Studying Differential Mental Health Expressions in India

Anonymous ACL submission

Abstract

Psychosocial stressors and the symptomatology of mental disorders vary across cultures. Mental health expressions on social media, however, are primarily informed by studies in the WEIRD (Western, Educated, Industrial, Rich, and Democratic) contexts. In this paper, we analyze mental health posts on Reddit made by individuals in India, to identify variations in online depression language specific to the Indian context compared to users from the Rest of the World (ROW). Unlike in Western samples, mental health discussions in India additionally express sadness, use negation, are present-focused, and are related to work and achievement. *Illness* is exclusively correlated to India, reaffirming the link between somatic symptoms and mental disorders in Indian patients. Two clinical psychologists validated the findings from social media posts and found 95% of the top-20 topics associated with mental health discussions as *prevalent* in Indians. Significant linguistic variations in online mental health-related language in India compared to ROW, highlight the need for precision culturally-aware mental health models. These findings have important implications for designing culturally appropriate interventions to reduce the growing diagnosis and treatment gap for mental disorders in India.

1 Introduction

Over 197 million individuals in India are diagnosed with mental health disorders (Sagar et al., 2020), a disproportionate majority of whom do not receive treatment (Singh, 2018). The treatment gap for mental health disorders goes up to 95% in India, which is the highest across Asian countries and more severe compared to the gap of 78% in the United States (US) (Murthy, 2017; Naveed et al., 2020). Stigma, limited accessibility to mental healthcare, and a lack of awareness of mental health disorders, cumulatively contribute to diagnostic barriers for mental health care (Meshvara, 2002; Lahariya, 2018; Krendl and Pescosolido, 2020).

Cognitive impairments such as self-rumination and negative evaluation owing to depression can be measured from language. For instance, increasing use of the first-person singular pronouns reflects *self-rumination* (Spasojević and Alloy, 2001; Holtzman et al., 2017), and phrases expressing self-deprecation indicate *negative self-evaluation* (Kovacs and Beck, 1978). Social media platforms are increasingly used for mental health-related conversations and seeking support in India (Akbar et al., 2020). Language in social media posts can be harnessed to measure cultural differences in the manifestation of depression (Chancellor and De Choudhury, 2020; Guntuku et al., 2017; Eichstaedt et al., 2018).

While it is long known that depression manifests differently across cultures (Manson, 1995), but, *how* is still an unanswered research question. Lacking awareness of cultural differences in the manifestation of depression leads to misdiagnosis even in clinical settings (Bailey et al., 2009). This study is the first to focus on depression expressions in social media language posted by users in India. A related previous study by De Choudhury et al. (2017) grouped India with other developing nations - for example, South Africa.Another work by Pendse et al. (2019) studying the Indian population, examined help-seeking patterns rather than depression markers, to inform social media platform design. More importantly, we collaborate with clinical psychologists in India to validate trends and themes obtained from social media posts, which bolsters this computational study.

Understanding how the language of people with depression varies across cultures calls for interpretable features such as Linguistic Inquiry and Word Count (LIWC) (Boyd et al., 2022) rele-

[*] These authors served as co-first authors, alphabetically ordered by last name.

vant to other stakeholders such as psychologists (Aguirre et al., 2021; Loveys et al., 2018; Burkhardt et al., 2022; Rai et al., 2024). Using a blend of psychology-informed features and machine learning models, we address the overarching question, *if and how the mental health expressions of Indian users on social media are different from the rest of the world* by answering the following:

- How do linguistic expressions in Reddit posts of individuals experiencing mental health challenges in India differ from individuals in the rest of the world?
- How well do data-driven insights on mental health expressions align with the experience of clinical psychologists in India?
- Do the linguistic expressions of depression specific to India differ enough to reliably distinguish individuals from MH-India compared to those from MH-RoW?

This paper bridges two critical gaps from previous literature. First, the paper identifies mental health expressions specific to India, the world's most populous country, by mining Reddit threads. Second, it engages with clinical psychologists practicing in India to validate the empirical findings, providing a culturally informed assessment of cross-country comparisons of mental health expressions.

2 Data

Reddit offers a platform for individuals to share their mental health journey and seek support anonymously, thereby making it a rich source to understand the symptomatology and sequelae of mental health (De Choudhury and De, 2014; Park et al., 2018; Gaur et al., 2018; Lokala et al., 2022). Other more widely used social media platforms such as Mx Takatak, Moj, Tiktok in India lack APIs for data collection (oos). Previously, Reddit posts have been used for identifying shifts to suicidal ideation (De Choudhury et al., 2016), depression symptoms (Gaur et al., 2018; Liu et al., 2023), and the mental health expressions of immigrants (Mittal et al., 2023b), among others.

2.1 Subreddits: Mental Health vs Control

We extracted $3,195,310$ posts and comments from mental health-related subreddits (See Appendix A for list of subreddits) using the PushShift

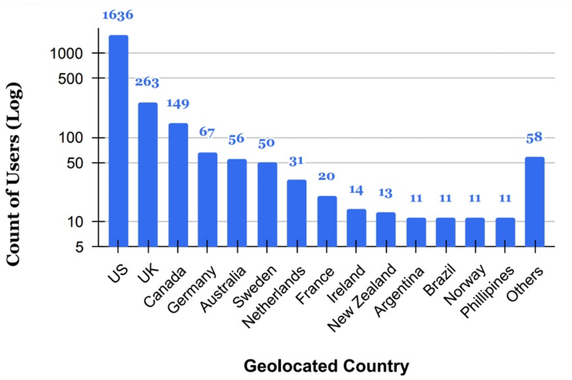

Figure 1: The count of users for each country in the Rest of World control group (log scale). Demonstrates that the large majority of users in the ROW group are geolocated to Western countries. The "Others" Category contains countries with less than 10 users, including Belgium (9), Italy (9), Mexico (6), Malaysia (5), Romania (4), Croatia (4), UAE (2), South Africa (2), China (2), Spain (2), Greece (2), Denmark (1), Finland (1), Iceland (1), Japan (1), South Korea (1), Poland (1), Russia (1), Singapore (1), Thailand (1), Turkey (1) and Vietnam (1).

