A Survey on Multilingual Mental Disorders Detection from Social Media Data

Anonymous ACL submission

Abstract

The increasing prevalence of mental health disorders globally highlights the urgent need for effective digital screening methods that can be used in multilingual contexts. Most existing studies, however, focus on English data, overlooking critical mental health signals that may be present in non-English texts. To address this important gap, we present the first survey on the detection of mental health disorders using multilingual social media data. We investigate the cultural nuances that influence online language patterns and self-disclosure behaviors, and how these factors can impact the performance of NLP tools. Additionally, we provide a comprehensive list of multilingual data collections that can be used for developing NLP models for mental health screening. Our findings can inform the design of effective multilingual mental health screening tools that can meet the needs of diverse populations, ultimately improving mental health outcomes on a global scale.

1 Introduction

It is estimated that nearly half of the population will develop at least one mental disorder by the age of 75 (McGrath et al., 2023). Unfortunately, many people do not seek psychiatric help for mental health issues due to stigma, which manifests itself differently between cultures and is influenced by different cultural norms, religious beliefs and social attitudes (Ahad et al., 2023). Due to the stigma associated with mental health and the limited access to professional care around the world, the World Health Organization (WHO) advocates for improved delivery of mental health services, including digital technologies to deliver remote care.[1] There is a pressing need for the integration of remote screening tools and the delivery of culturally adapted digital interventions (Bond et al., 2023). Remote screening relies on processing language patterns associated with mental disorders, which can be identified from short essay writing (Rude et al., 2004), text messages (Nobles et al., 2018), or social media (Eichstaedt et al., 2018).

The first well-known study on the detection of mental disorders using social media was conducted by De Choudhury et al. (2013). Multiple other studies have shown that the language used on Facebook can predict future depression diagnoses found in medical records, indicating that social media data could serve as a valuable complement to depression screening (Eichstaedt et al., 2018). The current methods used for social media screening focus mainly on English data (Skaik and Inkpen, 2020; Harrigian et al., 2021). Additionally, there have been multiple workshops and shared tasks addressing NLP applications to mental health primarily on English data such as eRisk (Parapar et al., 2024), CLPsych (Chim et al., 2024) and LT-EDI (Kayalvizhi et al., 2023).

There are important limitations in current NLP models when processing multilingual mental health-related data. Various studies analyzing English data from social media have shown that there are cultural differences in online language markers of mental disorders (De Choudhury et al., 2017; Aguirre and Dredze, 2021; Rai et al., 2024) and that the NLP models used for detection do not generalize on data from non-Western cultures (Aguirre et al., 2021; Abdelkadir et al., 2024). Even one of the best predictors of depression in language, the use of the first person pronoun "I" (Rude et al., 2004), for example, has different degrees of association with the severity of depression across different demographic groups (Rai et al., 2024). This suggests that markers of mental disorders in social media language are not universal. One reason for this variation is that self-disclosure rates differ between cultures; collectivist cultures tend

[1] https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care

to have lower self-disclosure rates than individualist cultures in online settings (Tokunaga, 2009). Furthermore, non-native English speakers tend to use their native language for more intimate self-disclosures on social media, with higher rates of negative disclosure compared to posts in English (Tang et al., 2011). This could have substantial implications for English-based social media screening tools, as they can overlook important signals of mental health disorders that are present in posts that are not written in English.

Recently, there have been efforts to develop detection models that focus on languages other than English, such as Portuguese (Santos et al., 2024), German (Zanwar et al., 2023), Arabic (Almouzini et al., 2019), and Chinese (Zhu et al., 2024). There have also been shared tasks specifically designed to address these issues, such as MentalRiskES (Mármol-Romero et al., 2023), which focuses on the early detection of depression, suicide, and eating disorders in Spanish. To further contribute to these important efforts, we present the first survey on mental disorders detection from multilingual social media data. This survey aims to promote the development of multilingual NLP models that take into account cross-cultural and cross-language differences in online language.

This paper makes the following **contributions**:

1. We investigate cross-cultural and cross-language differences in the manifestations of mental disorders in social media.

2. We provide a comprehensive list of multilingual mental health datasets that capture linguistic diversity and can be used for developing multilingual NLP models.[2]

3. We identify and describe several research gaps and future directions in the detection of multilingual mental disorders using online data.

2 Related Surveys

In this section, we analyze related surveys on the analysis of mental disorders from social media data. Calvo et al. (2017) is considered one of the first comprehensive surveys, presenting the datasets and NLP techniques used for mental health status detection and intervention. The survey explores research on various mental health conditions and states, including depression, mood disorders, psychological distress, and suicidal ideation, specifically in non-clinical texts such as user-generated content from social media and online forums. Similarly, recent surveys from Skaik and Inkpen (2020); Harrigian et al. (2021); Ríssola et al. (2021); Zhang et al. (2022); Garg (2023); Bucur et al. (2025) present the datasets, features, and models used to detect mental disorders from online content, focusing mainly on English language data. In addition to these surveys, Chancellor and De Choudhury (2020) provides a critical review of the study design and methods used to predict mental health status, along with recommendations to improve research in this field. Dhelim et al. (2023); Bucur et al. (2025) focus on studies that were published during the COVID-19 pandemic. It focuses on general mental well-being, loneliness, anxiety, stress, PTSD, depression, suicide, and other mental disorders.

Our paper fills an important gap in the literature by offering the first comprehensive survey of research on detecting mental disorders in languages other than English. The most related survey to ours is the one by Garg (2024), which focuses exclusively on low-resource languages. Our survey, however, has a broader scope as it discusses work on many languages irrespective of their resourcefulness.

3 Mental Disorders Detection Tasks Overview

In this section, we discuss the most common tasks related to predicting mental health disorders. When available, we include references to studies that focus on languages other than English. The prediction of mental health issues through social media is typically approached as a supervised classification task (Figure 1). The most common focus is on the **binary classification** of mental disorders. In this process, a collection of social media posts is used to train an NLP model, which then predicts a binary label that indicates the presence or absence of a mental disorder. Binary classification can be performed at the post-level, which is often used to predict suicidal ideation (Huang et al., 2019) and depression (Uddin et al., 2019). However, relying solely on a single post for decision making can lead to inaccurate predictions. Therefore, predictions can be made at the user level to detect conditions like depression (Hiraga, 2017), anxiety (Zarate et al., 2023), bipolar disorder (Sekulić et al., 2018), etc. Binary classification at the user level

[2]We make the list available online upon publication, and we will continuously update it.

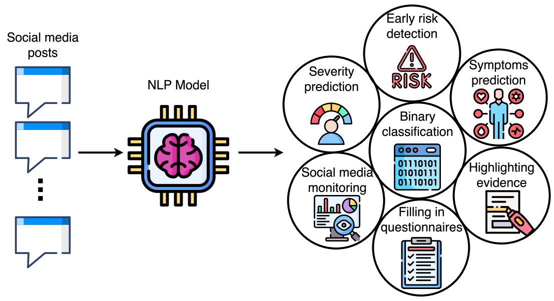

Figure 1: Overview of tasks related to detecting mental health problems from social media.

can also be modeled as an **early risk prediction task**, which aims to accurately label users as soon as possible, allowing the model to make a prediction or wait for more data before deciding (Losada and Crestani, 2016; Parapar et al., 2021).

Another important task is **severity prediction**, which can be modeled either as an ordinal regression / classification task or as a multiclass classification task. It is used primarily to predict the severity of depression (Naseem et al., 2022; Kabir et al., 2023; Sampath and Durairaj, 2022) or the risk of suicide attempts (Benjachairat et al., 2024). Social media posts can be modeled longitudinally to detect **moments of change** in the mental health status of individuals. These shifts or escalations in mood can be used as a warning signal for potential suicidal behavior (Tsakalidis et al., 2022b).

There are tasks designed to improve the explainability of the field, such as **symptom prediction** for mental disorders (Liu et al., 2023; Yadav et al., 2020). Another step toward improving the explainability of model predictions is **highlighting evidence** for mental disorders (Chim et al., 2024; Varadarajan et al., 2024). Mental health indicators from the social media timeline of an individual can be used to **fill in validated questionnaires**, with the goal of estimating symptoms of mental disorders that are usually assessed through survey-based methods such as the Beck's Depression Inventory-II (BDI-II)[3] for depression assessment (Parapar et al., 2021) or the Eating Disorder Examination Questionnaire (EDE-Q)[4] for eating disorders (Parapar et al., 2024).

Finally, **mental health monitoring** aggregated results from detection systems can be used to estimate the prevalence of mental disorders within a population. This approach was used during the COVID-19 pandemic to assess mental health burden, with results comparable to traditional survey-based methods (Cohrdes et al., 2021).

4 Shared Tasks

Shared tasks have encouraged interdisciplinary collaborations between psychologists and computer scientists, resulting in systems that help detect mental disorders through social media analysis. These shared tasks have provided benchmark datasets that the research community continues to use, even beyond the official competitions.

The Early Detection of Mental Disorders Risk in Spanish (**MentalRiskES**) is the only shared task focused on detecting mental disorders in languages other than English. MentalRiskES includes tasks such as the detection of depression, anxiety, eating disorders, and suicidal risk in the Spanish language (Mármol-Romero et al., 2023).

Other shared tasks are focused only on social media data in English. The Early Risk Prediction on the Internet Lab (**eRisk**) is an annual competition focusing mainly on the early detection of mental disorders, including depression, self-harm, pathological gambling, and eating disorders (Parapar et al., 2024). The Workshop on Computational Linguistics and Clinical Psychology (**CLPsych**) includes various tasks, such as detecting depression and PTSD (Coppersmith et al., 2015), labeling crisis posts (Milne et al., 2016), and identifying moments of change (Tsakalidis et al., 2022a) The Workshop on Language Technology for Equality, Diversity, and Inclusion (**LT-EDI**) organized tasks for predicting the severity of depression (Kayalvizhi et al., 2023).

5 Methodology

To identify datasets for modeling the manifestations of mental disorders in languages other than English, we conducted a systematic search

[3] https://naviauxlab.ucsd.edu/wp-content/uploads/2020/09/BDI21.pdf

[4] https://www.corc.uk.net/media/1273/ede-q_quesionnaire.pdf

Language	Resource	Datasets
Arabic	High	Almouzini et al. (2019); Alghamdi et al. (2020); Alabdulkreem (2021); Musleh et al. (2022), CairoDep (El-Ramly et al., 2021), Almars (2022); Maghraby and Ali (2022); Baghdadi et al. (2022), Arabic Dep 10,000 (Helmy et al., 2024), Al-Haider et al. (2024); Abdulsalam et al. (2024); Al-Musallam and Al-Abdullatif (2022)
Chinese	High	Zhang et al. (2014); Huang et al. (2015); Cheng et al. (2017); Shen et al. (2018); Wu et al. (2018); Cao et al. (2019); Wang et al. (2019); Peng et al. (2019); Huang et al. (2019); Li et al. (2020), WU3D (Wang et al., 2020), Yao et al. (2020); Yang et al. (2021); Chiu et al. (2021); Sun et al. (2022); Cai et al. (2023); Li et al. (2023); Guo et al. (2023); Wu et al. (2023); Lyu et al. (2023); Yu et al. (2023); Zhu et al. (2024)
French	High	Tabak and Purver (2020)
German	High	Cohrdes et al. (2021); Baskal et al. (2022); Tabak and Purver (2020), SMHD-GER (Zanwar et al., 2023)
Japanese	High	Tsugawa et al. (2015); Hiraga (2017); Niimi (2021); Cha et al. (2022); Wang et al. (2023)
Spanish	High	Leis et al. (2019), SAD (López-Úbeda et al., 2019), Valeriano et al. (2020); Ramírez-Cifuentes et al. (2020, 2021); Villa-Pérez et al. (2023), MentalRiskES (Romero et al., 2024), Cremades et al. (2017); Coello-Guilarte et al. (2019)
Brazilian Portuguese	Mid to High	von Sperling and Ladeira (2019); Mann et al. (2020); Santos et al. (2020); de Carvalho et al. (2020), SetembroBR (Santos et al., 2024), Mendes and Caseli (2024); Oliveira et al. (2024)
Dutch	Mid to High	Desmet and Hoste (2014, 2018)
Code-Mixed Hindi-English	Mid to High	Agarwal and Dhingra (2021)
Italian	Mid to High	Tabak and Purver (2020)
Korean	Mid to High	Lee et al. (2020); Park et al. (2020); Kim et al. (2022b,a); Cha et al. (2022)
Polish	Mid to High	Wołk et al. (2021)
Russian	Mid to High	Stankevich et al. (2019); Baskal et al. (2022); Narynov et al. (2020); Stankevich et al. (2020); Ignatiev et al. (2022)
Turkish	Mid to High	Baskal et al. (2022)
Bengali	Mid	Uddin et al. (2019); Victor et al. (2020); Kabir et al. (2022); Tasnim et al. (2022), BanglaSPD (Islam et al., 2022), Ghosh et al. (2023); Hoque and Salma (2023), BSMDD (Chowdhury et al., 2024)
Indonesian	Mid	Oyong et al. (2018); Yoshua and Maharani (2024)
Filipino	Mid	Tumaliuan et al. (2024); Astoveza et al. (2018)
Greek	Mid	Stamou et al. (2024)
Hebrew	Mid	Hacohen-Kerner et al. (2022)
Roman Urdu	Mid	Rehmani et al. (2024); Mohmand et al. (2024)
Thai	Mid	Katchapakirin et al. (2018); Hemtanon and Kittiphattanabawon (2019); Kumnunt and Sornil (2020); Hemtanon et al. (2020); Wongaptikaseree et al. (2020); Hämäläinen et al. (2021); Mahasiriakalayot et al. (2022); Boonyarat et al. (2024); Benjachairat et al. (2024)
Cantonese	Low	Gao et al. (2019)
Norwegian	Low	Uddin et al. (2022); Uddin (2022)
Sinhala	Rare	Rathnayake and Arachchige (2021), EmoMent (Atapattu et al., 2022), Herath and Wijayasiriwardhane (2024)

