Facebook postings of substance use disorder treatment retention versus discontinuation. *The American Journal of Drug and Alcohol Abuse* 48, 5 (2022), 573–585.

[41] Tingting Liu, Pallavi V. Kulkarni, Brenda L Curtis, Garrick T. Sherman, Kenna Yadeta, Louis Tay, Johannes C. Eichstaedt, and Sharath Chandra Guntuku. 2022. Head versus heart: social media reveals differential language of loneliness from depression. *npj Mental Health Research* 1 (2022).

[42] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv* abs/1907.11692 (2019).

[43] Matthew Matero, Akash Idnani, Youngseo Son, Salvatore Giorgi, Huy Vu, Mohammadzaman Zamani, Parth Limbachiya, Sharath Chandra Guntuku, and H Andrew Schwartz. 2019. Suicide risk assessment with multi-level dual-context language and BERT. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology*. 39–44.

[44] Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. (2002). http://mallet.cs.umass.edu.

[45] Saba Moussavi, Somnath Chatterji, Emese Verdes, Ajay Tandon, Vikram Patel, and Bedirhan Ustun. 2007. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *The Lancet* 370, 9590 (2007), 851–858.

[46] Laura Orsolini, Roberto Latini, Maurizio Pompili, Gianluca Serafini, Umberto Volpe, Federica Vellante, Michele Fornaro, Alessandro Valchera, Carmine Tomasetti, Silvia Fraticelli, et al. 2020. Understanding the complex of suicide in depression: from research to clinics. *Psychiatry investigation* 17, 3 (2020), 207.

[47] A Norcini Pala, P Steca, R Bagrodia, L Helpman, V Colangeli, P Viale, and ML Wainberg. 2016. Subtypes of depressive symptoms and inflammatory biomarkers: an exploratory study on a sample of HIV-positive patients. *Brain, behavior, and immunity* 56 (2016), 105–113.

[48] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015.* Technical Report.

[49] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate G. Blackburn. 2015. The Development and Psychometric Properties of LIWC2015.

[50] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society* 7, 2 (2021), 20563051211019004.

[51] Tom Pyszczynski, Kathleen Holt, and Jeff Greenberg. 1987. Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. *Journal of personality and social psychology* 52, 5 (1987), 994.

[52] Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. 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*. 99–107.

[53] Rafael Salas-Zárate, Giner Alor-Hernández, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde, Maritza Bustos-López, and José Luis Sánchez-Cervantes. 2022. Detecting depression signs on social media: a systematic literature review. In *Healthcare*, Vol. 10. MDPI, 291.

[54] James W Salazar, Karl Meisel, Eric R Smith, Aaron Quiggle, David B McCoy, and Matthew R Amans. 2019. Depression in patients with tinnitus: a systematic review. *Otolaryngology–Head and Neck Surgery* 161, 1 (2019), 28–35.

[55] Kiran Saqib, Amber Fozia Khan, Zahid Ahmad Butt, et al. 2021. Machine learning methods for predicting postpartum depression: Scoping review. *JMIR mental health* 8, 11 (2021), e29838.

[56] H Andrew Schwartz, Salvatore Giorgi, Maarten Sap, Patrick Crutchley, Lyle Ungar, and Johannes Eichstaedt. 2017. Dlatk: Differential language analysis toolkit. In *Proceedings of the 2017 conference on empirical methods in natural language processing: System demonstrations*. 55–60.

[57] Kerri Smith and IBC De Torres. 2014. A world of depression. *Nature* 515, 181 (2014), 10–1038.

[58] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, and Bernd Löwe. 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 166, 10 (2006), 1092–1097.

[59] Chrissoula Stavrakaki and Beverley Vargo. 1986. The relationship of anxiety and depression: a review of the literature. *The British Journal of Psychiatry* 149, 1 (1986), 7–16.

[60] Jackson G Thorp, Andries T Marees, Jue-Sheng Ong, Jiyuan An, Stuart MacGregor, and Eske M Derks. 2020. Genetic heterogeneity in self-reported depressive symptoms identified through genetic analyses of the PHQ-9. *Psychological Medicine* 50, 14 (2020), 2385–2396.

[61] Karan Wanchoo, Matthew Abrams, Raina M Merchant, Lyle Ungar, and Sharath Chandra Guntuku. 2023. Reddit language indicates changes associated with diet, physical activity, substance use, and smoking during COVID-19. *Plos one* 18, 2 (2023), e0280337.

[62] David G Weeks, John L Michela, Letitia A Peplau, and Martin E Bragg. 1980. Relation between loneliness and depression: a structural equation analysis. *Journal of personality and social psychology* 39, 6 (1980), 1238.

[63] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6

[64] Amir Hossein Yazdavar, Hussein S Al-Olimat, Monireh Ebrahimi, Goonmeet Bajaj, Tanvi Banerjee, Krishnaprasad Thirunarayan, Jyotishman Pathak, and Amit Sheth. 2017. 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*. 1191–1198.