API (Baumgartner et al., 2020). The largest portion of users (36.1%) were members of *r/depression*. The subreddits for the Control group were queried across all remaining subreddits external to the mental health subreddits.

Most posts and comments in the dataset were posted between 2019 and 2020. As a preprocessing step, we removed deleted usernames and null messages. Assuming that users posting in India-specific subreddits are likely to be Indians, we identified India-focused subreddits (See Appendix A) and then grouped together the users who posted in these subreddits. After this step and filtering users who posted at least 500 words (excluding comments), we grouped users into four groups:

- MH-India (4185 users): "Individuals geolocated in India and posting in Mental Health Subreddits",
- MH-ROW (5588 users): "Individuals geolocated outside India (i.e., Rest of World) and posting in Mental Health Subreddits",
- Control-India (2622 users): "Individuals geolocated in India and not posting in mental health subreddits" and,
- Control-ROW (5594 users): "Individuals geolocated outside India and not posting in mental health subreddits".

The first group (MH-India) is our *group of interest*; the remaining are controls. We specifically chose posts as the scope of this study was to obtain expressions of experiences with mental health challenges rather than interactions with others' mental health challenges.

2.2 Geolocation - India vs ROW

We used the geolocation inference approach (Harrigian, 2018) as a second layer of verification for user location. The geolocation model is a location estimation model that utilizes word usage, the frequency distribution of subreddit submissions, and the temporal posting habits of each user to determine their location. Specifically, we use the pre-trained GLOBAL inference model[1] to geolocate users in our dataset. We removed any users not geolocated to their group based on subreddit classification. For example, users in MH-India who are not geolocated to India and users in MH-RoW who are geolocated to India were removed. This functioned as a two-step verification to ensure that users in MH-India were from India. Ultimately, 1200 users out of the initial 4185 users were left in the MH-India group, and 930 users out of 2622 were left in the Control-India group. Most users in the ROW group were geolocated to the US (See Fig 1), affirming the dominance of West-centric data on Reddit.

We evaluated the quality of geolocation by manually verifying the self-disclosed location for randomly sampled 100 users. We found that the model's estimate of the individual's country matched the self-disclosed location, even though the state or city estimate was not always accurate.

2.3 Matching Control groups with users in MH-India

Age and gender are well-known confounders in behavioral health studies (Schwartz et al., 2013). We estimated age and gender for every user in our dataset using a machine-learning approach described in Appendix B to perform matching. We matched the samples from our group of interest, i.e., MH-India, with the samples in control groups (MH-ROW, and Control-ROW) on these two covariates. Matching was not performed for Control-India group due to the small sample size. The age distribution across the four groups was fairly similar before matching, with the average age being 25 for the MH-India, Control-India, and Control-ROW groups and 24 for the MH-ROW group.

[1] https://github.com/kharrigian/smgeo/tree/master#models

Group	# Distinct Users	# Posts
MH-India	1200	50928
Control-India	930	69957
MH-ROW	1200	54666
Control-ROW	1200	122654
Total	4530	298205

Table 1: Number of users and posts in each of the four groups of our dataset.

Ideally, the focus and control group samples should have indiscernible covariates. However, exact matching (Rosenbaum, 2020) is difficult to achieve without dropping a large set of samples. Coarsened Exact Matching (CEM) (Iacus et al., 2009) is a softer version of Exact Matching, which stretches the matching criteria wide enough to avoid dropping samples that are similar but not an exact match. We implement CEM using MatchIt package (Stuart et al., 2011) in R and set the distance to 'Mahalanobis' for one-to-one matching. The quality of matching was evaluated using Standard Mean Differences and Kolmogorov-Smirnov Statistics (See Appendix C). The mean age was 24.7 (sd= 3.41). The mean gender score was -0.97 (sd= 0.93), where a higher positive score indicates female. Table 1 shows the total number of posts and users in each of the four groups after CEM.

3 Methods

3.1 Language Features

We extracted n-grams, psychosocial word categories from LIWC, and topics (word clusters derived using LDA) to examine language correlated with depression.

1. We extracted **1-3 grams** from posts and created a normalized bag-of-words representation for each user. We filtered out 1-3 grams having point-wise mutual information (PMI)$\leq$5.

2. **Linguistic Inquiry Word Count 2022** (LIWC-22) is a dictionary comprising 102 word categories based on psycho-social states (e.g., Cognition, Social Processes, Affect, etc.). These word categories in LIWC are counted for each user, and the count is normalized by the total number of 1-grams for each user, thereby representing each user as

a vector of 102 normalized psychosocial categories.

3. We used Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to extract latent **topics** in users' timeline data. While recent neural topic modeling methods such as BERTopic (Grootendorst, 2022) show superior predictive accuracy, LDA has been shown to provide qualitatively robust topics (Dixon et al., 2022).

