Cross-Cultural Differences in Mental Health Expressions on Social Media

Sunny Rai[*†], **Khushi Shelat**[*†], **Devansh R Jain**[◇], **Kishen Sivabalan**[†], **Young Min Cho**[†],
Maitreyi Redkar[♠], **Samindara Sawant**[♣], **Lyle H. Ungar**[†] **& Sharath Chandra Guntuku**[†]
[†]University of Pennsylvania, [◇]Carnegie Mellon University,
[♠]Indian Institute of Technology, Bombay, [♣]Disha Counseling Center
{sunnyrai, sharathg}@seas.upenn.edu

Abstract

Culture moderates the way individuals perceive and express mental distress. Current understandings of mental health expressions on social media, however, are predominantly derived from WEIRD (Western, Educated, Industrialized, Rich, and Democratic) contexts. To address this gap, we examine mental health posts on Reddit made by individuals geolocated in India, to identify variations in social media language specific to the Indian context compared to users from Western nations. Our experiments reveal significant psychosocial variations in emotions and temporal orientation. This study demonstrates the potential of social media platforms for identifying cross-cultural differences in mental health expressions (e.g. seeking advice in India vs seeking support by Western users). Significant linguistic variations in online mental health-related language emphasize the importance of developing precision-targeted interventions that are culturally appropriate.

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 mental healthcare (Singh, 2018). Generative AI technologies can facilitate affordable and easily accessible mental health assessment and support, especially in under-resourced contexts such as India (Stade et al., 2024).

Mental disorders, however, manifest differently across cultures (Manson, 1995). For example, AI models trained on Black individuals' language fail to detect depression in Black individuals (Rai et al., 2024). Moreover, the frequency of language markers such as self-focus and negative evaluation indicative of depression do not increase with depression in Black individuals' language. AI systems lacking awareness of cross-cultural differences in mental health communications may lead to misdiagnosis and health inequity (Bailey et al., 2009; Lewis-Fernandez et al., 2005).

In this paper, we examine *how mental health expressions specific to Indian context vary compared to the rest of the world* (RoW). Previously, De Choudhury et al. (2017) examined depression patterns in the online language of individuals from majority world countries including India. Another related work examined help-seeking patterns to inform social media platform design (Pendse et al., 2019). This paper bridges two critical gaps from previous literature. First, we focus on mental health expressions specific to India, the world's most populous country, by mining Reddit threads. Second, we collaborate with clinical psychologists practicing in India to validate the empirical findings, providing a culturally informed assessment of cross-cultural comparisons of mental health expressions.

To inform the research on culturally competent mental health models (Sue, 1998), we adopt interpretable features that are comprehensible to stakeholders such as psychologists and policymakers, for modeling cross-cultural variations in mental health language. We use psychosocial word categories (e.g., Linguistic Inquiry and Word Count (LIWC)) and topic modeling (word clusters derived using LDA) to examine variations in language correlated with depression (Aguirre et al., 2021; Loveys et al., 2018; Burkhardt et al., 2022; Rai et al., 2024). Using a blend of cross-disciplinary interpretable language 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 (RoW)* by answering the following:

- How do linguistic expressions in Reddit posts of individuals experiencing mental health challenges in India differ from individuals in the

---

[*]These authors contributed equally to this work

RoW?

- 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 significantly to distinguish individuals from India compared to those from RoW?

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. India ranked 3rd in Reddit's website traffic globally after the US and the UK in 2024[*]. Previously, Reddit posts have been used for identifying shifts to suicidal ideation (De Choudhury et al., 2016), depression symptoms (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 API (Baumgartner et al., 2020). The largest portion of users (36.1%) were members of *r/depression*. We queried subreddits external to the mental health subreddits to create a control group. We removed deleted usernames and null messages. We considered users who posted at least 500 words (excluding comments) to ensure sufficient linguistic richness in users' language for further analysis. We limit the scope of this study to Reddit posts to obtain expressions of experiences with mental health challenges rather than interactions (as in comments) with others' mental health challenges.