Table 1: Available multilingual datasets for detecting mental disorders.

on major publication databases, including ACL Anthology, ACM Digital Library, IEEE Xplore, Springer Nature Link, ScienceDirect, and Google Scholar. Initially, 405 studies were identified through database searches. After screening the abstracts, 215 papers were excluded because they did not mention the language of the data, or mention that the data is in English. Thus, following a review of the main body of the papers, the number of eligible studies was narrowed down to 108, which represents the final count of papers presenting datasets. Papers that did not present new data collections in languages other than English were excluded during the screening process. The PRISMA flow diagram for the survey is presented in Figure 3 in the Appendix.

6 Multilingual Datasets

The languages most frequently represented in the data collections are three high-resource languages: Chinese, Arabic, and Spanish. Although approximately half of the datasets were published in unranked venues, leading to low visibility for the research, the other half were published in high-ranking journals and conferences (Figure 4 in Appendix A.

6.1 Data Sources

Most of the datasets in English are sourced from Twitter[5] and Reddit (Harrigian et al., 2021). Most non-English datasets in this section were also primarily collected from Twitter. However, Reddit was not as widely used for these data collections in non-English contexts. People use social media platforms differently. Twitter provides community and safety, helping raise awareness and combat stigma around mental health (Berry et al., 2017). In contrast, Reddit allows for greater anonymity with "throwaway" accounts, encouraging users to openly share their experiences in detailed posts on specific subreddits (De Choudhury and De, 2014). This longer format supports post-level mental health analysis (Chowdhury et al., 2024), while Twitter's shorter posts favor user-level insights, requiring longitudinal data to identify patterns (Tumaliuan et al., 2024). The data presented in this survey come from various populations and regions, and some of the sources are platforms that are exclu-

[5]All the datasets were collected before Twitter changed its name to X, so we refer to it as 'Twitter' in this paper.

sive to specific countries, such as Sina Weibo[6] used in China, VKontakte[7] used in Russia, Pantip[8] in Thailand, or Everytime[9] in Korea.

6.2 Languages

Table 1 presents all the datasets with multilingual data. A more detailed version of the table can be found in Appendix A, Table 2. For classifying resource types, we used the framework proposed by Joshi et al. (2020). Figure 4 illustrates that most of the languages used in the data collections belong to some of the largest language families by number of speakers, specifically the Indo-European, Sino-Tibetan and Afro-Asiatic language families. The languages most frequently represented in the data collections are high-resource languages: Chinese appears in 25 data collections, Arabic is found in 11 datasets, and Spanish is included in 10 datasets. Even if most of the languages covered in the data are from high-, mid to high- and mid-resourced languages, we also have some languages with fewer resources, such as Cantonese and Norwegian. The Cantonese data collection was gathered by Gao et al. (2019) from Youtube comments and annotated for the risk of suicide. The Norwegian datasets related to depression were collected from a public online forum in Norway (Uddin et al., 2022; Uddin, 2022). Sinhala language, which was classified as rare by Joshi et al. (2020) is represented in three research papers. One of the papers contains Facebook data annotated for suicide ideation (Herath and Wijayasiriwardhane, 2024), while another contains depression-related data from Twitter and Facebook (Rathnayake and Arachchige, 2021). The third dataset contains data from Facebook, with more fine-grained labeled data on the presence of mental illness, anxiety, suicidal ideation, emotions, psychosomatic symptoms, and other manifestations (Atapattu et al., 2022).

6.3 Mental Disorders

Figure 2 shows the distribution of mental disorders in different languages within the datasets. Depression is the most common mental disorder and is well-represented in the data. The languages that lack data on depression are Cantonese, Dutch, Hebrew, Hindi, and Turkish. Suicide is another mental disorder that frequently appears in collections. In contrast, the mental health problems that are least represented include eating disorders, obsessive-compulsive disorder (OCD), attention deficit / hyperactivity disorder (ADHD), autism spectrum disorder (ASD), anxiety, bipolar disorder, and schizophrenia.

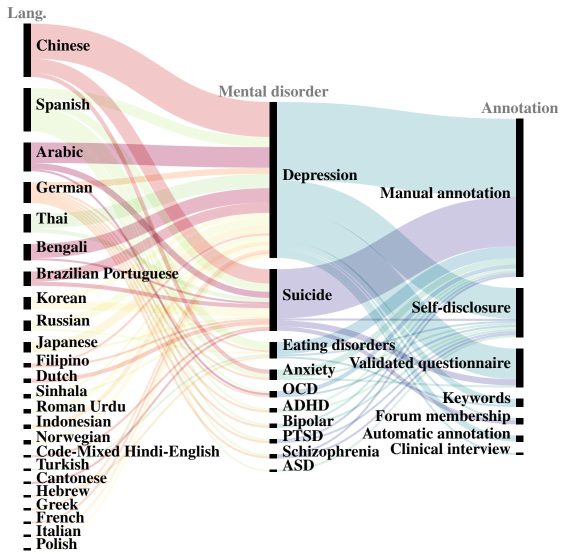

Figure 2: Overview of the mental disorders addressed in each dataset, along with the annotation procedures.

6.4 Annotation Procedure

Most data collections were manually annotated (Figure 2). Manual annotation was carried out by mental health experts or psychologists (Narynov et al., 2020; de Oliveira et al., 2022), graduate students who are native speakers of the language of interest (Boonyarat et al., 2024; Uddin et al., 2019), or nonexpert individuals. However, some datasets do not specify who the annotators were or what guidelines they followed during the annotation process. Most datasets that collect user-level data from online platforms rely on the self-disclosure of mental health statuses. For example, they rely on explicit mentions of diagnoses (e.g. "I was diagnosed with depression") (Tabak and Purver, 2020; Villa-Pérez et al., 2023). The third most common annotation method involves asking social media users to complete validated questionnaires to diagnose mental disorders. The most frequently used survey-based methods include the CES-D (Tsugawa et al., 2015; Lyu et al., 2023), BDI-II (Sun et al., 2022; Stankevich et al., 2019; Ignatiev et al., 2022) or tools specifically designed for certain populations, such as the TMHQ[10] (Katchapakirin et al., 2018).

[6]https://weibo.com
[7]https://vk.com/
[8]https://pantip.com/
[9]https://everytime.kr/

[10]Thai Mental Health Questionnaire

Another reliable annotation approach is conducting clinical interviews to assess mental health problems (Wołk et al., 2021). Less common and noisier annotation methods include identifying posts based on the presence of specific keywords (López-Úbeda et al., 2019), by forum membership (Agarwal and Dhingra, 2021), or automatic annotation through another model trained on mental health data (Cohrdes et al., 2021).

6.5 Availability of Data Collections

Of the 108 datasets listed in Table 1, only 23 are publicly available for download without any restrictions. These datasets focus on the detection of depression, suicide, and anorexia and are in various languages, including Arabic, Bengali, Brazilian Portuguese, Chinese, Hebrew, Hindi, Spanish, Russian, Roman Urdu, and Thai. For 15 of the datasets, access can be obtained by contacting the authors of the respective research papers, while four datasets require users to complete a data agreement to gain access. Additionally, four datasets are unavailable due to the sensitive nature of the data. For the remaining datasets, the research papers do not provide any information on data availability. Details about the availability of data collections can be found in Appendix A, Table 2.

7 Mental Disorders Detection Approaches

In this section, we present the NLP methods proposed for the datasets in Section 6. Most approaches are monolingual and specifically target only one non-English language.

Classical approaches Most approaches use Bag-of-Words, TF-IDF, or Word2Vec for text representation, which are then used as input for classical machine learning models (Almouzini et al., 2019; Alghamdi et al., 2020; Helmy et al., 2024) or deep learning models (Mann et al., 2020; Tasnim et al., 2022; Ghosh et al., 2023).

Pre-trained transformer-based models While multilingual models like XLM-Roberta and Multilingual BERT demonstrate strong performance in downstream tasks, only two studies focus exclusively on these models (Kabir et al., 2022; Hoque and Salma, 2023). In contrast, twelve of the papers in Section 6 rely on pre-trained monolingual models specific to the target language, such as Chinese BERT (Yao, 2024), AraBERT (Abdulsalam et al., 2024), German BERT (Zanwar et al., 2023), Bangla BERT (Chowdhury et al., 2024) and others. In addition, seven research papers evaluate both language-adapted and multilingual models (Hacohen-Kerner et al., 2022; Oliveira et al., 2024).

Translation Zahran et al. (2025) presented a comprehensive evaluation of LLMs on Arabic data related to depression, suicidal ideation, anxiety, and others. The authors found that LLMs performed better on original Arabic datasets compared to data that had been translated into English. Other works also rely on the detection using data translated from the target language to English (Vajrobol et al., 2023). However, Schoene et al. (2025) has shown that automatically translating suicide dictionaries from English to low-resource languages often leads to spelling errors and fails to capture the cultural nuances of the speakers of the target language. When developing mental health models in other languages, some studies rely on translation from English to the target language, such as Greek (Skianis et al., 2024) or various Indian languages (Rajderkar and Bhat, 2024).

Multilingual approaches Methods developed for multiple languages simultaneously utilize cross-lingual embeddings and make use of information from languages with more mental health-related resources, such as English, to make predictions on Spanish data (Coello-Guilarte et al., 2019). Lee et al. (2020) developed a cross-lingual model for suicidal ideation by translating data from Korean to English and Chinese. They used existing dictionaries related to suicidal ideation in these languages to inform predictions on the Korean language.

8 Cross-cultural and Cross-language Differences in Mental Health Expression

Culture influences the sources of distress, how it is expressed, how it is interpreted, the process of seeking help, and the responses of others (Kirmayer et al., 2001). In addition, the way people perceive themselves influences their mental health. In Western cultures, there is a strong emphasis on personal narratives, and people tend to express their emotions more openly, a trend that is reflected in online posts (Tokunaga, 2009). In contrast, in Asian societies, individuals often internalize their emotional struggles or express them indirectly, influenced by their collectivist values (Broczek et al., 2024). Although negative self-thoughts are a common characteristic of depression, in East Asian contexts, self-criticism is often viewed as a sign of healthy functioning (Gotlib and Hammen, 2008).

Symptoms of mental disorders Cultural differences in the interpretation of mental health symptoms can lead individuals of certain backgrounds to minimize the psychological effects of mental distress. Instead, they may report more socially acceptable somatic symptoms (Kirmayer et al., 2001). Somatic symptoms are common across various cultures, but the ways in which they are reported or understood can differ. In addition, there are culturally specific idioms of distress associated with mental disorders. One such example is the term "nervios" (translated as "nerves" in English), which is a syndrome of distress primarily studied in Latin American communities. This syndrome manifests with psychological and somatic symptoms and has a high comorbidity with anxiety and mood disorders (De Snyder et al., 2000). The DSM-V (American Psychiatric Association, 2013), which is used for the assessment of mental disorders, includes cultural concepts of distress to help clinicians recognize how individuals from various cultures express psychological issues.