   We generated three sets of topics by setting the number of topics $= [200, 500, 2000]$. We did not experiment with higher topic numbers to avoid the curse of dimensionality. We evaluated the topics' quality using Topic Uniqueness (TU) (Nan et al., 2019). TU represents the number of times a set of keywords is repeated across topics; a higher TU corresponds to a rarely repeated word, indicating that topics are diverse, which is favorable. Additionally, three co-authors independently reviewed the quality of topics. We set the number of topics to 2000 based on the automated (See Table A3) and manual evaluation. While this is a large number, we evaluated 200, 500 and 2000 topics and found that 2000 topics were best able to perform in the Topic Uniqueness (TU) measure, indicating that they are able to display minimally overlapping topics from the corpus. We believe this is a crucial metric because of the large size of our corpus, which calls for a comprehensive latent topic analysis where each topic provides as unique a grouping as possible.

3.2 Correlation Analysis

To understand the association between language and the groups (MH-India, MH-ROW, Control-India, and Control-ROW), we performed ordinary least squares regressions with the three language feature sets (i.e., 1-3 grams, LIWC, and Topics) and MH-India vs control groups as the response variable to study language difference in mental health expression across groups. In this regression, the feature sets were independent variables. We calculated Pearson r to measure the association of each feature to each group in a one-vs-all setting. This is important in supporting an understanding of how each feature set correlates with the presence of mental health challenges in India. p-values were corrected using Benjamini-Hochberg correction for multiple hypothesis testing. 102 word categories for LIWC, 2000 for LDA topics, and 23,344 1-3 grams were considered for p-value correction.

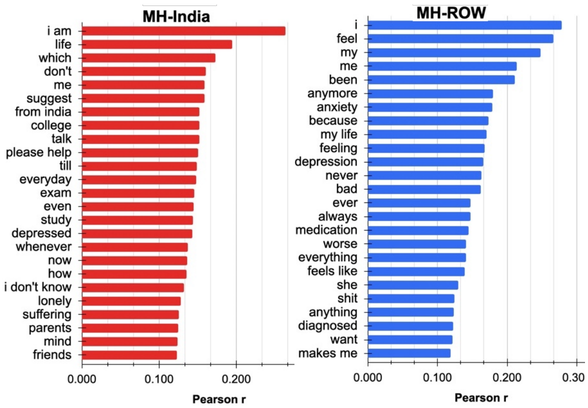

Figure 2: Top 25 statistically significant N-grams by effect size for both MH-India and MH-ROW. Significant at $p < .001$, two-tailed t-test, Benjamini-Hochberg corrected. Repeated N-grams are omitted.

3.3 Predictive Model

To examine whether the linguistic expressions of the MH-India group sufficiently differ to discriminate it from other groups, we trained 'one vs rest' logistic regression models in a 10-fold cross-validation setting (Rifkin and Klautau, 2004). More sophisticated methods (such as XGBoost) could potentially provide higher performance, but the focus of the study is not to achieve state-of-the-art performance for group prediction but to test if sufficient discriminating evidence exists across groups to warrant culturally aligned depression models. We report the Area Under the Receiver Operating Curves (AUC) for each feature for the MH-India and MH-ROW groups.

4 Results

4.1 Mental Health Expression: India vs ROW

4.1.1 N-grams

Sixty-one 1-3 grams were significantly ($p < 0.05$) correlated with the MH-India group, and 156 were correlated with the MH-ROW group. Figure 2 illustrates the top 25 1-3 grams arranged in decreasing order of Pearson r for both groups. First-person singular pronouns ('i am', 'i'), and mentions of depression ('depressed', 'depression'), are prevalent in both groups. Help-seeking phrases ('suggest', 'advice', 'talk', 'please help') are prevalent in MH-India, whereas MH-ROW discuss feelings, use clinical language ('diagnosed', 'medication') and express negative emotions ('anxiety', 'bad',

MH-India				MH-ROW			
	Category	r	**Top Words**		**Category**	r	**Top Words**
	Sadness	0.428	depression, sad, depressed, cry, lonely		Substances	0.645	drunk, wine, marijuana, vape, cbd
	Negative Tone	0.324	bad, wrong, lost, hate, depression		Health - Mental	0.625	depressed, addiction, bipolar, paranoid
Affect	Anxiety	0.197	scared, fear, afraid, worried, anxious,		Health - General	0.519	pain, fat, tired, depression,sick
	Negation	0.326	not, don't, no, never, can't	**Physical**	Feeling	0.613	feel, hard, felt, feeling, cool
	Personal Pronouns	0.274	i, my, you, me, they		Personal Pronouns	0.614	i, you, my, me, i'm
Time	*Present Focus*	0.298	is, are, can, am, i'm		Conjunctions	0.432	and, but, as, so, or
	Health - Mental	0.290	depressed, addiction, bipolar, adhd		Common Verbs	0.424	is, have, was, be, are
	Health - General	0.257	depression, pain, tired, sick, fat		Common Adjectives	0.327	more, other, only, much, new
	Feeling	0.174	feel, hard, feeling, felt, pain	**Linguistic**	Impersonal Pronouns	0.308	it, that, this, what, it's
Physical	*Illness*	0.171	pain, sick, covid, painful, recovery		Negative Tone	0.571	bad, wrong, lost, hit, hate
	Causation	0.254	how, because, make, why, since	**Affect**	Anxiety	0.559	fear, worried, scared, afraid, worry
	All-or-none	0.222	all, no, never, every, always	**Motives**	Allure	0.450	have, like, out, get, time
Cognition	Insight	0.195	how, know, feel, think, find		All-or-none	0.412	all, no, never, every, always
	Communication	0.237	thanks, said, say, tell, talk		Certitude	0.389	really, actually, completely, simply
Social	*Politeness*	0.195	please, thanks, hi, thank, ms	**Cognition**	Insight	0.292	how, know, think, feel, find
Motives	Allure	0.237	have, like, get, know, now	**Time**	Past Focus	0.368	was, had, been, i've, were
States	Want	0.200	want, wanted, hope, wish, wants		Want	0.280	want, wanted, hope, wants, wish
Lifestyle	*Work*	0.200	work, job, edit, working, school	**States**	Acquire	0.278	get, got, take ,getting, took
Drives	*Achievement*	0.179	work, better, tried, best, able	**Social Ref.**	Friends	0.273	bf, mate, buddies, mates, ally