2.2 Geolocation - India vs RoW

For labeling users' location, we performed two-level analysis. First, we identified India-focused subreddits (See Appendix A) since users posting in India-specific subreddits are likely to be Indians and grouped together users into four groups:

- MH-India (4185 users): "Individuals posting in India-specific Subreddits and posting in Mental Health Subreddits",
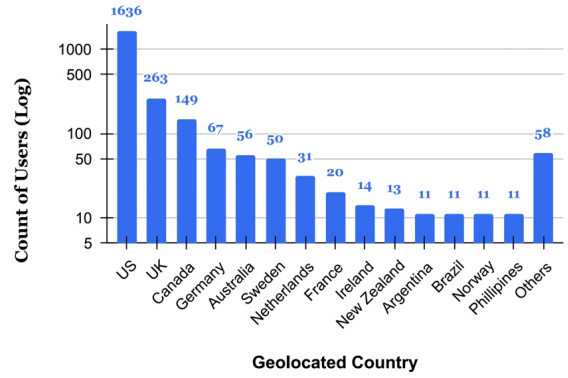

Figure 1: The count of users for each country in the Rest of World control group (log scale). Majority of users in the RoW group are geolocated to **Western countries**. The "Others" Category contains countries with less than 10 users, including Belgium (9), Italy (9), Mexico (6), Malaysia (5), Romania (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).

- MH-RoW (5588 users): "Individuals NOT posting in India-specific Subreddits and posting in Mental Health Subreddits",

- Control-India (2622 users): "Individuals posting in India-specific Subreddits and NOT posting in mental health subreddits" and,

- Control-RoW (5594 users): "Individuals NOT posting in India-specific Subreddits and NOT posting in mental health subreddits".

The first group (MH-India) is our *group of interest*; the remaining are controls.

Second, 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[*] 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.
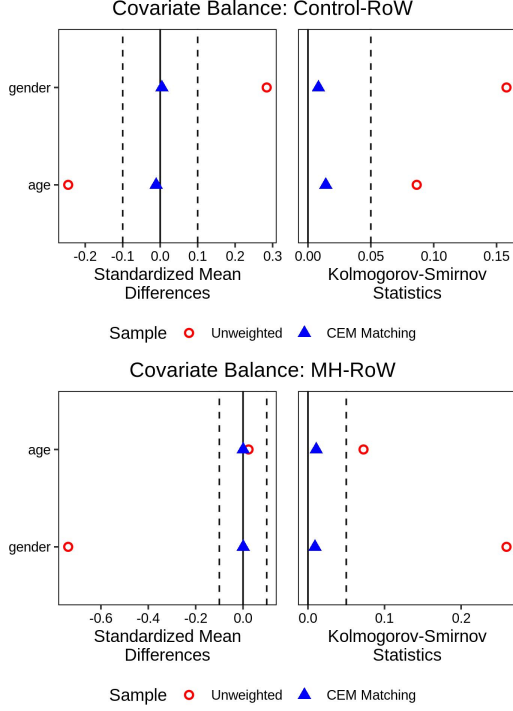
133

Figure 2: Differences in Covariates before and after CEM for groups "Control-RoW" and "MH-RoW". A control group is considered balanced with the treatment group if the difference is close to zero. Matching was not performed for Control-India due to smaller sample size.

Manual Evaluation 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.

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. The majority of the users ($\approx 95\%$) in the RoW group were geolocated to Western nations (See Fig 1), affirming the dominance of West-centric data on Reddit. From now on, we use the terms RoW and Western nations interchangeably.

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 first estimated the age and gender of every user in our dataset using a machine-learning approach (see Appendix B for method and evaluation). We then matched the users from our group of interest,

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.

i.e., MH-India, with those in control groups (MH-RoW, and Control-RoW) on these two covariates. Owing to the smaller sample size, matching was not performed for the Control-India group. However, 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.

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 Fig. 2). 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 matching.

3 Modeling Cross-Cultural Variations in Mental Health Language

3.1 Language Features

We extracted a diverse set of interpretable language features widely used in Psychology literature to understand the unique markers of mental health among Indians.

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. N-grams may reveal prevalent *cul-

tural idioms of distress (Desai and Chaturvedi, 2017) such as *tension* (Weaver, 2017) that are unique to a culture.

