Studying Differential Mental Health Expressions in India

Anonymous ACL submission

Abstract

Psychosocial stressors and the symptomatology of mental disorders are known to vary with socio-cultural environment. 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 made on Reddit by individuals in India, the most populous country in the World, to identify psycho-social categories and themes specific to the Indian context compared to Reddit users located in the Rest of the World (ROW), predominantly the United States. Contrary to findings in Western samples, mental health discussions in India are present-focused and are about work and achievement-related topics. Psycho-social category, *illness* is exclusively correlated with mental health posts originating from India, reaffirming the link between somatic symptoms and mental disorders in Indian patients. Two clinical psychologists practicing in India labeled 95% of the top-20 topics associated with mental health discussions as *prevalent* in Indians. Both Psychologists are are female Indian citizens working with patients for over 5 years. Significant linguistic variations in online mental health-related language originating from India vs. ROW, highlight the need for precision culturally-aware machine learning 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). The reasons for these staggering statistics include stigmatization of mental disorders, a shortage of mental healthcare providers, and a lack of awareness of mental health disorders, cumulatively contributing to diagnostic barriers for mental health care (Meshvara, 2002; Lahariya, 2018; Krendl and Pescosolido, 2020).

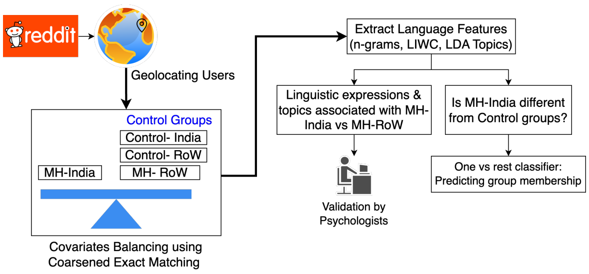

Figure 1: Study overview to investigate the linguistic expressions and topics specific to Indian users posting in Mental Health (MH) related subreddits (See Appendix A) compared to users across the rest of the world (ROW).

Prior works used social media data to identify linguistic markers of depression and other behavioral disorders for automated risk-screening (Chancellor and De Choudhury, 2020; Guntuku et al., 2017; Eichstaedt et al., 2018). Reddit, in particular, provides unique affordances, such as the choice to be anonymous, for individuals to openly discuss their experiences without any character limit and seek support in dedicated threads (De Choudhury and De, 2014; Boettcher, 2021). The use of social media platforms for mental health-related conversations is growing rapidly in India (Akbar et al., 2020). Automated analyses of user-generated content could enable early detection of mental health disorders and facilitate targeted assessments, including support and treatment, especially in under-resourced contexts such as India, alleviating the challenges associated with traditional assessment methods (Organization et al., 2021).

Computational models for detecting mental disorders trained on social media datasets (Coppersmith et al., 2015; Benton et al., 2017) comprise mostly White users (Aguirre et al., 2021). However, studies have shown that variations in mental health expression exist at racial (Rai et al., 2023), gender (Aguirre and Dredze, 2021), and geographic levels (De Choudhury et al., 2017).

In this paper, 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* (See Figure 1 for study overview) by answering the following:

- How do the psychosocial language markers and thematic content 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?
- Can language features extracted from social media posts reliably predict membership in mental health and India-related subgroups?

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 Background

2.1 Mental Health in India

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). Insufficient government funding, limited availability of mental healthcare providers, and cultural taboos are major catalysts behind public mental health crisis in India (Khandelwal et al., 2004; Srivastava et al., 2016; Hossain and Purohit, 2019). Familial struggles are the primary barrier to mental health recovery in India (Biswas et al., 2016). A recent survey comprising 3556 Indian respondents revealed that 71% participants exhibit stigma when answering questions about mental health (Foundation, 2018). Relatedly, somatic symptoms, hypochondriasis, anxiety, and agitation are more commonly seen in Indian patients (Gada, 1982). While the extent of the problem in India regarding mental health stigma and treatment availability is known, our study will contribute to an understanding of the lived experiences of those in India suffering from mental health challenges.