Mental health expressions in online language Online expression varies between cultures and has been extensively studied among English-speaking individuals from different regions (De Choudhury et al., 2017; Aguirre and Dredze, 2021; Rai et al., 2024). When analyzing data from a peer-support mental health community, Loveys et al. (2018) found that manifestations of negative emotions differ between demographic groups. Moreover, Pendse et al. (2019) found that users in the US, UK, and Canada employed more clinical language to express mental distress compared to users from India, Malaysia, and the Philippines.

Variation of features across cultures The tendency for self-focused attention, often referred to as "I"-language, is considered one of the strongest predictors of depression in language (Mihalcea et al., 2024). As a result, the frequency of the pronoun "I" has been used in previous studies as a feature for detecting depression in English. However, it is crucial to carefully consider the applicability of this marker to non-English languages. This association has not been observed in non-Western individuals (Rai et al., 2024) or in speakers of Chinese (Lyu et al., 2023) or Romanian (Trifu et al., 2024). While the pronoun "I" serves as a significant indicator of depression in English, its usage in other languages requires special attention due to linguistic differences. For example, English requires nouns or pronouns to be explicitly included as subjects in sentences. In contrast, some languages, such as Chinese and Romanian, are pro-drop languages, which allow the subject of the action to be omitted (Koeneman and Zeijlstra, 2019). This can result in a lower frequency of the personal pronoun "I" in these languages.

Mental health metaphors Indicators of mental disorders are often displayed through metaphors. Depression is often described as weight, pressure, or darkness, and is often portrayed using containment metaphors (Charteris-Black, 2012). Metaphors are often used by individuals to articulate their experience and psychologists in the therapeutic process (Mould et al., 2010). Mental illness metaphors have been extensively studied in English (Charteris-Black, 2012; Lazard et al., 2016) and have been used to predict mental states (Shi et al., 2021; Zhang et al., 2021). With the exception of research in Spanish (Coll-Florit and Climent, 2023), there is a notable lack of resources to understand metaphors of mental illness in other languages.

It is essential to consider the various cultural and multilingual differences when developing automated methods to predict mental disorders based on language. These differences may explain why many studies have shown that models designed to predict mental illnesses often fail to generalize (Aguirre et al., 2021; Abdelkadir et al., 2024).

9 Research Gaps

In this section, we highlight several research gaps that we hope will be explored in future studies.

Lack of mental health-related data for low-resource languages As presented in Section 6, most data collection in non-English languages are often from mid- and high-resourced languages, with the exception of Cantonese, Norwegian, and Sinhala. Currently, many languages remain underrepresented, including high-resourced languages like French and mid-to-high resource languages such as Finnish, Croatian, and Vietnamese. Moreover, there is a lack of data collections for low-resource languages, which may hinder the development of online screening tools for individuals who speak these languages. Although few studies have used automatic translation for building datasets in languages other than English, it cannot accurately capture the cultural nuances of native speakers of the target language (Schoene et al., 2025).

Cross-lingual expressions in underrepresented mental disorders Although there are mental health-related datasets available in non-English data, most of them primarily focus on depression and suicide. Other mental disorders, such as anxiety, OCD, bipolar disorder, and PTSD, are underrepresented. To gain a better understanding of how these disorders manifest in the online language, the research community needs more linguistically diverse collections that encompass a wider range of mental disorders. This approach would not only facilitate a broader exploration of mental health expressions in various languages, but also help develop more inclusive and effective online mental health screening tools worldwide.

Multilingual approaches As highlighted in Section 7, most NLP approaches have focused on processing data in a single target language, with multilingual approaches addressing multiple languages being almost nonexistent. Most existing NLP models developed for mental disorders detection do not support multiple languages effectively, which limits their applicability in multicultural and multilingual settings where mental health issues may manifest differently.

Annotation transparency and consistency Although most of the datasets presented in this paper rely on manual annotation for labeling the data related to mental disorders, it is often unclear who did the annotations. The authors of the research papers should provide specific details about the annotation process, such as whether the annotators are mental health experts or non-experts, if they are native speakers of the target language, and whether they understand the cultural differences in the manifestations of mental disorders. These factors significantly impact the quality and reliability of the data, as understanding cultural nuances is essential in interpreting mental health expressions.

Explainability While many mental health studies in English emphasize the importance of explainable approaches (Yang et al., 2023a; Souto et al., 2023; Yang et al., 2023b), there is a significant opportunity for applying explainable approaches to non-English languages. Currently, few studies have examined model explainability in Bengali (Ghosh et al., 2023) and Thai (Vajrobol et al., 2023). These methods may help in understanding the various manifestations of mental disorders.

10 Conclusion

In this paper, we presented a comprehensive review of research for mental disorders detection from multilingual data sourced from social media. We highlight cross-cultural and multilingual differences in mental health expressions and provide a comprehensive list of data collections that can be used to develop multilingual NLP models for online mental health screening. Our focus was on non-English resources, as most previous research has focused on English (Skaik and Inkpen, 2020; Harrigian et al., 2021). Lastly, we identified several gaps in current research that we hope will be addressed in future interdisciplinary studies.

Future Directions and Call to Action We aim to encourage researchers to develop mental health datasets in low-resource languages, fostering interdisciplinary collaborations with experts from psychology and mental health organizations, as seen in successful previous projects like REMO COST Action[11], and PsyMine (Ellendorff et al., 2016), which have primarily focused on English. By involving community members, multilingual shared tasks can be organized to identify mental disorders across different languages, inspired by successful SemEval multilingual tasks for offensive language (Zampieri et al., 2020) and emotion detection (Muhammad et al., 2025). Researchers can work together to annotate data in underrepresented languages while adhering to ethical protocols. By participating in these tasks, members of the ACL community can gain access to data collections that are essential for developing multilingual models. Such initiatives will improve the visibility of multilingual mental disorder detection and encourage further collaborations, providing researchers with more opportunities to address challenges in this field. Researchers can focus on building data collections for underrepresented mental disorders beyond depression and suicide, adhering to ethical guidelines and providing transparency in the annotation process (Benton et al., 2017). Recent advances in explainability can also be applied to better understand the cultural manifestations of mental disorders.

[11] https://projects.tib.eu/remo

Limitations

Our paper aims to provide a comprehensive review of cross-cultural language differences and the datasets available for developing multilingual NLP models. We included 108 data collections in this study and carefully reviewed each paper cited in our survey. However, it is possible that we may have overlooked some works that do not explicitly mention in their title or abstract that they focus on non-English languages.

Ethical Considerations

Data Collection We recognize that using online data to identify mental disorders is a promising approach for early screening, but it also presents several ethical challenges (Benton et al., 2017; Chancellor and De Choudhury, 2020). To ensure that research protocols in this area comply with ethical guidelines, researchers must take the following steps: (1) obtain Institutional Review Board (IRB) approval, (2) follow ethical research protocols to protect sensitive data, as outlined by Benton et al. (2017), (3) obtain consent from participants, (4) de-anonymize the data and store it on a secure server. Any further sharing of the data with other researchers must adhere to the same ethical protocols. From our survey of 108 datasets, we found that only 18 received ethical approval from an IRB. In addition, 19 papers indicated that they anonymized the data to protect user privacy. It is concerning that only about 35% of the papers adhered to ethical practices in their research, highlighting the urgent need for a greater emphasis on ethical standards, especially since ethical disclosures were expected to gradually increase over time (Ajmani et al., 2023).

Potential Consequences Moreover, the ethical implications extend beyond data collection and storage. Researchers should consider the potential consequences of their findings on the populations studied and ensure that their work does not inadvertently stigmatize or harm individuals with mental health disorders. Incorrect predictions can have harmful effects on individuals' lives. For instance, if a system falsely predicts that someone shows signs of mental disorders, it can adversely impact their well-being due to the stigma associated with such labels. This may lead individuals to believe there is something wrong with them, ultimately lowering their self-esteem (Chancellor et al., 2019b). A false negative prediction occurs when the system fails to identify significant signs of distress, preventing the individual from receiving the necessary treatment or interventions. False negative predictions are particularly critical in cases of suicidal ideation, where a person's life may be at risk. Chancellor et al. (2019a) critically discuss how subjects are represented in this area of research, highlighting the risk of inadvertently dehumanizing individuals. The language used in mental health-related papers can unintentionally perpetuate stigma, often referring to those involved in data collection as "sufferers" of mental disorders while labeling others as "normal." Engaging with the community and stakeholders during the research process can help mitigate these risks and foster a more responsible approach to using online data in mental health research (Chancellor et al., 2019b).

Model Validity There are ongoing concerns regarding the construct validity of models trained on data collected from social media, specifically whether these models effectively measure the manifestations of mental disorders (Chancellor and De Choudhury, 2020). The datasets used in this survey predominantly rely on manual annotation or labeling through validated questionnaires, which are considered more reliable methods for annotation. However, it is essential to conduct interdisciplinary research and ground the constructs being measured in both theoretical and clinical frameworks. For example, clinical depression (or major depressive disorder) is fundamentally different from merely "feeling depressed." The latter may refer to temporary feelings, while clinical depression encompasses a range of persistent symptoms. These symptoms may include depressed mood, loss of interest in previously enjoyed activities, changes in body weight, sleep disturbances, fatigue, psychomotor agitation or retardation, feelings of guilt, and thoughts of death or suicidal ideation (American Psychiatric Association, 2013). To be diagnosed with depression, these symptoms must be persistent and significantly impair an individual's ability to function. Prioritizing interdisciplinary collaboration and rigorous validation methods is essential in addressing the complexities of mental health.

Representativeness It is important to note that individuals active on social media represent only a subset of the overall population. As a result, there may be differences in how mental disorders are expressed among social media users compared to the general population. Using social media data can introduce bias, as it tends to reflect the experiences of younger and more technologically literate individuals who are more likely to engage with these platforms (Chancellor et al., 2019b). In addition, datasets that include self-disclosure of a mental health diagnosis often come from individuals who are more likely to have sought professional help for their diagnosis and/or treatment. Furthermore, not everyone feels comfortable sharing sensitive information about their mental health online (Chancellor et al., 2019b).

Cultural and Linguistic Variation Understanding cultural and linguistic variations is crucial when developing automated methods for predicting mental disorders, as they help explain why many predictive models struggle to generalize effectively on data from different demographics (Aguirre et al., 2021; Aguirre and Dredze, 2021; Abdelkadir et al., 2024). Furthermore, each individual's experience with depression is unique, and it is important to consider their distinct experiences and symptomatology. Algorithmic representations and abstractions play a crucial role in the understanding of mental illness and well-being by providing a framework for generalization (Chancellor et al., 2019a). While these simplifications can help identify trends and better understand complex individual experiences, they also risk oversimplifying those experiences. It is important to recognize that generalizing can sometimes lead to misunderstandings regarding the unique nuances of mental health experiences and symptoms. Each person's experience with mental health disorders is unique, and acknowledging this is essential for a deeper understanding of mental health.

References

Nuredin Ali Abdelkadir, Charles Zhang, Ned Mayo, and Stevie Chancellor. 2024. Diverse perspectives, divergent models: Cross-cultural evaluation of depression detection on twitter. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 672–680.

Asma Abdulsalam, Areej Alhothali, and Saleh Al-Ghamdi. 2024. Detecting suicidality in arabic tweets using machine learning and deep learning techniques. *Arabian Journal for Science and Engineering*, pages 1–14.

Kaustubh Agarwal and Bhavya Dhingra. 2021. Deep learning based approach for detecting suicidal ideation in hindi-english code-mixed text: Baseline and corpus. In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*, pages 100–105.

Carlos Aguirre and Mark Dredze. 2021. Qualitative analysis of depression models by demographics. In *Proceedings of CLPsych Workshop, NAACL*, pages 169–180.

Carlos Aguirre, Keith Harrigian, and Mark Dredze. 2021. Gender and racial fairness in depression research using social media. In *Proceedings of EACL*, pages 2932–2949.

Ahmed A Ahad, Marcos Sanchez-Gonzalez, and Patricia Junquera. 2023. Understanding and addressing mental health stigma across cultures for improving psychiatric care: a narrative review. *Cureus*, 15(5).

Leah Hope Ajmani, Stevie Chancellor, Bijal Mehta, Casey Fiesler, Michael Zimmer, and Munmun De Choudhury. 2023. A systematic review of ethics disclosures in predictive mental health research. In *Proceedings of ACM FAccT*, pages 1311–1323.

Malak Fahad Al-Haider, Ali Mustafa Qamar, Hasan Shojaa Alkahtani, and Hafiz Farooq Ahmad. 2024. Classification of obsessive-compulsive disorder symptoms in arabic tweets using machine learning and word embedding techniques. *Journal of Advances in Information Technology*, 15(7).