2. LIWC-22 is a closed dictionary comprising 102 **psychosocial categories** to measure cognitive processes in language. 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 to capture themes behind mental distress specific to a culture. We do not use neural models or embeddings due to their skew toward West-centric data and inferences (Havaldar et al., 2023).

We generated three sets of topics by setting the number of topics = $[200, 500, 2000]$. 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 and manual evaluation. Human annotators preferred high topic granularity because it highlights subtle linguistic variations across cultures.

3.2 Correlation Analysis

To measure the association between language and the groups (i.e., MH-India vs MH-RoW), we performed ordinary least squares regressions with the feature sets. We calculated Pearson r to measure the association of each feature to each group in a one-vs-all setting. p-values were corrected using Benjamini-Hochberg correction for multiple hypothesis testing. 23,344 1-3 grams, 102 word categories for LIWC, 2000 LDA topics, were considered for p-value correction.

3.3 Predictive Model

To examine whether the linguistic expressions of the MH-India group sufficiently differ from control groups, we trained 'one vs rest' logistic regression models in a 10-fold cross-validation setting (Rifkin
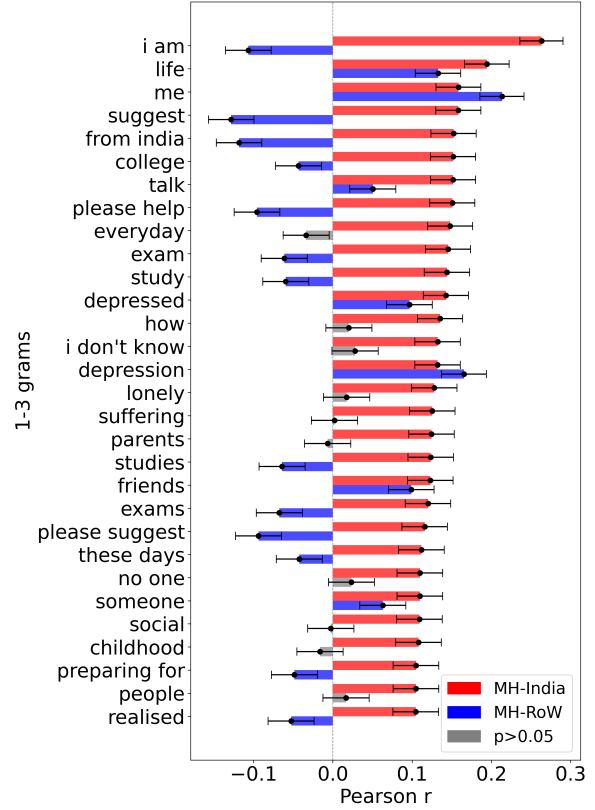

Figure 3: Top 30 statistically significant n-grams by effect size for MH-India and their corresponding correlation with MH-RoW. Significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected. Repetitive phrases (e.g. *life* vs *my life*) and function words are removed. See Fig A1 for top n-grams for MH-RoW.

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 models for estimating mental health risks. 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 Expressions: 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 3 illustrates the top 30 1-3 grams arranged in decreasing order of Pearson r for MH-India. Introductory phrases (*i am, from india*), negation (*don't, don't know*), and help-seeking phrases (*suggest, please*
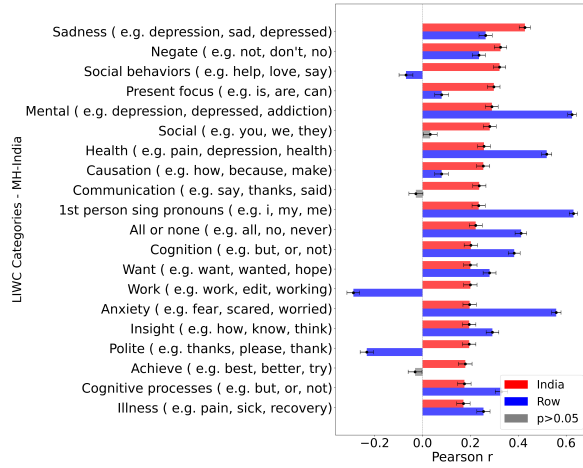
Figure 4: Pearson r for top 20 LIWC categories (with top-3 words) correlated with MH-India with corresponding correlation with MH-RoW. Bars in gray color indicate insignificant correlation ($p > 0.05$). p-values were corrected using Benjamini-Hochberg correction. Sadness, social behaviors, and present focus are the most strongly correlated categories.

