A Comparative Study of Affective and Linguistic Traits in Online Depression and Suicidal Discussion Forums

Salim Sazzed
Department of Computer Science
Old Dominion University
Norfolk, VA, USA
salim.sazzed@gmail.com

ABSTRACT

Depression is a type of mental illness that negatively impacts the lives of millions of people worldwide. Extreme depression is related to increasingly hopeless and worthless feelings, which may lead to suicidal attempts. The widespread use of social media, coupled with the anonymity it provides, enables individuals to freely express and share their frustrations and low emotions on these platforms. As a preliminary study, here, we investigate how the user-generated content regarding the two mental-health issues, depression and suicidal tendencies, are related at linguistic levels based on two Reddit mental-health forums. By collecting user posts from two Reddit social media forums, *r/depression* and *r/suicidal watch*, we seek to find the (dis)similarity of the various affective, grammatical, and semantic attributes in these two groups. We find that while some of the affective features exhibit some differences, overall, most attributes yield similar patterns in these two groups. The results suggest that it is very challenging to separate depressive posts from suicidal posts at the linguistic level as they possess similar traits. Hence, it is imperative to monitor the content of the depression forum vigilantly (likewise the suicidal forum) to identify any suicidal tendencies.

CCS CONCEPTS

• **Information systems → World Wide Web**; • **Computing methodologies → Natural language processing**.

KEYWORDS

Mental health, depression, suicidal thought, social health text, linguistic analysis

ACM Reference Format:
Salim Sazzed. 2023. A Comparative Study of Affective and Linguistic Traits in Online Depression and Suicidal Discussion Forums. In *34th ACM Conference on Hypertext and Social Media (HT '23), September 4–8, 2023, Rome, Italy.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3603163.3609059

1 INTRODUCTION

Depression is one of the leading causes of disability (e.g., partial, mild, and extreme conditions) [10] that affects over 300 million people worldwide and negatively impacts their family, social and professional life. A depressed person is vulnerable to inflicting self-harm [36] and renders the possibility of being a serious threat to public safety. Unfortunately, the worldwide occurrence of depression is soaring [16], with an increase of 18% between 2005 and 2015. Suicide is one of the leading causes of death, especially for young people in the age group between 15 and 24 years [1]. According to World Health Organization (WHO), over 0.7 million people die from suicide each year [2]; indeed, suicide is the fourth leading cause of death. Many warning signs of possible suicidal feelings could be related to extreme depression. Early detection of mental illness such as depression and determining its severity can help prevent its negative effects by taking appropriate measures [8, 19].

The increasing popularity of social media platforms (e.g., Twitter [28], Reddit, and Facebook [30]) makes them significant data sources for mental health research as people frequently share their thoughts, feelings, and opinions there [20, 29]. User-generated content on social media portrays behavioral aspects related to individuals' moods, communication styles, activities, and social interactions. The emotional content and language employed in social media posts can serve as potential indicators of mental health issues, including symptoms such as feelings of worthlessness, guilt, helplessness, and self-hatred that are commonly associated with major depression. Previous research has shown that individuals with mental illnesses may utilize social media as an outlet for expressing their emotional state and seeking relief from their condition [25], revealing a potential link between the content of social media posts and mental health.

In recent years, several studies have concentrated on automatically identifying depressive or suicidal text in social media using different techniques [4, 26]. For instance, Pirina and Çöltekin [26] applied the support vector machine (SVM) classifier to analyze depression in Reddit posts. Orabi et al. [24] categorized depressive posts on two Twitter benchmark datasets, CLPsych2015 [5] and Bell Letters Talk datasets utilizing convolutional neural network (CNN) and recurrent neural network (RNN). More recently, Zogan et al. (2021) utilized a fusion of two asymmetric parallel networks representing user behavior and user post history to automatically identify depression. Trifan et al. [34] conducted an investigation to comprehend the impact of various psycholinguistic patterns in writings to classify depressed users. The authors combined these psycholinguistic features with a rule-based estimator and assessed their effects on this classification problem. In a related study, Tadesse

HT '24, September 4–8, 2023, Rome, Italy

ACM ISBN 979-8-4007-0232-7/23/09.
https://doi.org/10.1145/3603163.3609059

[1] https://www.nami.org/Your-Journey/Kids-Teens-and-Young-Adults/What-You-Need-to-Know-About-Youth-Suicide
[2] https://www.who.int/news-room/fact-sheets/detail/suicide