Table 2: Top LIWC categories for MH-India and MH-ROW along with Pearson r effect sizes and top 5 words by frequency in our dataset. All categories shown are statistically significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected. The italicized text represents categories exclusive to MH-India group.

'worse') associated with depression. Interestingly, the MH-India group simultaneously discusses the feelings of loneliness and social relationships with parents and friends. Academic-related stress ('college', 'exam', 'study') is exclusively seen in discussions of the MH-India group. This is particularly interesting, considering users in both groups were matched for age, yet discussions around student-life challenges are prevalent only in MH-India.

Overall, the discussion in MH-India subreddits is centered around seeking help, whereas negative feelings are more commonly discussed in the MH-ROW group.

4.1.2 LIWC

Fifty-two LIWC categories were significantly associated ($p < 0.05$) with the MH-India group, whereas sixty categories were found to be correlated with the MH-ROW group. We provide the Top 20 LIWC categories for both groups in Table 2. While negative tone, anxiety, and personal pronouns are correlated with depression in both groups, *sadness*, an additional affect is only seen in the MH-Indian group. Depression-related discourse in MH-ROW group is past-focused, which is a known marker of depression whereas it is present-focused in India. Social behavioral attributes (*communication, politness*), work, and achievement are correlated with depression in the MH-India group but not in MH-ROW. Somatic symptoms/illness (pain, sick) are seen in the MH-India group, whereas substance abuse/addiction is correlated with the MH-ROW group.

4.1.3 LDA Topics

Of 2000 topics, 109 were found to be significant ($p < 0.05$) for the MH-India group and 216 for MH-ROW group. The top topics and the corresponding Pearson r, are provided in Table **??**. The most prevalent topics in the MH-India group discussed negative/suicidal thoughts (*life, parents, family, hate, die*), academic and job stressors ((*college, exam, study, university, engineering*), *job, degree, college, school, career*) and family and relationships (*love, heart, loved, beautiful, happiness*). In contrast, the prevalent themes in MH-ROW are negative feelings and emotions ((*feel, myself, feeling, depression, anymore*);(*went, didn, crying, mad, stayed*); (*feeling, body, feels, heart, scared*)). Topics such as mental disorders (*anxiety, depression, mental, medication, disorder*), goodbyes (*sister, ma, papa, clutching, plane*) and anger (*hate, angry, rant, respect, ugly*) are common in both sets.

4.2 Validation

We validated the significantly correlated topics for our group of interest (MH-India) by showing the top words to two clinical psychologists with significant practical experience with seeing patients and mental healthcare in India. Specifically, we asked the following question:

To what extent the open vocabulary topics having a significant correlation with the MH-India group are prevalent in Indian patients? - A Likert scale of 0-5 is provided where 5 indicates 'Highly Prevalant' and '0' indicates 'Not observed at all'.

Prevalence While independently labeling topics

MH-India			MH-ROW		
Theme	**Top Words**	**r**	**Theme**	**Top Words**	**r**
Negative/ Suicidal thoughts	life, parents, family, hate, die	0.280	Negative Feelings	feel, myself, feeling, depression, anymore	0.358
Depression mentions	anymore, depression, tired, depressed, everyday	0.229	Negative Feelings	went, didn, crying, mad, stayed	0.340
Negative Feelings	feel, myself, feeling, depression, anymore	0.227	Depression mentions	anymore, depression, tired, depressed, everyday	0.338
Friends & Relationships	friends, talk, social, anxiety, alone	0.226	Negative Feelings	feeling, body, feels, heart, scared	0.322
Friends & Relationships	love, heart, loved, beautiful, happiness	0.207	Symptoms	sick, woke, stomach, switched, asleep	0.318

Table 3: Top Words and Correlation Coefficient (r) by Theme for MH-India and MH-ROW

	1-3 grams	LIWC	LDA Topics
MH-India	0.853	0.776	0.758
MH-ROW	**0.881**	0.818	0.811

Table 4: AUCs for Logistic regression one vs. rest models predicting group membership.

for prevalence, the clinical psychologists agreed with each other 81.49% of the time. Of the top 20 topics significantly associated with the MH-India group, 95% were ranked either extremely or somewhat prevalent (4 or 5 on a scale of 1 - 5) in India by at least one of the two clinical psychologists, and 80% were ranked as prevalent (a score of 4 or 5) by both evaluators. Of the 109 topics significantly associated with the MH-India group, 56% were annotated as prevalent by at least one evaluator.