Figure 5: Pearson r for top 20 LIWC categories (with top-3 words) correlated with MH-RoW with corresponding correlation with MH-India. Bars in gray color indicate insignificant correlation ($p > 0.05$). p-values were corrected using Benjamini-Hochberg correction. Substances, 1st person sing. pronouns, and Mental are the most strongly correlated categories.

help) are exclusively correlated with MH-India, reflecting the struggle of users in MH-India for accessing mental help. Academic-related stressors (*college, exam, study*) are exclusively seen in discussions in MH-India. This is particularly interesting, considering users in both groups were matched for age, yet discussions around student-life challenges are prevalent exclusively in MH-India. Self-referential pronouns such as *I* and *me*, negative feelings and symptoms mentions are more frequently seen in MH-RoW (see Fig. A1). Overall, the discussion in MH-India subreddits is centered around seeking help, whereas symptoms and diagnosis 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 LIWC categories for MH-India and MH-RoW in Fig 4 and 5. Sadness, a subset of negative emotions is more strongly correlated with MH-India whereas anxiety (also a type of negative emotion) is more correlated with MH-RoW. Mental health-related discourse is more present-focused in India whereas in MH-RoW group, it is past-focused. This aligns with n-grams findings i.e. MH-India's emphasis on seeking help and MH-RoW's on discussing symptoms. Social aspects including the use of 2nd/3rd person pronouns (e.g. you, we), communication,
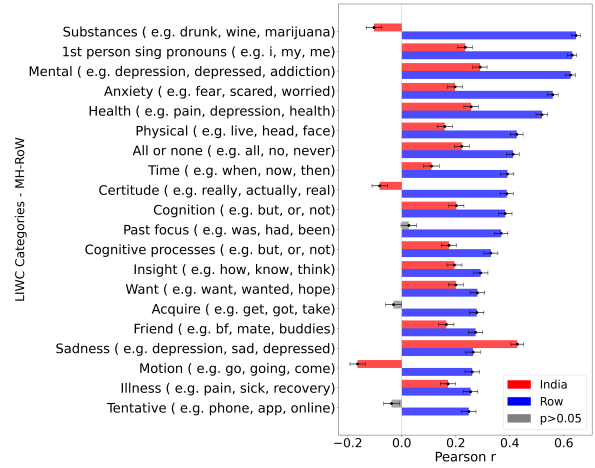
work, politeness, and achievement are positively correlated with only MH-India – reflecting the potential sociocultural expectations (such as collectivism and *high-power* distance in communication (Robert et al., 2000)) in Indian society.

Discussions in MH-RoW use more clinical language with concepts from psychosocial categories such as mental, health, and illness (also observed in Fig. A1) - reflecting awareness about mental health disorders. Substances are the most strongly correlated psychosocial category with MH-RoW however, it has a negative correlation with MH-India. The correlation with 1st person singular pronouns is also three times more in MH-RoW compared to MH-India. This aligns with recent findings that the use of self-referential pronouns is more likely a marker of depression in European Americans compared to other demographic groups (Rai et al., 2024). Similarly, cognitive processes such as all or none (also a type of cognitive distortion (Mercan et al., 2023)) and certitude are more strongly correlated with MH-RoW.

Categories such as *illness* and *friend* are similarly correlated with both groups.