Norah Al-Musallam and Mohammed Al-Abdullatif. 2022. Depression detection through identifying depressive arabic tweets from saudi arabia: machine learning approach. In *2022 Fifth National Conference of Saudi Computers Colleges (NCCC)*, pages 11–18. IEEE.

Eatedal Alabdulkreem. 2021. Prediction of depressed arab women using their tweets. *Journal of Decision Systems*, 30(2-3):102–117.

Norah Saleh Alghamdi, Hanan A Hosni Mahmoud, Ajith Abraham, Samar Awadh Alanazi, and Laura García-Hernández. 2020. Predicting depression symptoms in an arabic psychological forum. *IEEE access*, 8:57317–57334.

Abdulqader M Almars. 2022. Attention-based bi-lstm model for arabic depression classification. *Computers, Materials & Continua*, 71(2).

Salma Almouzini, Asem Alageel, et al. 2019. Detecting arabic depressed users from twitter data. *Procedia Computer Science*, 163:257–265.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*, 5th ed. edition. Autor, Washington, DC.

Ghelmar Astoveza, Randolph Jay P Obias, Roi Jed L Palcon, Ramon L Rodriguez, Bernie S Fabito, and Manolito V Octaviano. 2018. Suicidal behavior detection on twitter using neural network. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 0657–0662. IEEE.

Thushari Atapattu, Mahen Herath, Charitha Elvitigala, Piyanjali de Zoysa, Kasun Gunawardana, Menasha Thilakaratne, Kasun de Zoysa, and Katrina Falkner. 2022. Emoment: An emotion annotated mental health corpus from two south asian countries. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6991–7001.

Nadiah A Baghdadi, Amer Malki, Hossam Magdy Balaha, Yousry AbdulAzeem, Mahmoud Badawy, and Mostafa Elhosseini. 2022. An optimized deep learning approach for suicide detection through arabic tweets. *PeerJ Computer Science*, 8:e1070.

Christina Baskal, Amelie Elisabeth Beutel, Jessika Keberlein, Malte Ollmann, Esra Üresin, Jana Vischinski, Janina Weihe, Linda Achilles, and Christa Womser-Hacker. 2022. Data sets of eating disorders by categorizing reddit and tumblr posts: A multilingual comparative study based on empirical findings of texts and images. In *Proceedings of the Workshop on Dataset Creation for Lower-Resourced Languages within the 13th Language Resources and Evaluation Conference*, pages 10–18.

Pantaporn Benjachairat, Twittie Senivongse, Nattasuda Taephant, Jiratchaya Puvapaisankit, Chonlakorn Maturosjamnan, and Thanakorn Kultananawat. 2024. Classification of suicidal ideation severity from twitter messages using machine learning. *International Journal of Information Management Data Insights*, 4(2):100280.

Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the EthNLP Workshop*, pages 94–102.

Natalie Berry, Fiona Lobban, Maksim Belousov, Richard Emsley, Goran Nenadic, Sandra Bucci, et al. 2017. # whywetweetmh: understanding why people use twitter to discuss mental health problems. *JMIR*, 19(4):e6173.

Raymond R Bond, Maurice D Mulvenna, Courtney Potts, Siobhan O'Neill, Edel Ennis, and John Torous. 2023. Digital transformation of mental health services. *Npj Mental Health Research*, 2(1):13.

Panchanit Boonyarat, Di Jie Liew, and Yung-Chun Chang. 2024. Leveraging enhanced bert models for detecting suicidal ideation in thai social media content amidst covid-19. *Information Processing & Management*, 61(4):103706.

Katarzyna Milana Broczek, Marie-Christine Gely-Nargeot, and Pietro Gareri. 2024. Editorial: Depression across cultures and linguistic identities. *Frontiers in Psychology*, 15.

Ana-Maria Bucur, Andreea-Codrina Moldovan, Krutika Parvatikar, Marcos Zampieri, Ashiqur R KhudaBukhsh, and Liviu P Dinu. 2025. On the state of nlp approaches to modeling depression in social media: A post-covid-19 outlook. *Journal of Biomedical and Health Informatics*.

Yicheng Cai, Haizhou Wang, Huali Ye, Yanwen Jin, and Wei Gao. 2023. Depression detection on online social network with multivariate time series feature of user depressive symptoms. *Expert Systems with Applications*, 217:119538.

Rafael A Calvo, David N Milne, M Sazzad Hussain, and Helen Christensen. 2017. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, 23(5):649–685.

Lei Cao, Huijun Zhang, Ling Feng, Zihan Wei, Xin Wang, Ningyun Li, and Xiaohao He. 2019. Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1718–1728.

Junyeop Cha, Seoyun Kim, and Eunil Park. 2022. A lexicon-based approach to examine depression detection in social media: the case of twitter and university community. *Humanities and Social Sciences Communications*, 9(1):1–10.

Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. 2019a. Who is the" human" in human-centered machine learning: The case of predicting mental health from social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–32.

Stevie Chancellor, Michael L Birnbaum, Eric D Caine, Vincent MB Silenzio, and Munmun De Choudhury. 2019b. A taxonomy of ethical tensions in inferring mental health states from social media. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 79–88.

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):1–11.

Jonathan Charteris-Black. 2012. Shattering the bell jar: Metaphor, gender, and depression. *Metaphor and Symbol*, 27(3):199–216.

Qijin Cheng, Tim MH Li, Chi-Leung Kwok, Tingshao Zhu, and Paul SF Yip. 2017. Assessing suicide risk and emotional distress in chinese social media: a text mining and machine learning study. *Journal of medical internet research*, 19(7):e243.

Jenny Chim, Adam Tsakalidis, Dimitris Gkoumas, Dana Atzil-Slonim, Yaakov Ophir, Ayah Zirikly, Philip Resnik, and Maria Liakata. 2024. Overview of the clpsych 2024 shared task: Leveraging large language models to identify evidence of suicidality risk in online posts. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)*, pages 177–190.

Chun Yueh Chiu, Hsien Yuan Lane, Jia Ling Koh, and Arbee LP Chen. 2021. Multimodal depression detection on instagram considering time interval of posts. *Journal of Intelligent Information Systems*, 56(1):25–47.

Ahmadul Karim Chowdhury, Saidur Rahman Sujon, Md Shirajus Salekin Shafi, Tasin Ahmmad, Sifat Ahmed, Khan Md Hasib, and Faisal Muhammad Shah. 2024. Harnessing large language models over transformer models for detecting bengali depressive social media text: A comprehensive study. *Natural Language Processing Journal*, 7:100075.

Laritza Coello-Guilarte, Rosa María Ortega-Mendoza, Luis Villaseñor-Pineda, and Manuel Montes-y Gómez. 2019. Crosslingual depression detection in twitter using bilingual word alignments. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings 10*, pages 49–61. Springer.

Caroline Cohrdes, Seren Yenikent, Jiawen Wu, Bilal Ghanem, Marc Franco-Salvador, Felicitas Vogelgesang, et al. 2021. Indications of depressive symptoms during the covid-19 pandemic in germany: comparison of national survey and twitter data. *JMIR mental health*, 8(6):e27140.

Marta Coll-Florit and Salvador Climent. 2023. Metaphor repositories: the case of the mental health metaphor dictionary. *Digital Scholarship in the Humanities*, 38(4):1440–1452.

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.

S Zafra Cremades, Jose M Gomez Soriano, and Borja Navarro-Colorado. 2017. Design, compilation and annotation of a corpus for the detection of suicide messages in social networks. *Procesamiento del Lenguaje Natural*, 59:65–72.

Vinícios Faustino de Carvalho, Bianca Giacon, Carlos Nascimento, and Bruno Magalhães Nogueira. 2020. Machine learning for suicidal ideation identification on twitter for the portuguese language. In *Brazilian Conference on Intelligent Systems*, pages 536–550. Springer.

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.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of ICWSM*.

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 ACM CSCW*, pages 353–369.

Adonias C de Oliveira, Evandro JS Diniz, Silmar Teixeira, and Ariel S Teles. 2022. How can machine learning identify suicidal ideation from user's texts? towards the explanation of the boamente system. *Procedia Computer Science*, 206:141–150.

V Nelly Salgado De Snyder, Ma de Jesus Diaz-Perez, and Victoria D Ojeda. 2000. The prevalence of nervios and associated symptomatology among inhabitants of mexican rural communities. *Culture, Medicine and Psychiatry*, 24:453–470.

Bart Desmet and Véronique Hoste. 2014. Recognising suicidal messages in dutch social media. In *9th international conference on language resources and evaluation (LREC)*, pages 830–835.

Bart Desmet and Véronique Hoste. 2018. Online suicide prevention through optimised text classification. *Information Sciences*, 439:61–78.

Sahraoui Dhelim, Liming Chen, Sajal K Das, Huansheng Ning, Chris Nugent, Gerard Leavey, Dirk Pesch, Eleanor Bantry-White, and Devin Burns. 2023. Detecting mental distresses using social behavior analysis in the context of covid-19: A survey. *ACM Computing Surveys*.

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.

Mohammad El-Ramly, Hager Abu-Elyazid, Youseef Mo'men, Gameel Alshaer, Nardine Adib, Kareem Alaa Eldeen, and Mariam El-Shazly. 2021. Cairodep: Detecting depression in arabic posts using bert transformers. In *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pages 207–212. IEEE.

Tilia Ellendorff, Simon Foster, and Fabio Rinaldi. 2016. The psymine corpus-a corpus annotated with psychiatric disorders and their etiological factors. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3723–3729.

Jiahui Gao, Qijin Cheng, and Philip LH Yu. 2019. Detecting comments showing risk for suicide in youtube. In *Proceedings of the Future Technologies Conference (FTC) 2018: Volume 1*, pages 385–400. Springer.

Muskan Garg. 2023. Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, 30(3):1819–1842.

Muskan Garg. 2024. Towards mental health analysis in social media for low-resourced languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(3):1–22.

Tapotosh Ghosh, Md Hasan Al Banna, Md Jaber Al Nahian, Mohammed Nasir Uddin, M Shamim Kaiser, and Mufti Mahmud. 2023. An attention-based hybrid architecture with explainability for depressive social media text detection in bangla. *Expert Systems with Applications*, 213:119007.

Ian H Gotlib and Constance L Hammen. 2008. *Handbook of depression*. Guilford Press.

Zhihua Guo, Nengneng Ding, Minyu Zhai, Zhenwen Zhang, and Zepeng Li. 2023. Leveraging domain knowledge to improve depression detection on chinese social media. *IEEE Transactions on Computational Social Systems*, 10(4):1528–1536.

Yaakov Hacohen-Kerner, Natan Manor, Michael Goldmeier, and Eytan Bachar. 2022. Detection of anorexic girls-in blog posts written in hebrew using a combined heuristic ai and nlp method. *IEEE Access*, 10:34800–34814.

Mika Hämäläinen, Pattama Patpong, Khalid Alnajjar, Niko Partanen, and Jack Rueter. 2021. Detecting depression in thai blog posts: a dataset and a baseline. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 20–25.

Keith Harrigian, Carlos Aguirre, and Mark Dredze. 2021. On the state of social media data for mental health research. In *Proceedings of CLPsych Workshop, NAACL*, pages 15–24.

Mariam Hassib, Nancy Hossam, Jolie Sameh, and Marwan Torki. 2022. Aradepsu: Detecting depression and suicidal ideation in arabic tweets using transformers. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 302–311.

AbdelMoniem Helmy, Radwa Nassar, and Nagy Ramdan. 2024. Depression detection for twitter users using sentiment analysis in english and arabic tweets. *Artificial intelligence in medicine*, 147:102716.

Siranuch Hemtanon, Saifon Aekwarangkoon, and Nichnan Kittphattanabawon. 2020. Behavior features for automatic detection of depression from facebook users. In *Machine Learning and Artificial Intelligence*, pages 12–20. IOS Press.

Siranuch Hemtanon and Nichnan Kittiphattanabawon. 2019. An automatic screening for major depressive disorder from social media in thailand. In *Proceeding National & International Conference*, volume 10, pages 103–113.

Sandamini Herath and Thareendra Keerthi Wijayasiriwardhane. 2024. A social media intelligence approach to predict suicidal ideation from sinhala facebook posts. In *2024 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, volume 7, pages 1–6. IEEE.

Misato Hiraga. 2017. Predicting depression for japanese blog text. In *Proceedings of ACL 2017, student research workshop*, pages 107–113.