2.2 Social Media and Mental Health

Social media data has significant predictive utility in identifying behavioral health conditions such as depression, anxiety, PTSD, and suicide ideation, among others (see (Guntuku et al., 2017; Chancellor and De Choudhury, 2020) for surveys on this topic). Psychosocial word categories (e.g., LIWC) and topics (word clusters derived using LDA) are commonly used approaches to examine language correlated with depression (G et al., 2017; M et al., 2019). For instance, the increasing use of self-referential pronouns and negative emotions in social media language is known to predict depression (Stamatis CA, 2022). However, there has been growing evidence that these markers vary with demographics (Rai et al., 2023; Aguirre and Dredze, 2021; De Choudhury et al., 2017). Mittal et al. (2023a) found that US immigrant mental health concerns are more aligned with conversations around race, politics, violence, employment, and affordability of day-to-day expenses such as rent than their non-immigrant counterparts in the US. De Choudhury et al. (2017) found that a group of India and South Africa-based users to be less candid in their posts and tend to regulate negative emotions in comparison to their Western counterparts. Another study looked at Indian, Malaysian, and Filipino users on Mental Health Support Forums such as Talklife (Pendse et al., 2019) and found that Indians discuss "wanting or needing friends" more than other countries. Recent cross-cultural studies provide extensions in the field of effective cultural comparisons using social media data but lack analysis focused on India or the Indian diaspora(De Choudhury et al., 2017; Pendse et al., 2019).

Prior work by De Choudhury et al. (2017) and Pendse et al. (2019) is the closest to this study. The aggregation of Indian and South African self-disclosure tweets in De Choudhury et al. (2017) limits understanding the cultural nuances of each country. Pendse et al. (2019) focused on identifying differences in clinical language across countries (India, Malaysia, The Philippines vs. USA, Canada, and the UK) on online mental health forums; however, not having an explicit control group (discussions about non-mental health-related topics) leaves more to be explored in terms of the breadth of the mental health language markers vs. culture-specific language markers.

In this paper, we compare the language from the entire Reddit timelines of individuals geolocated to India and who also post in mental health subreddits with that of a coarsened-exact matched control set consisting of (a) Indians who post in non-mental health subreddits and (b) individuals from other (mostly Western, see Fig. 2 for distribution) countries who post in mental health subreddits, and (c) in non-mental health subreddits. This provides an opportunity to obtain language markers associated with the mental health challenges of Indians going above and beyond the colloquial usage of terms within India and contrasting the mental health expressions of individuals outside.

3 Data

The most widely used social media platforms in India are Whatsapp, Instagram, and Facebook. However, these platforms do not offer APIs for data collection that are accessible to the same level as Reddit. Beyond the data accessibility for research purposes, Reddit offers a platform for individuals to share their mental health journey and seek support anonymously. 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. At the end of 2022, India ranked 5th in Reddit's website traffic globally with 240 million Indian users (Semrush).

3.1 Subreddits: Mental Health vs Control

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

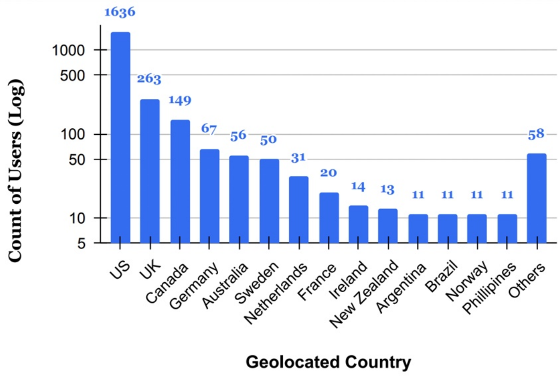

Figure 2: The count of users for each country in the Rest of World control group (log scale). Demonstrates that the large majority of users in the ROW group are geolocated to Western countries. The "Others" Category contains countries with less than 10 users, including Belgium (9), Italy (9), Mexico (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).

- 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.

3.2 Geolocation - India vs ROW

We used the geolocation inference approach introduced by 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 2), 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.

3.3 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.

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.

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.

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.

4 Methods

4.1 Language Features

We extracted three sets of language features as described below:

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 accu-

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

racy, 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 A1) and manual evaluation.

4.2 Statistical 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). In this regression, the feature sets were independent variables. Each of the four groups (MH-India, MH-ROW, Control-India, and Control-ROW) was a one-hot encoded dependent variable. 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. 102 word categories for LIWC, 2000 for LDA topics, and 23,344 1-3 grams were considered for p-value correction.

4.3 Thematic Annotations

Topics are clusters of semantically connected words that need to be contextualized for further analysis. However, determining the *theme* from the lens of mental disorders demands expert knowledge.