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 Fig. 6. Academic stressors (e.g. *college, exam, study, univer-*
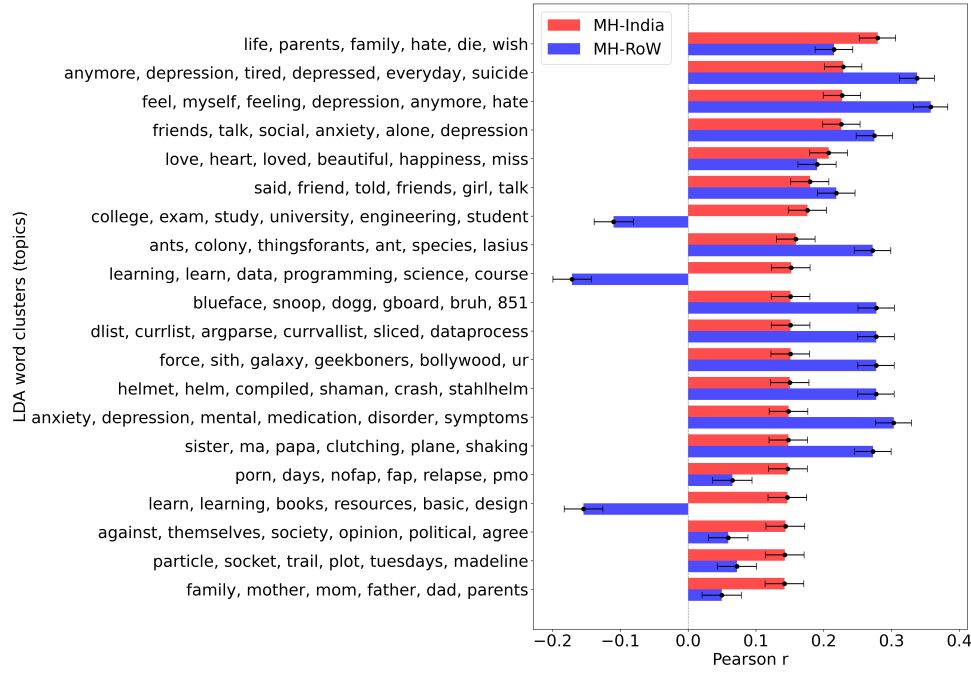
Figure 6: Top 20 LDA topics arranged in decreasing order of correlation with MH-India and its corresponding correlation with MH-RoW are shown. All topics shown are statistically significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected.

sity; *learning, learn, data, programming*) are exclusively related to discussions in MH-India. Family (*family, mother, mom, father*) and sexual content (*porn, days, nofap, fap*) are more frequent in MH-India compared to MH-RoW. Other topics including negative/suicidal thoughts (*life, parents, family, hate, die*), lack of belongingness (*friends, talk, social, anxiety, alone*; *said, friend, told, friends*), depression mentions (*anymore, depression, tired, depressed*; *anxiety, depression, mental, medication*) are overlapping however, they are more strongly correlated with MH-RoW. There are several topics (*ants, colony, thingsforants, ant*; *blueface, snoop, dogg, gboard*, etc,) which appear unrelated to mental health discussion but are discussed in both groups.

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 Prevalent' and '0' indicates 'Not observed at all'.

Prevalence While independently labeling topics

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

Table 2: 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.

4.2 Is the language in MH-India different from control groups?

High AUC scores (See Table 2) 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 mental health expressions at 1-3 gm level, in psychosocial categories as well as in latent topics of discussion.

137

5 Discussion

Our study uncovers significant cultural differences in the way users in the MH-India and MH-RoW groups discuss their mental health, underscoring the influence of sociocultural settings on how people perceive mental disorders and seek help. While MH-RoW discussions predominantly focus on feelings, symptoms, and peer support, MH-India situates their struggles with family, education, and work pressures, and are more likely to seek help or advice on social media platforms. Pendse et al. (2019) also found that Indians discuss "wanting or needing friends" on Mental Health Support Forum more than other countries.

Indian social media users often express immediate hopelessness or sadness with their current situation (Bahri et al., 2023). This differs from the in-person Indian patients having access to healthcare, who may ruminate more on past failures or anticipate future challenges. Similarly, somatization is widely observed among in-person patients in India (Kirmayer and Young, 1998), we, however, observe a weaker correlation for health-related mentions (such as pain, sick, etc.) with MH-India compared to MH-RoW. Social media data may underrepresent depression symptoms, as Indian users often seek advice rather than explicitly expressing distress. AI models trained on social media risk missing key depressive indicators, especially in cultures where emotional struggles are less directly verbalized. For instance, De Choudhury et al. (2017) demonstrated that 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.