4.3 Is depression language of MH-India different from control groups?

Our logistic regression model uses each feature set (i.e., n-grams, LIWC and LDA topics) to predict membership in MH-India vs MH-ROW. High AuC scores (See Table 4) demonstrate that users' language in the MH-India group significantly differs from those in the control groups, including MH-ROW. All language feature groups (i.e., n-grams, LIWC, and LDA topics) have fairly high AUC, indicating differences in depression expression at 1-3 gm level, in psycho-social categories as well as in latent topics of discussion.

5 Discussion

Our work reveals significant differences in the language markers of mental health expressions of Indian users compared to those from outside India. The association with *politeness* coupled with discussions around *family*, *work* and *achievement* indicates that the users in the MH-India group tend to associate mental health with their ability and social relationships as opposed to more mentions of swear words and feelings in the ROW group. *Academic* and *family pressures* are unique to Indian users, possibly due to the collectivist nature of Indian society (Chadda and Deb, 2013). The matching performed across control groups rules out the majority young demographic in social media data as the possible reason. The association with present focus words in MH-India contrasts with the widespread belief of self-focussed rumination when suffering from mental illness (Park et al., 2017). MH-India group also tends to reason ('causation' words - *how, because, why*) more in their language in contrast to expressing *feelings* in MH-ROW group. Communication (*phone, call, message, post/tweet/meme, sms/texting, chat*) is also exclusive to the MH-India group. We speculate high reliance on social media platforms for mental health support and privacy potentially due to stigma associated with mental health in face-to-face conversations (Shidhaye and Kermode, 2013).

Our work reaffirms cultural differences such as substance use in the MH-ROW population vs somatic complaints in the MH-India population. Unlike in Western cultures where substance use is a barrier to mental health support, it is familial concerns and the fear of stigma in MH-India (Biswas et al., 2016). We speculate based on these confirmations that culturally-aware mental health intervention techniques should seek to bridge the gap to treatment by addressing hypochondriacal ideas and familial embarrassment in particular.

Only 56% of 109 topics correlated with the MH-India group were labeled as *prevalent* in Indian patients by clinical psychologists. We speculate that some of the topics not labeled as "prevalent" are unseen or emerging themes. Of Top-20 topics in MH-India, "not prevalent" topics revolve around *Video Games/Online Content*, *Grooming/Physical Appearance*, and *Programming*, indicating the influence of digital content, growing isolation, and low self-esteem amongst the undiagnosed young population. These topics could be underrecognized concerns. The second group of topics includes *Environmental Impact of Energy Sources*, *Humorous content and reactions*, among others. Previous research has suggested that people (particularly young people) are increasingly climate anxious and that humor on social media is often used to cope with mental health challenges (Schneider, 2018; Sanson, 2022). Furthermore, certain pop culture references may be crucial to understanding the narrative - as one clinical psychologist pointed out, the top word *"Singh"* in one of the topics may correspond to the suicide of late Bollywood star Sushant Singh Rajput, a significant event that potentially catalyzed a range of important conversations surrounding mental health in India (Akbar et al., 2020).

The growing treatment gap for mental disorders is a major concern in Indian society. The economic loss from mental health conditions between 2012-2030 is estimated at USD 1.03 trillion[2]. Automated systems that could diagnose and support mental well-being can potentially alleviate the lack of resources, but they would only be useful when designed considering the cultural sensitivities and norms of society. This study establishes significant linguistic variations in the mental health-related language in social media posts by Indians compared to individuals from the rest of the world. It thus becomes imperative to test the computational depression detection models for cultural invariance before their deployment in clinical settings (Aguirre et al., 2021; Rai et al., 2023) and indicate the need for socio-culturally aware mental health models to prevent misdiagnosis.

6 Background

Depression and anxiety disorders are the most imminent mental health challenges, with the highest contribution to Indian Disability Adjusted Life Years (Sagar et al., 2020). Fear of embarrassing family members is a primary barrier to mental health recovery in India whereas, for example, substance abuse is the major hurdle in America (Biswas et al., 2016). 71% of Indians exhibited stigma when answering questions about mental health (Foundation, 2018). Relatedly, somatic symptoms, hypochondriasis, anxiety, and agitation are more commonly seen in Indian patients compared to psychological symptoms (Gada, 1982). While the extent of mental health stigma and treatment (un)availability is often studied, it remains unknown how individuals suffering from mental disorders in India express and seek support on social media.

There is accumulating evidence that suggests language markers of depression vary with demographics such as race (Rai et al., 2024; Aguirre and Dredze, 2021), immigrant status (Mittal et al., 2023a), and geographic location (De Choudhury et al., 2017). For instance, Indian and South Africa-based users are less candid in their posts and less likely to exhibit negative emotions in comparison to their Western counterparts (De Choudhury et al., 2017). Another study studied the language of Asian users including *Indian*, *Malaysian*, and *Filipino* on Mental Health Support Forum, Talklife (Pendse et al., 2019) and found that Indians discuss “wanting or needing friends” more than other countries.
.

Limitations

The text-based geolocation of individuals in this study could potentially label Indians who later moved to other countries as Indians residing in India. Further, the Reddit user sample does not represent the general population, as evidenced by the mostly English language data in our India samples, although India has over 100 languages. In particular, we note that the majority of users were geolocated to Karnataka (a southern state in India) and that the age (ranging between 12 and 48) distributions are not necessarily representative. Our work shows the significant cultural themes observed in Indian society. However, Reddit posts represent a small population of India. While this analysis provides correlational insight into the data, it does not offer causal claims.