Language models such as GPT-4 are increasingly discussed as potential alternatives of human experts for data annotation (Gilardi et al., 2023). However, there is a growing debate against whose perspective these language models represent when labeling texts that require cross-cultural knowledge (Havaldar et al., 2023; Atari et al., 2023). Our dataset, comprising multicultural discussions from users from all over the world, thus presents a unique testbed to assess the capability of language models in labeling the themes of behavioral health discussions. We generated thematic annotations using ChatGPT (OpenAI, 2021) for each significant LDA topic for our group of interest (MH-India) based on top words (See Table A3 for prompt). Two clinical psychologists practicing in India were asked (See Table A2 for annotation guidelines):

1. *To what extent a given topic (cluster of topic words) is prevalent in Indian patients?* - A Likert scale of 0-5 is provided where 5 indicates 'Highly Prevalant' and '0' indicates 'Not observed at all'.

2. *Does the machine-generated thematic label accurately capture the meaning of topic words?* The evaluators could mark *Yes, No* or *Unsure*. If *no* is selected, the evaluators were further prompted to suggest the correct label.

4.4 Predictive Model

To examine whether the language features 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. We report the Area Under the Receiver Operating Curves (AUC) for each feature for the MH-India and MH-ROW groups.

5 Results

5.1 Mental Health Expression: India vs ROW

5.1.1 N-grams

Out of the 23,344 unique 1-3 grams in our data, a total of 61 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 25 1-3 grams arranged in decreasing order of Pearson r for both groups. Personal 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 is more cognisant of their feelings, treatment ('diagnosed', 'medication') and commonly express negative emotions ('anxiety',

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
Linguistic	Auxiliary Verbs	0.198	is, have, was, be, are		Adverbs	0.533	so, just, about, there, when
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 20 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.

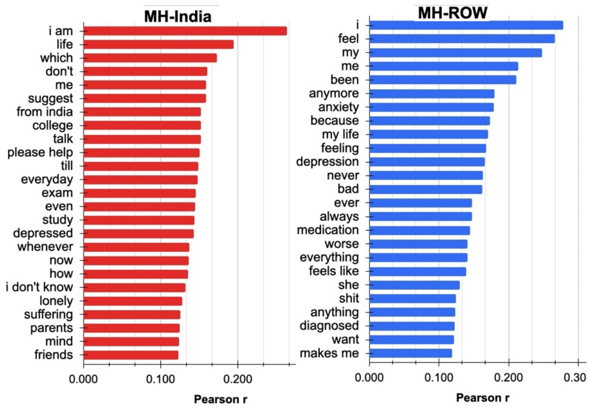

Figure 3: Top 25 statistically significant N-grams by effect size for both MH-India and MH-ROW. Significant at $p < .001$, two-tailed t-test, Benjamini-Hochberg corrected. Repeated N-grams are omitted.

'bad', '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 commonly discussed in the MH-ROW group.

5.1.2 LIWC

52 LIWC categories were significantly associated ($p < 0.05$) with the MH-India group, whereas 60 categories were found to be correlated with the MH-ROW group. We provide the Top 20 LIWC categories for both groups in Table 2. Negative tone, anxiety, and personal pronouns are correlated with depression in both groups. Past focus, a widely associated marker of depression, is exclusive to the MH-ROW group. Instead, discussions in the MH-India group are present-focus. Social behavioral attributes (*communication, politness*), work, and achievement are correlated with depression in the MH-India group but not in MH-ROW. *Sadness*, an additional affect is seen in the MH-Indian group along with negative tone and anxiety. Somatic symptoms/illness (pain, sick) are seen in the MH-India group, whereas substance abuse/addiction is correlated with the MH-ROW group.

5.1.3 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-20 topics and the corresponding Pearson *r*, p-values, and 95% confidence intervals are provided in Table 3. The most prevalent topics in the MH-India group discussed family struggles (*life, parents, family, hate, die*), academic and job stressors ((*college, exam, study, university, engineering*), *job, degree, college, school, career*) and relationships (*love, heart, loved, beautiful, happiness*). In contrast, the prevalent themes in MH-ROW are feelings (*feel, myself, feeling, depres-*