Regarding latent themes in Reddit discussions, only 56% of 109 LDA topics correlated with the MH-India group were labeled as *prevalent* in Indian patients by clinical psychologists.

Of Top-20 topics in MH-India which were labeled as "not prevalent" by psychologists comprise *Video Games/Hobbies* and *Programming/Learning*. This indicates the influence of digital content and growing isolation amongst the undiagnosed young population. These topics could be underrecognized concerns.

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[*]. Auto-

mated 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. Based on our findings, culturally competent mental health intervention techniques should seek to bridge the gap to treatment by addressing hypochondriacal ideas and familial embarrassment in particular. 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.

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 shame 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).

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

---

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

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

While Reddit data is public, it 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.

References

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.

Syamsul Bahri, Elisa Betty Manullang, Putri Syah Nadillah Sihombing, and Kevin Enzo Eleazar. 2023. Language change in social media. *Randwick International of Social Science Journal*, 4(3):713–721.

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.

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.

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.

Geetha Desai and Santosh K Chaturvedi. 2017. Idioms of distress. *Journal of neurosciences in rural practice*, 8(S 01):S094–S097.

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.

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

Shreya Havaldar, Sunny Rai, Bhumika Singhal, Langchen Liu, Sharath Chandra Guntuku, and Lyle Ungar. 2023. Multilingual language models are not multicultural: A case study in emotion. In *The 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, page 202.

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

Laurence J Kirmayer and Allan Young. 1998. Culture and somatization: clinical, epidemiological, and ethnographic perspectives. *Psychosomatic medicine*, 60(4):420–430.

Roberto Lewis-Fernandez, Amar K Das, César Alfonso, Myrna M Weissman, and Mark Olfson. 2005. Depression in us hispanics: diagnostic and management considerations in family practice. *The Journal of the American Board of Family Practice*, 18(4):282–296.

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.

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.

Neşe Mercan, Melisa Bulut, and Çiğdem Yüksel. 2023. Investigation of the relatedness of cognitive distortions with emotional expression, anxiety, and depression. *Current Psychology*, 42(3):2176–2185.

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.

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

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.

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.

Christopher Robert, Tahira M Probst, Joseph J Martocchio, Fritz Drasgow, and John J Lawler. 2000. Empowerment and continuous improvement in the united states, mexico, poland, and india: Predicting fit on the basis of the dimensions of power distance and individualism. *Journal of applied psychology*, 85(5):643.

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.

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.

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.

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.

Elizabeth C Stade, Shannon Wiltsey Stirman, Lyle H Ungar, Cody L Boland, H Andrew Schwartz, David B Yaden, João Sedoc, Robert J DeRubeis, Robb Willer, and Johannes C Eichstaedt. 2024. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPJ Mental Health Research*, 3(1):12.

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

Stanley Sue. 1998. In search of cultural competence in psychotherapy and counseling. *American psychologist*, 53(4):440.

Lesley Jo Weaver. 2017. Tension among women in north india: An idiom of distress and a cultural syndrome. *Culture, Medicine, and Psychiatry*, 41(1):35–55.

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.
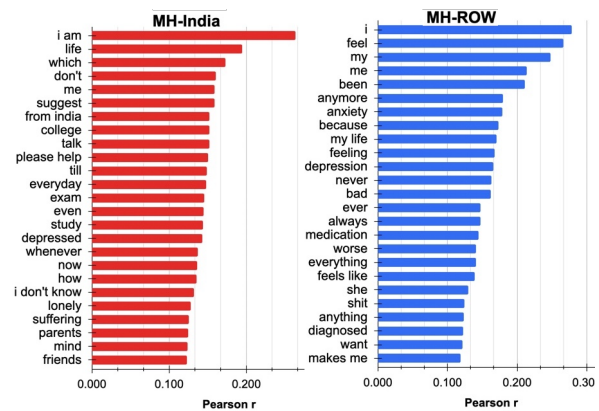

Figure A1: Top 25 1-3 grams in order of decreasing Pearson r for MH-India and MH-RoW.

C Communication with Clinical Psychologists

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

Table A3: Email communication with clinical psychologists who performed an informed review of the topics.

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