Ethical Considerations

Our university’s Institutional Review Board deemed this study exempt due to the public nature of all data. While Reddit data is public, it

[2] United Nations: https://www.who.int/india/health-topics/mental-health

may contain users' personal information, including city and town. We limited our analysis to country and state-level geolocation information to reduce the possibility of personally identifying individuals. Gender was predicted using a continuous scale, with extremes indicating masculinity and feminity. We exercised caution while presenting linguistic patterns and examples not to reveal any individual's timeline quotes. Members of our team have not viewed or worked with individual-level granular data. When done ethically with respect to user anonymity and privacy, we believe this line of research could assist in understanding diverse individuals' mental health challenges and developing personalized interventions that improve the well-being and mental health of under-resourced communities (Proferes et al., 2021).

Acknowledgment

We acknowledge the limited use of OpenAI's ChatGPT in the writing and refining this manuscript.

References

"social social media in india - 2023 stats platform trends". "considered as an approximation.".

Carlos Aguirre and Mark Dredze. 2021. Qualitative analysis of depression models by demographics. In *Proceedings of the seventh workshop on computational linguistics and clinical psychology: improving access*, pages 169–180.

Carlos Aguirre, Keith Harrigian, and Mark Dredze. 2021. Gender and racial fairness in depression research using social media. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2932–2949.

Syeda Zainab Akbar, Ankur Sharma, Himani Negi, Anmol Panda, and Joyojeet Pal. 2020. Anatomy of a rumour: Social media and the suicide of sushant singh rajput. *arXiv preprint arXiv:2009.11744*.

Rahn Kennedy Bailey, Holly L Blackmon, and Francis L Stevens. 2009. Major depressive disorder in the african american population: meeting the challenges of stigma, misdiagnosis, and treatment disparities. *Journal of the National Medical Association*, 101(11):1084–1089.

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*, volume 14, pages 830–839.

Jhilam Biswas, BN Gangadhar, and Matcheri Keshavan. 2016. Cross cultural variations in psychiatrists' perception of mental illness: a tool for teaching culture in psychiatry. *Asian journal of psychiatry*, 23:1–7.

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

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, pages 1–47.

Felix Burkhardt, Anabell Hacker, Uwe Reichel, Hagen Wierstorf, Florian Eyben, and Björn Schuller. 2022. A comparative cross language view on acted databases portraying basic emotions utilising machine learning. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1917–1924.

Rakesh K Chadda and Koushik Sinha Deb. 2013. Indian family systems, collectivistic society and psychotherapy. *Indian journal of psychiatry*, 55(Suppl 2):S299–S309.

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):43.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 71–80.

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*, pages 2098–2110.

Munmun De Choudhury, Sanket S Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 353–369.

Erica L Dixon, Sukanya M Joshi, William Ferrell, Kevin G Volpp, Raina M Merchant, and Sharath Chandra Guntuku. 2022. Covid-19 contact tracing app reviews reveal concerns and motivations around adoption. *Plos one*, 17(9):e0273222.

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):11203–11208.

Live Love Laugh Foundation. 2018. How india perceives mental health: Tlllf 2018 national survey report.

MT Gada. 1982. A cross cultural study of symptomatology of depression—eastern versus western patients—. *International Journal of Social Psychiatry*, 28(3):195–202.

Manas Gaur, Ugur Kursuncu, Amanuel Alambo, Amit 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. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 753–762.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

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*, 18:43–49.

Keith Harrigian. 2018. Geocoding without geotags: a text-based approach for reddit. *arXiv preprint arXiv:1810.03067*.

Nicholas S Holtzman et al. 2017. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, 68:63–68.

Stefano Iacus, Gary King, and Giuseppe Porro. 2009. Cem: Software for coarsened exact matching. *Journal of statistical Software*, 30:1–27.

Maria Kovacs and Aaron T Beck. 1978. Maladaptive cognitive structures in depression. *American Journal of psychiatry*, 135(5):525–533.

Anne C Krendl and Bernice A Pescosolido. 2020. Countries and cultural differences in the stigma of mental illness: the east–west divide. *Journal of Cross-Cultural Psychology*, 51(2):149–167.

C Lahariya. 2018. Strengthen mental health services for universal health coverage in india. *Journal of postgraduate medicine*, 64(1):7.

Tingting Liu, Devansh Jain, Shivani Reddy Rapole, Brenda Curtis, Johannes C Eichstaedt, Lyle H Ungar, and Sharath Chandra. 2023. Detecting symptoms of depression on reddit. In *Proceedings of the 15th ACM Web Science Conference 2023*, pages 174–183.

Usha Lokala, Aseem Srivastava, Triyasha Ghosh Dastidar, Tanmoy Chakraborty, Md Shad Akhtar, Maryam Panahiazar, and Amit Sheth. 2022. A computational approach to understand mental health from reddit: knowledge-aware multitask learning framework. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 640–650.

Kate Loveys, Jonathan Torrez, Alex Fine, Glen Moriarty, and Glen Coppersmith. 2018. Cross-cultural differences in language markers of depression online. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, pages 78–87.

Spero M Manson. 1995. Culture and major depression: Current challenges in the diagnosis of mood disorders. *Psychiatric Clinics of North America*, 18(3):487–501.