Md Nesarul Hoque and Umme Salma. 2023. Detecting level of depression from social media posts for the low-resource bengali language. *Journal of Engineering Advancements*, 4(02):49–56.

Xiaolei Huang, Xin Li, Lei Zhang, Tianli Liu, David Chiu, and Tingshao Zhu. 2015. Topic model for identifying suicidal ideation in chinese microblog. In *Proceedings of the 29th pacific asia conference on language, information and computation*, pages 553–562. Waseda University.

Yan Huang, Xiaoqian Liu, and Tingshao Zhu. 2019. Suicidal ideation detection via social media analytics. In *Human Centered Computing: 5th International Conference, HCC 2019, Čačak, Serbia, August 5–7, 2019, Revised Selected Papers 5*, pages 166–174. Springer.

Mika Hämäläinen, Pattama Patpong, Khalid Alnajjar, Niko Partanen, and Jack Rueter. 2021. Detecting depression in Thai blog posts: a dataset and a baseline. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 20–25, Online. ACL.

Nikolay Ignatiev, Ivan V Smirnov, and Maxim Stankevich. 2022. Predicting depression with text, image, and profile data from social media. In *ICPRAM*, pages 753–760.

Sabiha Islam, Md Shafiul Alam Forhad, and Hasan Murad. 2022. Banglasapm: A deep learning model for suicidal attempt prediction using social media content in bangla. In *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pages 1122–1126. IEEE.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293.

Woojin Jung, Donghun Kim, Seojin Nam, and Yongjun Zhu. 2023. Suicidality detection on social media using metadata and text feature extraction and machine learning. *Archives of suicide research*, 27(1):13–28.

Mohsinul Kabir, Tasnim Ahmed, Md. Bakhtiar Hasan, Md Tahmid Rahman Laskar, Tarun Kumar Joarder, Hasan Mahmud, and Kamrul Hasan. 2023. Deptweet: A typology for social media texts to detect depression severities. *Computers in Human Behavior*, 139:107503.

Muhammad Khubayeeb Kabir, Maisha Islam, Anika Nahian Binte Kabir, Adiba Haque, and Md Khalilur Rhaman. 2022. Detection of depression severity using bengali social media posts on mental health: study using natural language processing techniques. *JMIR Formative Research*, 6(9):e36118.

Kantinee Katchapakirin, Konlakorn Wongpatikaseree, Panida Yomaboot, and Yongyos Kaewpitakkun. 2018. Facebook social media for depression detection in the thai community. In *2018 15th international joint conference on computer science and software engineering (jcsse)*, pages 1–6. IEEE.

S Kayalvizhi, Durairaj Thenmozhi, Bharathi Raja Chakravarthi, SV Kogilavani, Pratik Anil Rahood, et al. 2023. Overview of the shared task on detecting signs of depression from social media text. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 25–30.

Donghun Kim, Woojin Jung, Seojin Nam, Hongjin Jeon, Jihyun Baek, and Yongjun Zhu. 2022a. Understanding information behavior of south korean twitter users who express suicidality on twitter. *Digital health*, 8:20552076221086339.

Nam Hyeok Kim, Ji Min Kim, Da Mi Park, Su Ryeon Ji, and Jong Woo Kim. 2022b. Analysis of depression in social media texts through the patient health questionnaire-9 and natural language processing. *Digital health*, 8:20552076221114204.

Laurence J Kirmayer et al. 2001. Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of clinical psychiatry*, 62:22–30.

Olaf Koeneman and Hedde Zeijlstra. 2019. Morphology and pro drop. In *Oxford Research Encyclopedia of Linguistics*.

Boriharn Kumnunt and Ohm Sornil. 2020. Detection of depression in thai social media messages using deep learning. In *DeLTA*, pages 111–118.

Allison J Lazard, Benita A Bamgbade, Jennah M Sontag, and Carolyn Brown. 2016. Using visual metaphors in health messages: A strategy to increase effectiveness for mental illness communication. *Journal of health communication*, 21(12):1260–1268.

Daeun Lee, Soyoung Park, Jiwon Kang, Daejin Choi, and Jinyoung Han. 2020. Cross-lingual suicidal-oriented word embedding toward suicide prevention. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2208–2217.

Angela Leis, Francesco Ronzano, Miguel A Mayer, Laura I Furlong, and Ferran Sanz. 2019. Detecting signs of depression in tweets in spanish: behavioral and linguistic analysis. *Journal of medical Internet research*, 21(6):e14199.

Genghao Li, Bing Li, Langlin Huang, Sibing Hou, et al. 2020. Automatic construction of a depression-domain lexicon based on microblogs: text mining study. *JMIR medical informatics*, 8(6):e17650.

Zepeng Li, Zhengyi An, Wenchuan Cheng, Jiawei Zhou, Fang Zheng, and Bin Hu. 2023. Mha: a multimodal hierarchical attention model for depression detection in social media. *Health information science and systems*, 11(1):6.

Tingting Liu, Devansh Jain, Shivani R Rapole, Brenda Curtis, Johannes C. Eichstaedt, Lyle H. Ungar, and Sharath Chandra Guntuku. 2023. Detecting symptoms of depression on reddit. In *Proceedings of ACM Web Science Conference*, WebSci '23, page 174–183, New York, NY, USA. ACM.

D. Losada and F. Crestani. 2016. A test collection for research on depression and language use. In *Proc. of Experimental IR Meets Multilinguality, Multimodality, and Interaction, 7th International Conference of the CLEF Association, CLEF 2016*, pages 28–39, Evora, Portugal.

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, New Orleans, LA. Association for Computational Linguistics.

Sihua Lyu, Xiaopeng Ren, Yihua Du, and Nan Zhao. 2023. Detecting depression of chinese microblog users via text analysis: Combining linguistic inquiry word count (liwc) with culture and suicide related lexicons. *Frontiers in psychiatry*, 14:1121583.

Pilar López-Úbeda, Flor Miriam Plaza Del Arco, Manuel Carlos Díaz Galiano, L Alfonso Urena Lopez, and M Teresa Martín-Valdivia. 2019. Detecting anorexia in spanish tweets. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 655–663.

Ashwag Maghraby and Hosnia Ali. 2022. Modern standard arabic mood changing and depression dataset. *Data in Brief*, 41:107999.

Suwaroj Mahasiriakalayot, Twittie Senivongse, and Nattasuda Taephant. 2022. Predicting signs of depression from twitter messages. In *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pages 1–6. IEEE.

Paulo Mann, Aline Paes, and Elton H Matsushima. 2020. See and read: detecting depression symptoms in higher education students using multimodal social media data. In *Proceedings of the International AAAI Conference on Web and social media*, volume 14, pages 440–451.

John J McGrath, Ali Al-Hamzawi, Jordi Alonso, Yasmin Altwaijri, Laura H Andrade, Evelyn J Bromet, Ronny Bruffaerts, José Miguel Caldas de Almeida, Stephanie Chardoul, Wai Tat Chiu, et al. 2023. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. *The Lancet Psychiatry*, 10(9):668–681.

Augusto R Mendes and Helena Caseli. 2024. Identifying fine-grained depression signs in social media posts. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8594–8604.

Rada Mihalcea, Laura Biester, Ryan L Boyd, Zhijing Jin, Veronica Perez-Rosas, Steven Wilson, and James W Pennebaker. 2024. How developments in natural language processing help us in understanding human behaviour. *Nature Human Behaviour*, 8(10):1877–1889.

David N Milne, Glen Pink, Ben Hachey, and Rafael A Calvo. 2016. Clpsych 2016 shared task: Triaging content in online peer-support forums. In *Proceedings of the third workshop on computational linguistics and clinical psychology*, pages 118–127.

Ruba Mohmand, Usman Habib, Muhammad Usman, Jamel Baili, and Yunyoung Nam. 2024. A deep learning approach for automated depression assessment using roman urdu. *IEEE Access*.

Tracy J Mould, Lindsay G Oades, and Trevor P Crowe. 2010. The use of metaphor for understanding and managing psychotic experiences: A systematic review. *Journal of Mental Health*, 19(3):282–293.

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Seid Muhie Yimam, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine De Kock, Tadesse Destaw Belay, Ibrahim Said Ahmad, Nirmal Surange, Daniela Teodorescu, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino Ali, Vladimir Araujo, Abinew Ali Ayele, Oana Ignat, Alexander Panchenko, Yi Zhou, and Saif M. Mohammad. 2025. SemEval-2025 task 11: Bridging the gap in text-based emotion detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, Vienna, Austria. Association for Computational Linguistics.

Dhiaa A Musleh, Taef A Alkhales, Reem A Almakki, Shahad E Alnajim, Shaden K Almarshad, Rana S Alhasaniah, Sumayh S Aljameel, and Abdullah A Almuqhim. 2022. Twitter arabic sentiment analysis to detect depression using machine learning. *Computers, Materials & Continua*, 71(2).

Alba María Mármol-Romero, Adrián Moreno-Muñoz, Flor Miriam Plaza-del Arco, María Dolores Molina-González, Maria Teresa Martín-Valdivia, Luis Alfonso Ureña-López, and Arturo Montejo-Raéz. 2023. Overview of mentalriskes at iberlef 2023: Early detection of mental disorders risk in spanish. *Procesamiento del Lenguaje Natural*, 71:329–350.

Sergazy Narynov, Daniyar Mukhtarkhanuly, and Batyrkhan Omarov. 2020. Dataset of depressive posts in russian language collected from social media. *Data in brief*, 29:105195.

Usman Naseem, Adam G. Dunn, Jinman Kim, and Matloob Khushi. 2022. Early identification of depression severity levels on reddit using ordinal classification. WWW '22, page 2563–2572, New York, NY, USA. ACM.

Yutaka Miyaji Yuka Niimi. 2021. Machine learning approach for depression detection in japanese. In *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation*, pages 346–353.

Alicia L Nobles, Jeffrey J Glenn, Kamran Kowsari, Bethany A Teachman, and Laura E Barnes. 2018. Identification of imminent suicide risk among young adults using text messages. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–11.

Adonias Caetano de Oliveira, Renato Freitas Bessa, and Ariel Soares Teles. 2024. Comparative analysis of bert-based and generative large language models for detecting suicidal ideation: a performance evaluation study. *Cadernos de Saúde Pública*, 40:e00028824.

Irwan Oyong, Ema Utami, and Emha Taufiq Luthfi. 2018. Natural language processing and lexical approach for depression symptoms screening of indonesian twitter user. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pages 359–364. IEEE.

Javier Parapar, Patricia Martin-Rodilla, David E Losada, and Fabio Crestani. 2021. Overview of erisk 2021: Early risk prediction on the internet.

Javier Parapar, Patricia Martín-Rodilla, David E. Losada, and Fabio Crestani. 2024. Overview of erisk 2024: Early risk prediction on the internet. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 73–92, Cham. Springer Nature Switzerland.

Sungjoon Park, Kiwoong Park, Jaimeen Ahn, and Alice Oh. 2020. Suicidal risk detection for military personnel. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2523–2531.

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.

Zhichao Peng, Qinghua Hu, and Jianwu Dang. 2019. Multi-kernel svm based depression recognition using social media data. *International Journal of Machine Learning and Cybernetics*, 10:43–57.

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.

Viraj Rajderkar and Aruna Bhat. 2024. Multilingual depression detection in online social media across eight indian languages. In *2024 3rd International Conference for Innovation in Technology (INOCON)*, pages 1–6. IEEE.

Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Joaquim Puntí, Pilar Medina-Bravo, Diego Alejandro Velazquez, Josep Maria Gonfaus, and Jordi Gonzàlez. 2020. Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. *Journal of medical internet research*, 22(7):e17758.

Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Nadia Sanz Lamora, Aida Álvarez, Alexandre González-Rodríguez, Meritxell Lozano Rochel, Roger Llobet Vives, Diego Alejandro Velazquez, Josep Maria Gonfaus, et al. 2021. Characterization of anorexia nervosa on social media: Textual, visual, relational, behavioral, and demographical analysis. *Journal of medical Internet research*, 23(7):e25925.

Lashini Rathnayake and Isuri Anuradha Nanomi Arachchige. 2021. Supervised learning approach for detection of sinhala depressive posts based on twitter. In *2021 21st International Conference on Advances in ICT for Emerging Regions (ICter)*, pages 111–116. IEEE.

Filza Rehmani, Qaisar Shaheen, Muhammad Anwar, Muhammad Faheem, and Shahzad Sarwar Bhatti. 2024. Depression detection with machine learning of structural and non-structural dual languages. *Healthcare Technology Letters*.