[65] Amir Hossein Yazdavar, Mohammad Saeid Mahdavinejad, Goonmeet Bajaj, William Romine, Amit Sheth, Amir Hassan Monadjemi, Krishnaprasad Thirunarayan, John M Meddar, Annie Myers, Jyotishman Pathak, et al. 2020. Multimodal mental health analysis in social media. *Plos one* 15, 4 (2020), e0226248.

A APPENDIX

Source code and data available at https://github.com/devanshrj/depression-symptoms-reddit.

Table A1: Mapping Depression Symptoms to DSM-5 and RDoc Domains

Symptoms in current study	Corresponding RDoc domain	Corresponding DSM-5 criteria
Anhedonia	Negative Valence Systems	Markedly diminished interest or pleasure in most or all activities*
Sad mood	Negative Valence Systems	Depressed Mood*
Worthlessness	Negative Valence Systems	Feelings of worthlessness or excessive or inappropriate guilt
Suicidal thoughts and attempts	Negative Valence Systems	Recurrent thoughts of death (not just fear of dying), or suicidal ideation, plan, or attempt
Anxiety	Negative Valence Systems	N/A
Concentration deficit	Cognitive Systems	Diminished ability to think or concentrate, or indecisiveness
Fatigue	Arousal and Regulatory Systems	Fatigue or loss of energy
Sleep problem	Arousal and Regulatory Systems	Insomnia or hypersomnia
Disordered eating	Arousal and Regulatory Systems	Significant weight loss (or poor appetite) or weight gain
Anger	Arousal and Regulatory Systems	N/A
Somatic complaint	Arousal and Regulatory Systems	N/A
Self-loathing	Social Processes/Negative Valence Systems	N/A
Loneliness	Social Processes	N/A

* Core DSM-5 symptoms.

Table A2: Symptoms vs. All: 5-fold Cross Validation Prediction Performance (AUC) for Different Reddit Models

Symptom	LIWC	Reddit Topics	RoBERTa
Control	0.829	**0.852**	0.839
Anger	0.907	0.926	**0.949**
Anhedonia	0.791	0.898	**0.937**
Anxiety	0.857	**0.896**	0.888
Concentration deficit	0.805	0.897	**0.930**
Disordered eating	0.900	**0.948**	0.946
Fatigue	0.840	0.877	**0.928**
Loneliness	0.824	0.889	**0.894**
Sad mood	0.813	0.812	**0.853**
Self-loathing	0.821	0.885	**0.902**
Sleep problem	0.924	**0.979**	0.974
Somatic complaint	0.869	**0.960**	0.944
Suicidal thoughts and attempts	0.843	0.907	**0.916**
Worthlessness	0.727	0.800	**0.868**

Note: Model with best AUC for each symptom in **bold**.

Table A3: Symptoms vs. All: Validation Correlations with Self-Reported Assessments of Depression, Anxiety, Loneliness

	Depression			Anxiety			Loneliness		
Symptom	RoBERTa	FB Topics	Reddit Topics	RoBERTa	FB Topics	Reddit Topics	RoBERTa	FB Topics	Reddit Topics
Control	**-0.251*****	-0.196***	-0.191***	**-0.212*****	-0.149***	-0.155***	**-0.162*****	0.044*	-0.125***
Anger	**0.115*****	0.032	0.063***	**0.068*****	-0.002	0.04*	0.084***	0.063***	**0.098*****
Anhedonia	0.042*	**0.053****	0.011	0.029	**0.043***	0.008	0.037*	**0.087*****	0.066***
Anxiety	**0.103*****	0.088***	0.016	**0.07*****	0.07***	-0.003	0.067***	**0.088*****	0.066***
Concentration deficit	0.02	**0.096*****	-0.026	0.021	**0.089*****	-0.018	0.029	**-0.115*****	0.061***
Disordered eating	**0.162*****	0.132***	0.101***	**0.166*****	0.111***	0.103***	**0.059****	0.055**	0.024
Fatigue	-0.013	0.015	**-0.059****	-0.03	-0.014	**-0.077*****	0.014	**0.066*****	0.033
Loneliness	**0.143*****	0.091***	0.081***	**0.117*****	0.102***	0.089***	**0.082*****	0.035	0.035
Sad mood	0.032	0.034	-0.006	0.03	0.027	-0.004	0.013	0.003	-0.029
Self-loathing	**0.203*****	0.119***	0.086***	**0.182*****	0.108***	0.072***	**0.124*****	0.093***	0.084***
Sleep problem	**0.148*****	0.094***	0.028	**0.125*****	0.056**	0.001	0.096***	**0.098*****	0.081***
Somatic complaint	-0.005	**-0.058****	0.018	-0.039*	**-0.089*****	-0.006	-0.011	-0.026***	**0.061*****
Suicidal thoughts and attempts	**0.206*****	0.139***	0.187***	**0.156*****	0.084***	0.141***	0.141***	0.103	**0.143*****
Worthlessness	-0.01	**0.048****	-0.043	-0.013	0.033	-0.054**	0.031	0.032	0.006

Note: FB = Facebook. All correlations are Spearman correlations. Non-significant correlations are grayed out. All other correlations are significant at $p < 0.05$.