Deva Meshvara. 2002. Mental health and mental health care in asia. *World Psychiatry*, 1(2):118.

J. Mittal, A. Belorkar, V. Jakhetiya, V. Pokuri, and S. C. Guntuku. 2023a. 'language on reddit reveals differential mental health markers for individuals posting in immigration communities'. *15th ACM Web Science Conference 2023*.

Juhi Mittal, Abha Belorkar, Vinit Jakhetiya, Venu Pokuri, and Sharath Chandra Guntuku. 2023b. Language on reddit reveals differential mental health markers for individuals posting in immigration communities. In *Proceedings of the 15th ACM Web Science Conference 2023*, pages 153–162.

R Srinivasa Murthy. 2017. National mental health survey of india 2015–2016. *Indian journal of psychiatry*, 59(1):21.

Feng Nan, Ran Ding, Ramesh Nallapati, and Bing Xiang. 2019. Topic modeling with wasserstein autoencoders. *arXiv preprint arXiv:1907.12374*.

Sadiq Naveed, Ahmed Waqas, Amna Mohyud Din Chaudhary, Sham Kumar, Noureen Abbas, Rizwan Amin, Nida Jamil, and Sidra Saleem. 2020. Prevalence of common mental disorders in south asia: a systematic review and meta-regression analysis. *Frontiers in psychiatry*, 11:573150.

Albert Park, Mike Conway, et al. 2018. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. *Journal of medical Internet research*, 20(4):e8219.

Gregory Park, H Andrew Schwartz, Maarten Sap, Margaret L Kern, Evan Weingarten, Johannes C Eichstaedt, Jonah Berger, David J Stillwell, Michal Kosinski, Lyle H Ungar, et al. 2017. Living in the past, present, and future: Measuring temporal orientation with language. *Journal of personality*, 85(2):270–280.

Sachin R Pendse, Kate Niederhoffer, and Amit Sharma. 2019. Cross-cultural differences in the use of online mental health support forums. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–29.

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):20563051211019004.

Sunny Rai, Elizabeth C Stade, Salvatore Giorgi, Ashley Francisco, Lyle H Ungar, Brenda Curtis, and Sharath C Guntuku. 2023. Key markers of depression depend on race. *PsyArXiv. November 2. doi:10.31234/osf.io/uvamz.*

Sunny Rai, Elizabeth C Stade, Salvatore Giorgi, Ashley Francisco, Lyle H Ungar, Brenda Curtis, and Sharath C Guntuku. 2024. Key language markers of depression on social media depend on race. *Proceedings of the National Academy of Sciences*, 121(14):e2319837121.

Ryan Rifkin and Aldebaro Klautau. 2004. In defense of one-vs-all classification. *The Journal of Machine Learning Research*, 5:101–141.

Paul R Rosenbaum. 2020. Modern algorithms for matching in observational studies. *Annual Review of Statistics and Its Application*, 7:143–176.

Rajesh Sagar, Rakhi Dandona, Gopalkrishna Gururaj, RS Dhaliwal, Aditya Singh, Alize Ferrari, Tarun Dua, Atreyi Ganguli, Mathew Varghese, Joy K Chakma, et al. 2020. The burden of mental disorders across the states of india: the global burden of disease study 1990–2017. *The Lancet Psychiatry*, 7(2):148–161.

Koustuv Saha, Sindhu Kiranmai Ernala, Sarmistha Dutta, Eva Sharma, and Munmun De Choudhury. 2020. Understanding moderation in online mental health communities. In *International Conference on Human-Computer Interaction*, pages 87–107. Springer.

Dubika Sanson. 2022. “editorial: The climate and ecological mental health emergency – evidence and action. *Child and Adolescent Mental Health*, 27(1):3.

Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar, and H Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1146–1151.

Tran Schneider, Voracek. 2018. “a joke a day keeps the doctor away?” meta-analytical evidence of differential associations of habitual humor styles with mental health. *Scandinavian Journal of Psychology*, 59(3):300.

H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Eva Sharma and Munmun De Choudhury. 2018. Mental health support and its relationship to linguistic accommodation in online communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–13.

Rahul Shidhaye and Michelle Kermode. 2013. Stigma and discrimination as a barrier to mental health service utilization in india. *International health*, 5(1):6–8.

Om Prakash Singh. 2018. Closing treatment gap of mental disorders in india: Opportunity in new competency-based medical council of india curriculum. *Indian journal of psychiatry*, 60(4):375.

Jelena Spasojević and Lauren B Alloy. 2001. Rumination as a common mechanism relating depressive risk factors to depression. *Emotion*, 1(1):25.

Elizabeth A Stuart, Gary King, Kosuke Imai, and Daniel Ho. 2011. Matchit: nonparametric preprocessing for parametric causal inference. *Journal of statistical software*.

Ayah Zirikly, Philip Resnik, Ozlem Uzuner, and Kristy Hollingshead. 2019. Clpsych 2019 shared task: Predicting the degree of suicide risk in reddit posts. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology*, pages 24–33.

A Subreddits Used to Extract the Raw Data

Mental Health Subreddits: The mental health subreddit was obtained from prior works (Sharma and De Choudhury, 2018; Saha et al., 2020). These include: r/Anxiety, r/bipolar, r/BipolarReddit, r/depression, r/sad, r/SuicideWatch, r/addiction, r/opiates, r/ForeverAlone, r/BPD, r/selfharm, r/StopSelfHarm, r/OpiatesRecovery, r/Sadness, r/schizophrenia, r/AdultSelfHarm

Control Subreddits: All subreddits excluding Mental health subreddits.