Alba M Mármol Romero, Adrián Moreno Muñoz, Flor Miriam Plaza Del Arco, M Dolores Molina-González, María Teresa Martín Valdivia, L Alfonso Urena Lopez, and Arturo Montejo Ráez. 2024. Mentalriskes: A new corpus for early detection of mental disorders in spanish. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11204–11214.

Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8):1121–1133.

Esteban A Ríssola, David E Losada, and Fabio Crestani. 2021. A survey of computational methods for online mental state assessment on social media. *ACM Transactions on Computing for Healthcare*, 2(2):1–31.

Kayalvizhi Sampath and Thenmozhi Durairaj. 2022. Data set creation and empirical analysis for detecting signs of depression from social media postings. In *Computational Intelligence in Data Science*, pages 136–151, Cham. Springer International Publishing.

Wesley Santos, Amanda Funabashi, and Ivandré Paraboni. 2020. Searching brazilian twitter for signs of mental health issues. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6111–6117.

Wesley Ramos dos Santos, Rafael Lage de Oliveira, and Ivandré Paraboni. 2024. Setembrobr: a social media corpus for depression and anxiety disorder prediction. *Language Resources and Evaluation*, 58(1):273–300.

Annika Marie Schoene, John E Ortega, Rodolfo Joel Zevallos, and Laura Haaber Ihle. 2025. Lexicography saves lives (lsl): Automatically translating suicide-related language. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3179–3192.

Ivan Sekulić, Matej Gjurković, and Jan Šnajder. 2018. Not just depressed: Bipolar disorder prediction on reddit. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 72–78.

Tiancheng Shen, Jia Jia, Guangyao Shen, Fuli Feng, Xiangnan He, Huanbo Luan, Jie Tang, Thanassis Tiropanis, Tat Seng Chua, and Wendy Hall. 2018. Cross-domain depression detection via harvesting social media. Proceedings of IJCAI.

Nan Shi, Dongyu Zhang, Lulu Li, and Shengjun Xu. 2021. Predicting mental health problems with automatic identification of metaphors. *Journal of Healthcare Engineering*, 2021(1):5582714.

Ruba Skaik and Diana Inkpen. 2020. Using social media for mental health surveillance: a review. *ACM Computing Surveys*, 53(6):1–31.

Konstantinos Skianis, A Doğruöz, and John Pavlopoulos. 2024. Leveraging llms for translating and classifying mental health data. In *Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024)*, pages 236–241.

Eliseo Bao Souto, Anxo Pérez, and Javier Parapar. 2023. Explainability, interpretability, depression detection, social media. *arXiv preprint arXiv:2310.13664*.

Vivian Stamou, George Mikros, George Markopoulos, and Spyridoula Varlokosta. 2024. Establishing control corpora for depression detection in modern greek: Methodological insights. In *Proceedings of the Fifth Workshop on Resources and Process-Ing of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments@ LREC-COLING 2024*, pages 68–76.

Maxim Stankevich, Andrey Latyshev, Evgenia Kuminskaya, Ivan Smirnov, and Oleg Grigoriev. 2019. Depression detection from social media texts. In

Elizarov, A., Novikov, B., Stupnikov., S (eds.) Data Analytics and Management in Data Intensive Domains: XXI International Conference DAMDID/RCDL, page 352.

Maxim Stankevich, Ivan Smirnov, Natalia Kiselnikova, and Anastasia Ushakova. 2020. *Depression detection from social media profiles*, pages 181–194.

Lijing Sun, Yu Luo, et al. 2022. Identification and analysis of depression and suicidal tendency of sina weibo users based on machine learning. *Advances in Educational Technology and Psychology*, 6(9):108–117.

Tom Tabak and Matthew Purver. 2020. Temporal mental health dynamics on social media. In *Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.

Dai Tang, Tina Chou, Naomi Drucker, Adi Robertson, William C Smith, and Jeffery T Hancock. 2011. A tale of two languages: strategic self-disclosure via language selection on facebook. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 387–390.

Farzana Tasnim, Sultana Umme Habiba, Nuren Nafisa, and Afsana Ahmed. 2022. Depressive bangla text detection from social media post using different data mining techniques. In *Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021*, pages 237–247. Springer.

Robert S Tokunaga. 2009. High-speed internet access to the other: The influence of cultural orientations on self-disclosures in offline and online relationships. *Journal of Intercultural Communication Research*, 38(3):133–147.

Raluca Nicoleta Trifu, Bogdan Nemeș, Dana Cristina Herta, Carolina Bodea-Hategan, Dorina Anca Talaș, and Horia Coman. 2024. Linguistic markers for major depressive disorder: a cross-sectional study using an automated procedure. *Frontiers in Psychology*, 15:1355734.

Adam Tsakalidis, Jenny Chim, Iman Munire Bilal, Ayah Zirikly, Dana Atzil-Slonim, Federico Nanni, Philip Resnik, Manas Gaur, Kaushik Roy, Becky Inkster, et al. 2022a. Overview of the clpsych 2022 shared task: Capturing moments of change in longitudinal user posts. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 184–198.

Adam Tsakalidis, Federico Nanni, Anthony Hills, Jenny Chim, Jiayu Song, and Maria Liakata. 2022b. Identifying moments of change from longitudinal user text. In *Proc. of ACL*, pages 4647–4660.

Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. Recognizing depression from twitter activity. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 3187–3196.

Faye Beatriz Tumaliuan, Lorelie Grepo, and Eugene Rex Jalao. 2024. Development of depression data sets and a language model for depression detection: mixed methods study. *JMIR Data*, 5:e53365.

Abdul Hasib Uddin, Durjoy Bapery, and Abu Shamim Mohammad Arif. 2019. Depression analysis of bangla social media data using gated recurrent neural network. In *2019 1st International conference on advances in science, engineering and robotics technology (ICASERT)*, pages 1–6. IEEE.

Md Zia Uddin. 2022. Depression detection in text using long short-term memory-based neural structured learning. In *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, pages 408–414. IEEE.

Md Zia Uddin, Kim Kristoffer Dysthe, Asbjørn Følstad, and Petter Bae Brandtzaeg. 2022. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, 34(1):721–744.

Vajratiya Vajrobol, Nitisha Aggarwal, Unmesh Shukla, Geetika Jain Saxena, Sanjeev Singh, and Amit Pundir. 2023. Explainable cross-lingual depression identification based on multi-head attention networks in thai context. *International Journal of Information Technology*, pages 1–16.

Kid Valeriano, Alexia Condori-Larico, and José Sulla-Torres. 2020. Detection of suicidal intent in spanish language social networks using machine learning. *International Journal of Advanced Computer Science and Applications*, 11(4).

Vasudha Varadarajan, Allison Lahnala, Ganesan Adithya V, Gourab Dey, Siddharth Mangalik, Ana-Maria Bucur, Nikita Soni, Rajath Rao, Kevin Lanning, Isabella Vallejo, Lucie Flek, H. Andrew Schwartz, Charles Welch, and Ryan L Boyd. 2024. Archetypes and entropy: Theory-driven extraction of evidence for suicide risk. In *Proceedings of CLPsych Workshop, EACL*.

Debasish Bhattacharjee Victor, Jamil Kawsher, Md Shad Labib, and Subhenur Latif. 2020. Machine learning techniques for depression analysis on social media-case study on bengali community. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pages 1118–1126. IEEE.

Miryam Elizabeth Villa-Pérez, Luis A Trejo, Maisha Binte Moin, and Eleni Stroulia. 2023. Extracting mental health indicators from english and spanish social media: A machine learning approach. *IEEE Access*, 11:128135–128152.

Otto von Sperling and Marcelo Ladeira. 2019. Mining twitter data for signs of depression in brazil. In *Anais do VII Symposium on Knowledge Discovery, Mining and Learning*, pages 25–32. SBC.

Lidong Wang, Yin Zhang, Bin Zhou, Shihua Cao, Keyong Hu, and Yunfei Tan. 2024. Automatic depression prediction via cross-modal attention-based multi-modal fusion in social networks. *Computers and Electrical Engineering*, 118:109413.

Siqin Wang, Huan Ning, Xiao Huang, Yunyu Xiao, Mengxi Zhang, Ellie Fan Yang, Yukio Sadahiro, Yan Liu, Zhenlong Li, Tao Hu, et al. 2023. Public surveillance of social media for suicide using advanced deep learning models in japan: time series study from 2012 to 2022. *Journal of medical internet research*, 25:e47225.

Xiaofeng Wang, Shuai Chen, Tao Li, Wanting Li, Yejie Zhou, Jie Zheng, Yaoyun Zhang, and Buzhou Tang. 2019. Assessing depression risk in chinese microblogs: a corpus and machine learning methods. In *2019 IEEE International conference on healthcare informatics (ICHI)*, pages 1–5. IEEE.

Yiding Wang, Zhenyi Wang, Chenghao Li, Yilin Zhang, and Haizhou Wang. 2020. A multimodal feature fusion-based method for individual depression detection on sina weibo. In *2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC)*, pages 1–8. IEEE.

Agnieszka Wołk, Karol Chlasta, and Paweł Holas. 2021. Hybrid approach to detecting symptoms of depression in social media entries.

Konlakorn Wongaptikaseree, Panida Yomaboot, Kantinee Katchapakirin, and Yongyos Kaewpitakkun. 2020. Social behavior analysis and thai mental health questionnaire (tmhq) optimization for depression detection system. *IEICE TRANSACTIONS on Information and Systems*, 103(4):771–778.

En-Liang Wu, Chia-Yi Wu, Ming-Been Lee, Kuo-Chung Chu, and Ming-Shih Huang. 2023. Development of internet suicide message identification and the monitoring-tracking-rescuing model in taiwan. *Journal of affective disorders*, 320:37–41.

Min Yen Wu, Chih-Ya Shen, En Tzu Wang, and Arbee L. P. Chen. 2018. A deep architecture for depression detection using posting, behavior, and living environment data. *Journal of Intelligent Information Systems*, 54:225–244.

Shweta Yadav, Jainish Chauhan, Joy Prakash Sain, Krishnaprasad Thirunarayan, Amit P. Sheth, and Jeremiah Schumm. 2020. Identifying depressive symptoms from tweets: Figurative language enabled multitask learning framework. *CoRR*, abs/2011.06149.

Kailai Yang, Shaoxiong Ji, Tianlin Zhang, Qianqian Xie, and Sophia Ananiadou. 2023a. Towards interpretable mental health analysis with chatgpt. *arXiv preprint arXiv:2304.03347*.

Kailai Yang, Tianlin Zhang, Ziyan Kuang, Qianqian Xie, and Sophia Ananiadou. 2023b. Mentalllama: Interpretable mental health analysis on social media with large language models. *arXiv preprint arXiv:2309.13567*.

Tingting Yang, Fei Li, Donghong Ji, Xiaohui Liang, Tian Xie, Shuwan Tian, Bobo Li, and Peitong Liang. 2021. Fine-grained depression analysis based on chinese micro-blog reviews. *Information Processing & Management*, 58(6):102681.

Xiaoxu Yao, Guang Yu, Xianyun Tian, and Jingyun Tang. 2020. Patterns and longitudinal changes in negative emotions of people with depression on sina weibo. *Telemedicine and e-Health*, 26(6):734–743.

Zheng Yao. 2024. A multi-model approach to detection of depression in the chinese social media entries. In *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, pages 2148–2151. IEEE.

Elroi Yoshua and Warih Maharani. 2024. Depression detection of users in social-media twitter using decision tree with word2vec. *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, 9(1):95–100.

Yang Yu, Qi Li, and Xiaoqian Liu. 2023. Automatic anxiety recognition method based on microblog text analysis. *Frontiers in Public Health*, 11:1080013.

Noureldin Zahran, Aya E Fouda, Radwa J Hanafy, and Mohammed E Fouda. 2025. A comprehensive evaluation of large language models on mental illnesses in arabic context. *arXiv preprint arXiv:2501.06859*.

Marcos Zampieri, Preslav Nakov, Sara Rosenthal, Pepa Atanasova, Georgi Karadzhov, Hamdy Mubarak, Leon Derczynski, Zeses Pitenis, and Çağrı Çöltekin. 2020. SemEval-2020 Task 12: Multilingual Offensive Language Identification in Social Media (OffensEval 2020). In *Proceedings of SemEval*.

Sourabh Zanwar, Daniel Wiechmann, Yu Qiao, and Elma Kerz. 2023. Smhd-ger: a large-scale benchmark dataset for automatic mental health detection from social media in german. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1526–1541.