Depression
I cannot fucking feel a single fucking thing man . I bottle up every feeling and I am so far away from everyone in my life . I have no friends or anyone I can talk to , I feel like shit right this now , I do not want anything in life , I do not want to grow old , travel , make friends or whatever the fuck everyone says we should live for . I just want to end it , I wish I died in my sleep , everyday I wake up I feel shock and sadness . I did not choose to be born so why do I have to suffer for people who do not give a fuck about me . oh god !! just kill me nowMusic used to make me feel better (it was the only thing) , but now even music does nothing . I cannot tell anyone about how I feel , I feel stuck I cannot take this anymore . My parents think that I am very lazy and stupid so anything I tell them would be because I am m lazy . Like I told my mom n dad about how I do not find anything interest while talking about career and they said it is because I am lazy and stupid that I think too highly of myself and that I am overconfident .They also said that I am useless and would not accomplish anything in life . when actually I have low self-esteem and zero self-confidence . how am I supposed to tell them about my depression that I have got because of neglect and abuse from them in childhood . Like I tell them that I do not feel like doing anything , then they would just say it is because I am lazy , you do not want to do this then you are fucking lazy and stupid "I just want to die in my sleep , that is the only thing I want . I feel so alone , I wish I had someone to support me emotionally The worst thing about all this is being alone
I hate myself, I fucking suck. I am the most unstable fucker alive
Suicidal
there is.....foodAnd other things I will be judged for and for having weird views because maybe I am delusional I want to go to jail. Life would be better there I am not joking. Maybe suicide would not even cross my mind in jail
It sounds quite stupid, I know. For the last couple of days, I have prepared everything I need (I will not bring the details, but it is a very elaborate plan) except one: a note. I started brainstorming and drafting what is useless and what is not, decided whether to post or write by hand, and checked for grammar. But I felt like even my last piece of writing, however short, was not good enough. Nothing I do is ever good enough. that is how it has always been. I want to perfect something in my life and end myself in a grand finale. Yet, I do not know. Perhaps I can give it another few days. Today is a new one. I am planning to end it all, but as a writer, I keep editing my suicide note

Figure 1: Examples of posts from Depression and Suicidal groups

et al. (2019) compiled a list of highly common terms utilized by depressed users. Shen et al. [31] proposed a multimodal learning model to detect depressed users on Twitter.

Although existing studies analyzed social media texts to recognize the presence of depression or suicidal tendencies, they predominantly treated these topics as separate entities (except [19]). Additionally, these studies have largely focused on classification tasks, such as categorizing text into depressive and non-depressive categories [2, 27], or assessing the level of suicidal risk [37]. In contrast, this study investigates how various textual attributes in posts representing these two mental health-related issues (i.e., suicidal tendencies and generic depression) are related. By performing linguistic analysis, this study reveals the (dis)similarities in these two mental health-related issues in a popular social media forum, Reddit. The posts pertaining to suicidal tendencies were collected from the Reddit discussion forum *r/suicidalwatch*, which primarily consists of posts made by individuals who are contemplating suicide [32]. On the other hand, the depression-related posts were extracted from the *r/depression* forum. We conduct a comprehensive analysis of various psychological characteristics, such as sentiment and emotion, to investigate the extent to which they differ across these two types of posts. Additionally, we perform semantic, grammatical, and length-related analyses on these posts. The findings reveal that although some differences exist, there are significant similarities in the linguistic patterns between these two groups, making it challenging to discern text representing suicidal tendencies from generic depression.

2 DATASET

The dataset utilized in this study comprises user posts from two Reddit forums, namely */r/depression* and */r/suicidewatch*. Reddit is a popular and widely used social media platform where individuals participate in discussions on a wide range of topics. Due to the anonymous nature of Reddit, users often share posts related to stigmatized topics [33]. The considerable length of Reddit posts makes them a valuable resource for exploring linguistic and other textual features. The dataset used in this study was obtained from Kaggle [3], a web platform for data scientists and machine learning researchers. The */r/suicidewatch* forum contains 10007 posts, while the */r/depression* forum contains 10357 posts. After excluding posts with various issues, such as those with minimal content (less than five words), the final dataset consists of 9968 posts from the */r/suicidewatch* forum and around 10300 posts from the */r/depression* forum.

Each post in the dataset is annotated with a binary label of either 0 or 1, corresponding to the subreddit forums it represents. A label of 0 indicates that the post belongs to the */r/depression* forum (i.e., depression group), and a label of 1 indicates that it belongs to the */r/suicidewatch* forum (i.e., suicidal group)[32]. It is noteworthy that the guidelines of the */r/depression* forum explicitly state that suicidal thoughts should be directed to the */r/suicidewatch* forum. Additionally, the */r/suicidewatch* forum is widely recognized as a prominent suicide support forum, offering support for individuals struggling with vulnerable thoughts [11]. Thus, based on these observations, it is highly unlikely that posts related to suicidal ideation would appear in the */r/depression* forum, as opposed to the */r/suicidewatch* forum [7]. Fig 1 shows some examples of posts from two groups.