MH-India						MH-ROW		
Topic#	ChatGPT Label	Corrected Label	Top Words	Prevalent?	r	Topic#	Top Words	r
1807	struggling with mental health and suicidal thoughts		life, parents, family, hate, die	Yes	0.280	334	feel, myself, feeling, depression, anymore	0.358
730	*battling depression and seeking help*	battling depression, giving up any kind of help	anymore, depression, tired, depressed, everyday	Yes	0.229	501	went, didn, crying, mad, stayed	0.340
334	***dealing with loneliness and anxiety***	Experiencing feelings of loneliness and isolation	feel, myself, feeling, depression, anymore	Yes	0.227	730	anymore, depression, tired, depressed, everyday	0.338
1531	**struggling with social anxiety and loneliness**		friends, talk, social, anxiety, alone	Yes	0.226	1642	feeling, body, feels, heart, scared	0.322
560	**complex emotions related to love**		love, heart, loved, beautiful, happiness	Yes	0.207	375	sick, woke, stomach, switched, asleep	0.318
1872	navigating friendships and relationships		said, friend, told, friends, girl	Yes	0.180	851	didn, don, wasn, couldn, re	0.308
1757	*university life and studying*	University life, studies and academic pressure	college, exam, study, university, engineering	Yes	0.176	758	anxiety, depression, mental, medication, disorder	0.303
1221	*miscellaneous electronics and science-related topics*	learning programming and coding	learning, learn, data, programming, science	No	0.152	439	ve, don, re, ll, doesn	0.293
758	Mental health and related issues		anxiety, depression, mental, medication, disorder	Yes	0.148	1412	said, didn't, friend, told, asked	0.287
595	*Emotional goodbye to a family member*	Family Relationships	sister, ma, papa, clutching, plane	Yes	0.148	453	dad, mom, broke, suicide, crying	0.286
270	Discussion of pornography addiction and recovery		porn, days, nofap, fap, relapse	Yes	0.147	1262	buy, save, buck, cheap, bang	0.284
1736	*Learning resources and tutorials for design*	Learning resources and tutorials	learn, learning, books, resources, basic	No	0.147	1531	friends, talk, social, anxiety, alone	0.275
1492	Societal views, opinions, and political arguments		against, themselves, society, opinion, political	Yes	0.144	595	sister, ma, papa, clutching, plane	0.273
899	Family members and their relationships.		family, mother, mom, father, dad	Yes	0.142	1010	sooner, handful, crossed, figuring, span	0.262
1326	**Education and career paths**		job, degree, college, school, career	Yes	0.140	923	anime, manga, series, watched, japanese	0.260
439	Expressing Dislike and Judgments		ve, don, re, ll, doesn	Yes	0.140	1747	damage, level, attack, weapon, hit	0.231
832	***Negative attitudes towards others***	Negative attitudes and emotions towards others	hate, angry, rant, respect, ugly	Yes	0.139	1180	donate, donation, charity, donations, donating	0.229
1642	**Coping with fear and anxiety**		feeling, body, feels, heart, scared	Yes	0.133	1446	fucking, shit, fuck, hate, ass	0.229
419	Relationships and dating		relationship, together, wants, we've, ex	No	0.116	708	weeks, october, wednesday, waited, knocked	0.223
549	Human mind and spirituality		mind, human, universe, reality, self	No	0.122	832	hate, angry, rant, respect, ugly	0.222

Table 3: Top 20 topics and their top words by frequency for MH-India and MH-ROW are shown. All topics shown are statistically significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected. ChatGPT Labels in bold indicate "extremely prevalent" (i.e. a prevalence score of 5) topics and italicized text indicates labels marked incorrect by clinical psychologists.

sion, anymore) and negative emotions ((*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.

Prevalence While independently labeling topics 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.

Quality of ChatGPT generated Thematic Annotations 54% of ChatGPT-generated thematic annotations were marked as correct by both clinical psychologists, whereas 87.15% of thematic annotations were labeled as correct by at least one of the clinical psychologists. The labels predicted by ChatGPT for top 20 topics are provided in Table 3. Incorrect labels are italicized and alternative labels are also provided.

	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.

5.2 Predictive Modeling

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, with the highest performing feature being n-grams.

6 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).

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 LLMs generated thematic summaries for LDA topics were coherent, and more importantly, quick to obtain compared to traditional annotation, which requires annotators to find the relationship between keywords, which is time-consuming and requires expert domain knowledge.

Significant linguistic variations exist in the mental health-related language in social media posts by Indians compared to individuals from the rest of the world. Recent studies have indicated language variation behind the underperformance of mental health models on persons of color (Aguirre et al., 2021; Rai et al., 2023). These findings emphasize the need for socio-culturally aware mental health models to prevent misdiagnosis.

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. The language markers of depression vary across cultures and demographics, and our study affirms the urgency to culturally adapt healthcare technologies to prevent misdiagnosis and deliver inclusive care.

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

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.

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 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. Finally, while utilizing OpenAI models for labeling topics, only topic keywords were inputted into ChatGPT's interface, barring any individual identifying information. Furthermore, we were careful not to make inferences on the topics with ChatGPT's labeling alone, relying on a discussion of the labels suggested by mental health professionals with experience in the Indian subcontinent as a basis of comparison and analysis. 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).