Daniel Zarate, Michelle Ball, Maria Prokofieva, Vassilis Kostakos, and Vasileios Stavropoulos. 2023. Identifying self-disclosed anxiety on twitter: A natural language processing approach. *Psychiatry Research*, 330:115579.

Dongyu Zhang, Nan Shi, Ciyuan Peng, Abdul Aziz, Wenhong Zhao, and Feng Xia. 2021. Mam: a metaphor-based approach for mental illness detection. In *International Conference on Computational Science*, pages 570–583. Springer.

Lei Zhang, Xiaolei Huang, Tianli Liu, Ang Li, Zhenxiang Chen, and Tingshao Zhu. 2014. Using linguistic features to estimate suicide probability of chinese microblog users. In *International Conference on Human Centered Computing*, pages 549–559.

Tianlin Zhang, Annika M Schoene, Shaoxiong Ji, and Sophia Ananiadou. 2022. Natural language processing applied to mental illness detection: a narrative review. *NPJ digital medicine*, 5(1):46.

Zhenwen Zhang, Jianghong Zhu, Zhihua Guo, Yu Zhang, Zepeng Li, and Bin Hu. 2024. Natural language processing for depression prediction on sina weibo: Method study and analysis. *JMIR Mental Health*, 11:e58259.

Jianghong Zhu, Zhenwen Zhang, Zhihua Guo, and Zepeng Li. 2024. Sentiment classification of anxiety-related texts in social media via fuzing linguistic and semantic features. *IEEE Transactions on Computational Social Systems*.

A Appendix

A.1 Methodology details

Initially, 405 papers were retrieved through a database search across the ACL Anthology, ACM Digital Library, IEEE Xplore, Springer Nature Link, ScienceDirect, and Google Scholar. After screening and assessing their eligibility, 108 papers were included in this survey. The PRISMA flow diagram is presented in Figure 3.

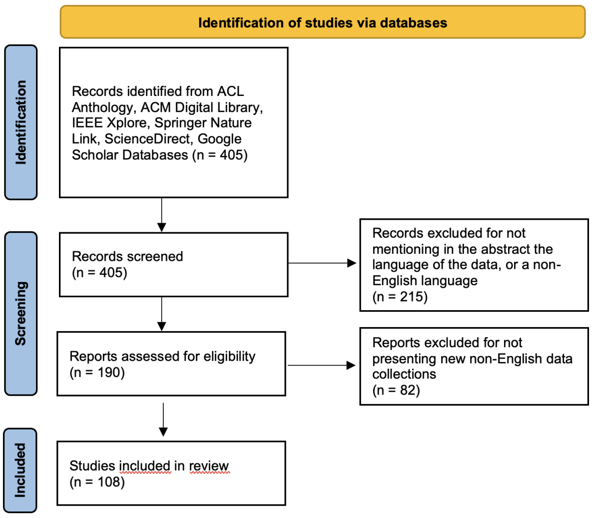

Figure 3: PRISMA flow diagram for our review.

A.2 Rankings of the publication venues for the multilingual datasets

Figure 4 presents an overview of these languages along with the ranking of the publications in which they appeared. The rankings for conferences are categorized as A*, A, B, and C, following the CORE Rankings Portal.[12] For journals, the rankings are classified as Q1, Q2, Q3, and Q4, based on the Journal Citation Reports.[13] There are also datasets published in unranked conferences or journals. While about half of the datasets appeared in unranked venues, leading to lower visibility for the research, the other half were published in high-ranking journals and conferences.

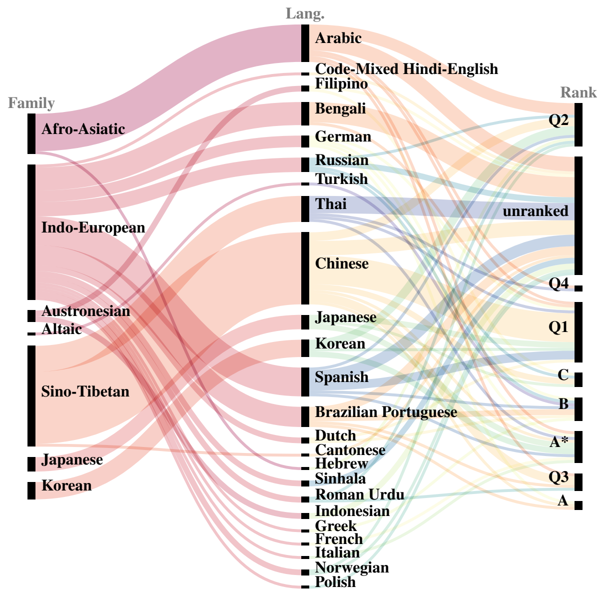

Figure 4: Overview of the languages in the datasets, their language families, and the ranking of their publication venues.

[12] https://www.core.edu.au/conference-portal
[13] https://jcr.clarivate.com/

Table 2: List of Non-English available datasets for mental disorders-related tasks using data posted on online platforms.

Dataset	Language	Mental disorder	Platform	Annotation Procedure	Label	Dataset Size	Availab.	Method	Performance
Almouzini et al. (2019)	Arabic	depression	Twitter	Self-disclosure	Binary	89 users, 2.7K posts	UNK	Bag-of-Unigrams, Linear SVM	Accuracy: 87.5%, F1-score: 87.5%
Alghamdi et al. (2020)	Arabic	depression	Online forums	Manual annotation	Binary	20K posts	UNK	Lexicon-based	Accuracy: 80.45%, F1-score: 80.81%
Alabdulkreem (2021)	Arabic	depression	Twitter	Manual annotation	Binary	200 users	UNK	Word2Vec, RNN-LSTM	Accuracy: 72%, F1-score: 69%
Musleh et al. (2022)	Arabic	depression	Twitter	CES-D and self-disclosure	Binary, DSM-5 symptoms	4.5K posts	UNK	TF-IDF, RF	Accuracy: 82.39%, F1-score: 82.53%
CairoDep (El-Ramly et al., 2021)	Arabic	depression	Twitter, Reddit, Online forums	Keywords, Manual annotation	Binary	2.4K posts	FREE	AraBERT	Accuracy: 96.93%, F1-score: 96.92%
Almars (2022)	Arabic	depression	Twitter	Manual annotation	Binary	6.1K posts	UNK	Attention BiLSTM	Accuracy: 83%, F1-score: 83%
Maghraby and Ali (2022)	Arabic	depression	Twitter	PHQ-9	PHQ-9 symptoms	1.2K posts	FREE	TF-IDF, RF	F1-score: 98%
AraDepSu (Hassib et al., 2022)	Arabic	depression, suicide	Twitter	Manual annotation	Depression, depression with suicidal ideation, or non-depression	20K posts	UNK	MARBERT	Accuracy: 91.20%, F1-score: 88.75%
Arabic Dep 10,000 (Helmy et al., 2024)	Arabic	depression	Twitter	Manual annotation	Binary	10K posts	FREE	TF-IDF, RBF SVM	F1-score: 96.6%
Al-Haider et al. (2024)	Arabic	OCD	Twitter	Manual annotation	Binary	8.7K posts	UNK	fastText, RF	F1-score: 80%
Baghdadi et al. (2022)	Arabic	suicide	Twitter	Manual annotation	Binary	2K posts	FREE	AraBERT	Accuracy: 96.06%, F1-score: 95.86%
Abdulsalam et al. (2024)	Arabic	suicide	Twitter	Manual annotation	Binary	5.7K posts	UNK	AraBERT	Accuracy: 91%, F1-score: 88%
Al-Musallam and Al-Abdullatif (2022)	Arabic	depression	Twitter	Manual annotation	Binary	6k posts	UNK	TF-IDF, LR	Accuracy: 82%, F1-score: 81%
Uddin et al. (2019)	Bengali	depression	Twitter	Manual annotation	Binary	1.1K posts	FREE	GRU	Accuracy: 75.7%
Victor et al. (2020)	Bengali	depression	Facebook, Twitter	Manual annotation	Binary	30K posts	UNK	TF-IDF, RF	Accuracy: 90%
Kabir et al. (2022)	Bengali	depression	Facebook	Manual annotation	Depression severity	5K posts	FREE	BiGRU	Accuracy: 81%, F1-score: 81%
Tasnim et al. (2022)	Bengali	depression	Facebook	Manual annotation	Binary	7K posts	UNK	BOW, TF-IDF, DT	Accuracy: 97%, F1-score: 97%
BanglaSPD Islam et al. (2022)	Bengali	suicide	Facebook	Manual annotation	Binary	1.7K posts	UNK	fastText, CNN-BiLSTM	Accuracy: 61%, F1-score: 61%
Ghosh et al. (2023)	Bengali	depression	Facebook, Twitter, YouTube	Manual annotation	Binary	15K posts	AUTH	fastText, BiLSTM-CNN	Accuracy: 94.32%
Hoque and Salma (2023)	Bengali	depression	Facebook	Manual annotation	Depression severity	2.5K posts	UNK	XLM-RoBERTa	Accuracy: 61.11%, F1-score: 60.89%
BSMDD (Chowdhury et al., 2024)	Bengali	depression	Reddit, Twitter	Manual annotation	Binary	28K posts	FREE	GPT 3.5	Accuracy: 97.96%, F1-score: 98.04%
von Sperling and Ladeira (2019)	Brazilian Portuguese	depression	Twitter	Self-disclosure	Binary	2.9K users	UNK	Hand-crafted features, SVM	F1-score: 79.8%

Dataset	Language	Mental disorder	Platform	Annotation Procedure	Label	Dataset Size	Availab.	Method	Performance
Mann et al. (2020)	Brazilian Portuguese	depression	Instagram	BDI	Binary	221 users	UNK	ELMo, ResNet, MLP	F1-score: 79%
Santos et al. (2020)	Brazilian Portuguese	depression	Twitter	Self-disclosure	Binary	224 users	UNK	TF-IDF, LR	F1-score: 69%
de Carvalho et al. (2020)	Brazilian Portuguese	suicide	Twitter	Manual annotation	Possibly/ Strongly concerning, Safe to ignore	2.4K posts	UNK	BERT-Portuguese	F1-score: 79%
SetembroBR (Santos et al., 2024)	Brazilian Portuguese	depression	Twitter	Self-disclosure	Binary	18.8K users	FREE	BERTimbau	F1-score: 63%
Mendes and Caseli (2024)	Brazilian Portuguese	depression symptoms	Facebook	Manual annotation	Depression symptoms	780 posts	UNK	BERTimbau	Precision: 76.14%
Oliveira et al. (2024)	Brazilian Portuguese	suicide	Twitter	Manual annotation	Binary	3.7K posts	FREE	BERTimbau	Accuracy: 96%
Gao et al. (2019)	Cantonese	suicide	Youtube	Manual annotation	Binary	5K posts	UNK	Word2vec, LSTM	Geometric mean of accuracies: 84.5%
Zhang et al. (2014)	Chinese	suicide	Sina Weibo	SPS	SPS score	697 users	UNK	LIWC, LR	RMSE: 11
Huang et al. (2015)	Chinese	suicide	Sina Weibo	Manual annotation	Binary	7.3K posts	UNK	Topic modeling, LibSVM	80.0%
Cheng et al. (2017)	Chinese	suicide	Sina Weibo	Suicide Probability Scale (SPS), DASS-21	Binary	974 users	UNK	LIWC, SVM	AUC: 0.61%
Shen et al. (2018)	Chinese	depression	Sina Weibo	Self-disclosure	Binary	1.1K users	UNK	Hand-crafted features, DNN	F1-score: 78.5%
Wu et al. (2018)	Chinese	depression	Facebook	CES-D	Binary	1.4K users	UNK	Word2vec, Hand-crafted features ,RNN	F1-score: 76.9%
Cao et al. (2019)	Chinese	suicide	Sina Weibo	Manual checking of self-report and/or appartenence to a suicide-related community	Binary	7K users	DUA	fastText, RNN	Accuracy: 91%
Wang et al. (2019)	Chinese	depression	Sina Weibo	Manual annotation	Depression severity	13.9K users	UNK	BERT	F1-score: 53.8%
Peng et al. (2019)	Chinese	depression	Sina Weibo	Manual annotation	Binary	387 users	UNK	TF-IDF, SVM	83.46%
Huang et al. (2019)	Chinese	suicide	Sina Weibo	Manual annotation	Binary	18.5K posts	UNK	LIWC, Dictionary, LR, DT, SVM	F1-score: 0.88%
Li et al. (2020)	Chinese	depression	Sina Weibo	Self-disclosure	Binary	1.8K users	FREE	Lexicon-based, RF	F1-score: 76%
WU3D (Wang et al., 2020)	Chinese	depression	Sina Weibo	Depression-related keywords	Binary	32K users	FREE XLNet mebeddings, BiGRU		F1-score: 96.85%
Yao et al. (2020)	Chinese	depression	Sina Weibo	Manual, automatic annotation	Binary	2.7K users	UNK	-	-
Yang et al. (2021)	Chinese	depression	Sina Weibo	Manual annotation	Depression severity	6.1K posts	AUTH	BERT-based	F1-score: 65.7%
Chiu et al. (2021)	Chinese, English	depression	Instagram	Depression-related keywords	Binary	520 users	UNK	Multimodal features, Adaboost	F1-score: 83.5%
Sun et al. (2022)	Chinese	suicide, depression	Sina Weibo	BDI, SDS, Manual annotation	Binary / Possibly/Strongly concerning, Safe to ignore	203 users, 1.2K posts	UNK	Gradient Boosting	Accuracy: 82.4%
Cai et al. (2023)	Chinese	depression	Sina Weibo	Self-disclosure and manual annotation	Binary	23K users	FREE	DNN	F1-score: 92.02%
Li et al. (2023)	Chinese	depression	Sina Weibo	Self-disclosure, manual annotation	Binary	4.8K users	UNK	Multimodal features, DNN	F1-score: 92.78%
Guo et al. (2023)	Chinese	depression	Sina Weibo	Manual annotation	Binary	3.1K users	UNK	Lexicon-based, XGBoost	F1-score: 93.22%
Wu et al. (2023)	Chinese	suicide	Dcard and PTT	Manual annotation	Risk levels	2K posts	UNK	-	-