3 FEATURE ANALYSIS AND SIGNIFICANCE TEST

We explore a diverse set of affective, grammatical, and semantic features (chosen based on earlier mental health-related research focusing on non-clinical text [3]) in the posts of two groups. We analyze the distributions of these features in the two types of posts and report their mean, median, and standard distribution (std.) values. In addition, we investigate whether the differences in the quantitative values for each attribute in the two types of posts are significant using the Mann-Whitney U test [21, 35]. The Mann-Whitney U test is often interpreted as a comparison between the

[3]https://www.kaggle.com/datasets/xavrig/reddit-dataset-rdepression-and-rsuicidewatch

medians of the two populations. The null hypothesis deems similar distributions in both sets, while the alternative hypothesis suggests the opposite. We use a p-value of 0.05 for the significance test. Note that unlike the t-test, which requires a normal distribution of the values, the U test does not have such constraints, therefore, is a more flexible measure.

3.1 Affective Feature

Psychological attributes such as sentiment and emotion can be closely related to mental health, as they can be indicators of a person's psychological well-being [9, 14]. We utilize several sentiment and emotion lexicons to find the prevalence of sentiment and emotion words in these two types of posts.

3.1.1 Coverage of sentiment lexicon. We examine the sentimental aspects, such as the presence of sentiment and opinion words in the posts of two forums based on two popular English sentiment lexicons, Opinion Lexicon [13] and VADER [15]. The Opinion Lexicon is a binary-level sentiment lexicon that includes approximately 6800 opinion words, each assigned with a polarity score of either -1 (negative) or +1 (positive). On the other hand, the VADER lexicon comprises around 7500 words and emoticons, with each term assigned with an integer polarity score ranging from -4 (strongly negative) to +4 (strongly positive).

Table 1: The presence (%) of sentiment words in the post of Depression and Suicidal groups based on two lexicons

Sentiment Lexicon	Depression	Suicidal	Statistically Significant
	Med./Mean/Std.	Med./Mean/Std.	
Opinion Lexicon	8.57/9.09/4.18	8.68/9.27/4.65	No
VADER	11.25/11.98/4.89	11.84/12.65/5.58	No

3.1.2 Coverage of emotion lexicon . We utilize an emotion lexicon, NRC Emotion Lexicon, proposed by Mohammad and Turney [22]. The NRC Emotion lexicon consists of a list of English words and corresponding values representing eight types of emotions (i.e., *anger, fear, anticipation, trust, surprise, sadness, joy,* and *disgust*). Here, we consider four negative emotions, such as *anger*, *fear*, *sadness*, and *disgust.*

Table 2: The presence (%) of four negative types of emotion words in the posts of Depression and Suicidal groups

Emotion type	Depression	Suicidal	Statistically Significant
	Med./Mean/Std.	Med./Mean/Std.	
Anger	1.94/2.38/2.36	2.16/2.60/2.58	No
Fear	2.38/2.87/2.72	2.99/3.61/3.22	Yes
Sadness	3.10/3.60/3.02	3.33/3.90/3.29	No
Disgust	1.37/1.80/2.10	1.55/2.03/2.37	Yes

3.1.3 Dominant emotion in post. In addition to determining the proportions of words representing different types of emotion, we also identify the dominant negative emotions at the post level. For this purpose, we utilize EmoNet [1], a free emotion recognition framework capable of identifying eight primary emotions: *joy, anticipation, surprise, trust, anger, disgust, fear,* and *sadness.* Considering that depressive or suicidal posts are likely to contain predominantly negative emotions, we specifically focus on the four negative emotions of *anger, disgust, fear,* and *sadness* in our analysis.

3.2 Grammatical and Semantic Feature

Linguistic features could be related to mental illness as shown in earlier studies [6, 17]. We consider a number of grammatical and semantic features in the two groups.

3.2.1 Subordinating conjunction. We study the presence of subordinating conjunctions that indicates the presence of complex sentence. Subordinating conjunctions are frequently employed to connect an independent clause with a dependent clause, resulting in the formation of complex sentences. Complex sentences are more difficult to process than simple sentences; nevertheless, they are likely to convey a clear and more informative message. A list of 50 commonly used subordinating conjunctions is considered in this study [4].

3.2.2 Adjectives & Verbs & Preposition & Article. The proportions of adjectives, verbs, and prepositions in each post from both forums, relative to the total words, are calculated, and various statistical measures are reported. The spaCy library [12] is employed to identify adjectives and verbs in the text. A list of the commonly used preposition is considered [5]. Besides, the percentages of articles (i.e., *a, an, the*) are investigated.