India focused Subreddits: r/india, r/mumbai, r/tamil, r/Hindi, r/Kerala, r/Urdu, r/delhi, r/pune, r/hyderabad, r/bangalore, r/kolkata, r/telugu, r/marathi, r/AskIndia, r/sanskrit, r/Kochi, r/Rajasthan, r/pali, r/Chandigarh, r/Chennai, r/karnataka, r/Bhopal, r/Coimbatore, r/kannada, r/TamilNadu, r/Trivandrum, r/gujarat, r/punjabi, r/Bengali, r/kolhapur, r/Vijaywada, r/Dehradun, r/sahitya, r/Uttarakhand, r/ahmedabad, r/bharat, r/nagpur, r/Agra, r/assam, r/Indore, r/surat, r/navimumbai, r/Goa, r/sikkim, r/lucknow, r/Bareilly, r/nashik, r/Allahabad, r/Durgapur, r/Jamshedpur, r/Asansol, r/indianews, r/IndianGaming, r/IndiaSpeaks, r/indiameme, r/dankinindia, r/indiasocial

B Age and Gender

We applied an open-source age and gender predictive lexica (Sap et al., 2014) to obtain continuous values of age and gender. This lexicon was built over a set of over 70,000 users from social media and blogs and predicted age with a Pearson r of 0.86 and gender with an accuracy of 0.91 and has been applied reliably on Reddit data in prior studies (Zirikly et al., 2019). We used the probabilities from this model to denote the gender attribute of users in our data and did not consider gender as a binary category.

To validate the machine-generated predictions of gender and age for users within the Reddit dataset, we looked for posts containing self-disclosures of gender and age per user. Examples of this include "(23F)" for a user who self identifies as a 23 year old female. We were able to identify 5,844 posts across 706 unique users (See Table A1 for distribution) who employed some form of gender self-identification, allowing us to measure the accuracy of provided gender predictions. Using this subset, the model's gender prediction holds at 91.89% (See Table A2 for groupwise performance).

Group Name	Gender	Age
MH India	224	427
MH ROW	195	378
Non-MH India	45	111
Non-MH ROW	242	388

Table A1: Distribution of Users who self-disclosed gender and age by Group

Group Name	Accuracy	MAE (in yrs)
MH India	87.67%	3.12
MH ROW	86.69%	4.65
Non-MH India	94.33%	4.38
Non-MH ROW	96.12%	6.48

Table A2: Model performance for predicting gender and age. MAE stands for Mean Absolute Error and is reported in years.

C Coarsened Exact Matching

A control group is considered balanced with the treatment group if the difference is close to zero. The focus group MH-India has 1200 users and a total of 1200 users were matched from the groups MH-ROW and Control-ROW. Matching was not performed for Control-India as the number of samples is 930 and we did not want to drop any samples.

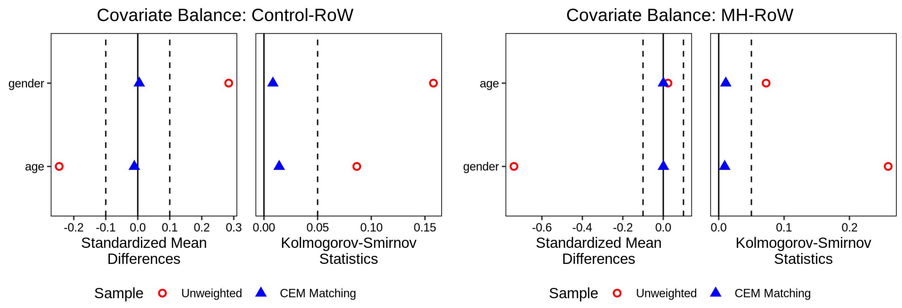

Figure 3: Differences in Covariates before and after CEM for groups “Control-ROW” and “MH-ROW”

D Topic Uniqueness Scores

Table A3 describes the TU scores generated by each of the tested combinations of number of topics and grouping of Posts and Comments. This was used to choose Posts Only, Number of Topics = 2000 for our analysis.

Message Type	Number of Topics	Value
Both Messages	200	0.7205
Both Messages	500	0.6194
Both Messages	2000	0.5676
Posts Only	200	0.7675
Posts Only	500	0.7162
Posts Only	2000	0.7694

Table A3: Topic Uniqueness Scores for different numbers of topics and message types.

E Communication with Clinical Psychologists

Table A3 shows the email text used for communication with each clinical psychologist.

The goal of this project is to study the manifestation of mental illness in Indians. As a part of this project, we have identified a set of 100 Topics/ Themes that Indians Users were found to commonly discuss on Reddit, a social media platform. We have labeled these topics as per our understanding and we now need your help in interpreting these topics from your perspective. The objective is to essentially identify

- Topic or theme of discussion in the context of mental illness in India,
- How often a theme is observed in an Indian patient suffering from a mental illness?

These identified topics are available in this Google sheet. Please read the below steps carefully:

- Peruse the top words given in Column-A. These are the top 10 common words comprising a single topic.
- In Column E, select the degree of prevalence of this topic amongst Indian patients. The options are Highly prevalent, somewhat prevalent, unsure, rarely observed, and Not observed at all.
- You may add your comments in Column -F

Table A4: Email communication with clinical psychologists who performed an informed review of the topics and ChatGPT-generated labels in this study.