Dataset	Language	Mental disorder	Platform	Annotation Procedure	Label	Dataset Size	Availab.	Method	Performance
Lyu et al. (2023)	Chinese	depression	Sina Weibo	CES-D	Binary	789 users	AUTH	LIWC, LR	Pearson correlation: 0.33
Yu et al. (2023)	Chinese	anxiety	Sina Weibo	Self-Rating Anxiety Scale	SAS score	1K users	N/A	LIWC, XGBoost	Pearson correlation: 0.32
Zhu et al. (2024)	Chinese	anxiety	Sina Weibo	Manual annotation	Binary	6K posts	UNK	LIWC, Word embeddings CNN	F1-score: 86.13%
Wang et al. (2024)	Chinese	depression	Sina Weibo	Manual annotation	Binary	14.8K users	AUTH	Multimodal features, DNN	F1-score: 89.15%
Yao (2024)	Chinese	depression	Sina Weibo	Manual annotation	Binary	200 users	AUTH	BERT, DNN	Accuracy: 90%
Zhang et al. (2024)	Chinese	depression	Sina Weibo	Manual annotation	Binary	1.6K users	UNK	Tencent Embeddings, HTN	F1-score: 95.43%
Desmet and Hoste (2014)	Dutch	suicide	Online forums	Manual annotation	Fine-grained labels	1.3K posts	UNK	BOW, SVM	F1-score: 85.6%
Desmet and Hoste (2018)	Dutch	suicide	Online forums	Manual annotation	Fine-grained labels	10K posts	UNK	BOW, Topic modeling, LibSVM	F1-score: 92.69%
Abdelkadir et al. (2024) Ali et al. (2024)	English, but from different populations	depression	Twitter	Self-disclosure, Manual annotation	Binary	531 users	UNK	MentalLongformer	F1-score: 62%
Tumaliuan et al. (2024)	Filipino, English	depression	Twitter	PHQ-9	Binary	72 users	AUTH	-	-
Astoveza et al. (2018)	Filipino, Taglish	suicide	Twitter	Manual annotation	Binary	2.1K posts	UNK	BOW ,MLP	Accuracy: 77.9%
Cohrdes et al. (2021)	German	depression	Twitter	Automatic annotation for PHQ-8 symptoms	Binary	88K posts	AUTH	–	–
SMHD-GER (Zanwar et al., 2023)	German	depression, ADHD, anxiety, bipolar, OCD, PTSD, schizophrenia	Reddit	Manual annotation	Labels for multiple disorders	28K posts	DUA	LIWC, BiLSTM	F1-score: 52.22%
Baskal et al. (2022)	German, Russian, Turkish, English	eating disorders	Reddit, Tumblr	Manual annotation	Binary	3K posts	AUTH	–	–
Tabak and Purver (2020)	German, French, Italian, Spanish, English	depression	Twitter	Self-disclosure	Binary	5K users	UNK	BOW, BiLSTM	F1-score: 69%
Hacohen-Kerner et al. (2022)	Hebrew	anorexia	Online forums	Manual annotation	Binary	200 posts	FREE	Hand-crafted features, RF	Accuracy: 90.63%
Agarwal and Dhingra (2021)	Code-Mixed Hindi-English	suicide	Reddit	Subreddit membership	Binary	6.4K posts	FREE	Indic BERT	Accuracy: 98.54%
Oyong et al. (2018)	Indonesian	depression	Twitter	Manual annotation	Binary	55 users	UNK	Hand-crafted depression score	F1-score: 0.50%
Yoshua and Maharani (2024)	Indonesian	depression	Twitter	DASS-42	Binary	184 users	UNK	Word2Vec, DT	F1-score: 94%
Tsugawa et al. (2015)	Japanese	depression	Twitter	CES-D, BDI	Binary	209 users	UNK	Hand-crafted features, Topic modeling, SVM	Accuracy: 66%
Hiraga (2017)	Japanese	depression	Online blogs	Self-disclosure	Binary	101 users	UNK	Part-of-speech, NB	Accuracy: 95.5%
Niimi (2021)	Japanese	depression	TOBYO	Blog theme	Binary	901 users	UNK	TF-IDF, SVM	F1-score: 96.2%
Wang et al. (2023)	Japanese	suicide	Twitter	Manual annotation	Binary	30K posts	N/A	–	–

Dataset	Language	Mental disorder	Platform	Annotation Procedure	Label	Dataset Size	Availab.	Method	Performance
Lee et al. (2020)	Korean	suicide	Naver Cafe	Membership in a forum	Binary	31K posts	UNK	Word2Vec, RNN	Accuracy: 87.49%
Park et al. (2020)	Korean	suicide	Online forums	Manual annotation	Risk levels	2.7K posts	AUTH	XLM-R	Accuracy: 88%
Kim et al. (2022a)	Korean	suicide	Twitter	Manual annotation	Binary	20K posts, 414 users	UNK	–	–
Kim et al. (2022b)	Korean	depression	Online forums	PHQ-9, Manual annotation	PHQ-9 score, PHQ-9 symptoms	60 users, 28K posts	UNK	BERT-based	Accuracy: 68.3%
Jung et al. (2023)	Korean	suicide	Twitter	Manual annotation	Binary	20k posts	UNK	Metadata, word count, XGBoost	F1-score: 83.57%
Cha et al. (2022)	Korean, Japanese, English	depression	Twitter, Everytime	Lexicon-based automatic annotation	Binary	26M posts, 22K posts	AUTH	BERT-based	F1-score:99%
Stamou et al. (2024)	Modern Greek	depression	Twitter	Self-disclosure	Binary	78 users	AUTH	–	–
Uddin (2022)	Norwegian	depression	Online forums	Manual annotation	Binary	21.8K posts	UNK	TF-IDF, LSTM	Accuracy: 99%
Uddin et al. (2022)	Norwegian	depression	Online forums	Manual annotation	Binary	30K posts	UNK	Hand-crafted depression features; LSTM	Accuracy: 99%
Wołk et al. (2021)	Polish	depression	Facebook, Reddit	Self-disclosure, clinical interview	Binary	262 users	UNK	Hybrid Model; BERT	Accuracy 71%
Rehmani et al. (2024)	Roman Urdu	depression	Facebook	Manual annotation	Depression severity	3K posts	AUTH	SVM	Accuracy: 84%
Mohmand et al. (2024)	Roman Urdu	depression	Twitter	Keywords-based annotations + Expert review	Depression severity	25K posts	FREE	Transfer learning; BERT	Accuracy: 99%
Stankevich et al. (2019)	Russian	depression	VKontakte	BDI	BDI score	531 users	UNK	Psycholinguistic Markers	F1 Score: 66%
Narynov et al. (2020)	Russian	depression	VKontakte	Manual annotation	Binary	34K posts	FREE	–	–
Stankevich et al. (2020)	Russian	depression	VKontakte	BDI	BDI score	1.3K users	UNK	–	–
Ignatiev et al. (2022)	Russian	depression	VKontakte	BDI	Binary	619 users	DUA	CatBoost	F1 Score: 69%
Rathnayake and Arachchige (2021)	Sinhala	depression	Twitter, Facebook	Manual annotation	Binary	1K posts	UNK	KNN	Accuracy: 70%
EmoMent (Atapattu et al., 2022)	Sinhala, English	mental illness	Facebook	Manual annotation	mental illness, sadness, suicidal, anxiety/stress, psychosomatic, other, irrelevant	2.8K posts	AUTH	RoBERTa	F1 Score: 76%
Herath and Wijayasiriwardhane (2024)	Sinhala	suicide	Facebook	Manual annotation	Binary	300 posts	UNK	Naive Bayes	Accuracy: 79%
Leis et al. (2019)	Spanish	depression	Twitter	Self-disclosure, manual annotation	Binary	540 users, 1K posts	FREE	–	–
SAD López-Úbeda et al. (2019)	Spanish	anorexia	Twitter	Hashtags	Binary	5.7K posts	FREE	SVM	Accuracy: 91.6%
Valeriano et al. (2020)	Spanish	suicide	Twitter	Manual annotation	Binary	2K posts	FREE	Word2Vec; LR	Accuracy: 79%
Ramírez-Cifuentes et al. (2020)	Spanish	suicide	Twitter	Manual annotation	Binary	252 users	N/A	–	–
Ramírez-Cifuentes et al. (2021)	Spanish	anorexia	Twitter	Manual annotation	Anorexia, control, under treatment, recovered, doubtful	645 users	N/A	–	–

Dataset	Language	Mental disorder	Platform	Annotation Procedure	Label	Dataset Size	Availab.	Method	Performance
Villa-Pérez et al. (2023)	Spanish, English	depression, ADHD, anxiety, ASD, bipolar, eating disorders, OCD, PTSD, schizophrenia	Twitter	Self-disclosure	Labels for multiple disorders	6K users	DUA	N-Grams; XGBoost	AUC: 71.2%
MentalRiskES Romero et al. (2024)	Spanish	depression, anxiety, suicide, eating disorders	Telegram	Manual annotation	Binary + suffer + in favour (sf), suffer + against (sa), suffer + other (so) for Depression	1.2K users	AUTH	Social media text; mDeBERTa	F1 Score: 46%
Cremades et al. (2017)	Spanish, English	suicide	Facebook, Twitter, Blogspot, Reddit, Pinterest	Manual annotation	Binary	97 posts	FREE	–	–
Coello-Guilarte et al. (2019)	Spanish, English	depression	Twitter	Self-disclosure	Binary	316 users	FREE	BA-LIWC	F1 Score: 65%
Katchapakirin et al. (2018)	Thai	depression	Facebook	TMHQ	Binary	35 users	UNK	RF	F1 Score: 88.9%
Hemtanon and Kittiphattanabawon (2019)	Thai	depression	Facebook	Manual annotation	Binary	1.5K posts	UNK	SVM	F1 Score: 94%
Kumnunt and Sornil (2020)	Thai	depression	Pantip	Hashtags	Binary	31K posts	UNK	CNN-LSTM	F1 Score: 83.1%
Hemtanon et al. (2020)	Thai	depression	Facebook	PHQ-9	Binary	160 users	UNK	Social media features	F1 Score: 91.4%
Wongaptikaseree et al. (2020)	Thai	depression	Facebook	TMHQ	Binary	600 users	UNK	–	–
Hämäläinen et al. (2021)	Thai	depression	Online blogs	Manual annotation	Binary	900 posts	FREE	BERT	Accuracy: 77.53%
Mahasiriakalayot et al. (2022)	Thai	depression	Twitter	Manual annotation	Depression symptoms	3.1K posts	UNK	LSTM	Accuracy: 89.17%
Boonyarat et al. (2024)	Thai	suicide	Twitter	Manual annotation	Binary + 6 emotions	2.4K posts	FREE	Linguistic features; BERT	F1-score: 90%
Benjachairat et al. (2024)	Thai	suicide	Twitter	Manual annotation	C-SSRS Labels	5.1K posts	AUTH	Text features; LSTM	F1-score: 93.88%