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.

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*.

Mohammad Atari, Mona J Xue, Peter S Park, Damián Blasi, and Joseph Henrich. 2023. Which humans?

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.

Adrian Benton, Margaret Mitchell, Dirk Hovy, et al. 2017. Multitask learning for mental health conditions with limited social media data. In *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017. Proceedings of Conference*. Association for Computational Linguistics.

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.

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

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.

Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. Clpsych 2015 shared task: Depression and ptsd on twitter. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 31–39.

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.

Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, Chua T-S, and Zhu W. 2017. " depression detection via harvesting social media: a multimodal dictionary learning solution.". *IJCAI*.

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.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd-workers for text-annotation tasks. *arXiv preprint arXiv:2303.15056*.

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*.

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.

Md Mahbub Hossain and Neetu Purohit. 2019. Improving child and adolescent mental health in india: Status, services, policies, and way forward. *Indian Journal of Psychiatry*, 61(4):415.

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

Sudhir K Khandelwal, Harsh P Jhingan, S Ramesh, Rajesh K Gupta, and Vinay K Srivastava. 2004. India mental health country profile. *International review of psychiatry*, 16(1-2):126–141.

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.

Matero M, Idnani A, Son Y, Giorgi S, Vu H, Zamani M, Limbachiya P, Guntuku S, and Schwartz HA. 2019. "suicide risk assessment with multi-level dual-context language and bert.". *Proceedings of the sixth workshop on computational linguistics and clinical psychology*.

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.

OpenAI. 2021. Chatgpt. Accessed on May 15, 2023.

World Health Organization et al. 2021. Ethics and governance of artificial intelligence for health: Who guidance.

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*.

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.

2023 Semrush. "reddit traffic stats 2023". "considered as an approximation.".

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.

Kalpana Srivastava, Kaushik Chatterjee, and Pookala Shivaram Bhat. 2016. Mental health awareness: The indian scenario. *Industrial psychiatry journal*, 25(2):131.

Liu T Sherman G Wang H Liu T Curtis B Ungar LH Mohr DC Stamatis CA, Meyerhoff J. 2022. "prospective associations of text-message-based sentiment with symptoms of depression, generalized anxiety, and social anxiety". *Depression and Anxiety*.

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.

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.

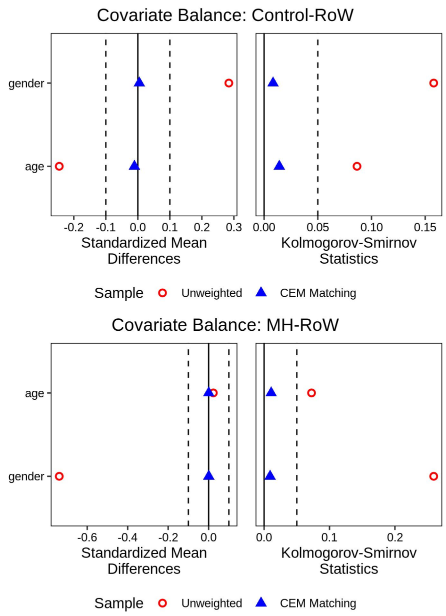

Figure 4: Differences in Covariates before and after CEM for groups "Control-ROW" and "MH-ROW"

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 A1: Topic Uniqueness Scores for different numbers of topics and message types.

D Topic Uniqueness Scores

Table A1 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.

E Communication with Clinical Psychologists

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

F Prompt for Topic Labeling

Table A3 shows the prompt used for topic labeling.

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. • Read the topic label provided in Column B, which is intended to capture the essence of the topic corresponding to these top 10 words. **This label is not given by a psychologist.** • Select "Yes," "No" or unsure in Column C, based on whether you believe the topic label provided is coherent with the topic/theme suggested by the topwords • Select "Unsure" only if the topic given doesn't make sense to you (eg. illegible words, unrelated words) • If you selected "No," please suggest a new topic label in Column D that you believe will better capture the given 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 A2: Email communication with clinical psychologists who performed an informed review of the topics and ChatGPT-generated labels in this study.

Prompt: In each row, please find the relationship between words and conclude a topic with one short phrase. Examples: 1 <feel, myself, feeling, depression, anymore, hate, depressed, anxiety, alone, worse > - *struggling with loneliness and anxiety*. 2 <love, heart, loved, beautiful, happiness, miss, sad, joy, sadness, forever> - *mixed emotions*

Table A3: Prompt to generate thematic annotations for a given set of topics. Two examples (2-shot learning) along with the prompt were provided as listed above.