3.2.3 Negation words. The proportions of negation words in both groups are determined using the extended VADER [15] negation word list as a reference.

3.2.4 Presence of named entity. A named entity (NE) refers to a real-world entity, such as the name of a person or location. In this study, the analysis focuses on the three most common types of entities: person, location, and organization, as found in two types of posts. The mentions of person names (PER), geographical entities (GE) such as countries, cities, or similar references, and organizations (ORG) are retrieved using the spaCy library [12].

3.2.5 Deontic modals. Deontic modals are auxiliary verbs that express some kind of necessity, obligation, or moral recommendation [18]. We aim to find whether the content of these two forums shows any difference related to necessities or obligations based on the following four deontic modals: *must*, *should*, *ought*, and *need.*

3.3 Length Statistics

In addition, the following length-related statistics of the posts of two groups are compared: i) average post length (#word); ii) average post length (#sentence); iii) average sentence length (#words).

[4]https://github.com/sazzadcsedu/LinguisticAnalysis/blob/main/50_subordinate_clause.txt
[5]https://github.com/sazzadcsedu/LinguisticAnalysis/blob/main/preposition.png

Table 3: Distributions of dominant emotions (%) in Depression and Suicidal groups

Group	Anger	Disgust	Fear	Sadness
Depression	4628 (44.68%)	277 (2.67%)	1441 (13.91%)	4011 (38.72%)
Suicidal	4940 (49.55%)	309 (3.09%)	1477 (14.81%)	3242 (32.52%)

Table 4: Presence (%) of various linguistic and semantic features in the posts of Depression and Suicidal groups

Type	Depression	Suicidal	Statistically Significant
	Median/Mean/Std.	Median/Mean/Std.	
Subordinating Conjunction	6.36/ 6.41/ 3.20	6.12/ 6.14/ 3.50	No
Verb	14.72/ 14.99/ 3.60	15.28/ 15.64/ 4.40	No
Adjective	5.88/ 6.08/ 2.90	5.58/ 5.79/ 3.2	No
Preposition	8.28/ 8.27/ 3.19	8.27/ 8.30/ 3.50	No
Article	3.71/ 3.78/ 2.24	3.61/ 3.71/ 2.52	No
Negation	4.58/ 5.02/ 3.68	4.85/ 5.32/ 4.19	No
Deontic Modals	0.0/ 0.28/ 0.79	0.0 / 0.35 /1.07	No
Named Entity - GE	0.0/ 0.08/ 0.39	0.0/ 0.11/ 0.57	No
Named Entity - PER	0.0/ 0.15/ 0.55	0.0/ 0.15/ 0.86	No
Named Entity - ORG	0.0/ 0.16/ 0.68	0.0/ 0.16/ 0.94	No

Table 5: Statistics of length-related attributes in the Depression and Suicidal posts

Post Statistics	Depression	Suicidal	Statistically Significant
	Median/Mean/Std.	Median/Mean/Std.	
#Words	124/179.34/190.59	99.0/153.78/180.08	Yes
#Sentences	7/10.53/11.98	6.0/ 9.56/10.90	No
Sentence length (in words)	16.4/21.39/23.39	15.0/20.51/26.23	Yes

4 RESULTS, DISCUSSION AND FUTURE WORKS

As evident from Table 1, there are subtle differences in the sentiment features between the two groups, albeit not statistically significant as determined by the Mann-Whitney U Test. Specifically, mean and median values of sentiment words are slightly higher in the suicidal group based on both the Opinion Lexicon and VADER. Table 2 illustrates the presence of four types of emotion words, *anger*, *fear*, *sadness*, and *disgust* in the suicidal and depressive posts. We observe some differences in the presence of all four types of emotion words. The difference is more prominent for the *anger* and *fear* related emotion words, which are much higher in the *suicidal* group. The higher presence of all four types of emotion suggests that suicidal posts are likely to be more emotional. This finding is also in accord with the earlier observation by [23], who analyzed the difference between genuine and simulated suicidal notes.

In the analysis of dominant negative emotions using the emotion recognition framework (refer to Table 3), it is observed that *anger* is the most prevalent emotion among the four negative emotions in both the *suicidal* and *depression* groups. Furthermore, the dominant emotions in posts from both groups exhibit a similar pattern, with *anger*, *sadness*, *fear*, and *disgust* occurring in descending order of frequency. However, the dominance of *anger*-related emotions in the suicidal group is markedly higher compared to *sadness*, accounting for 49.55% of the emotions, as opposed to 32.52% for *sadness*. In contrast, the gap between *anger* and *sadness* in the depression group is relatively smaller, with *anger* accounting for 44.68% compared to 38.72% for *sadness*. These findings suggest that emotions associated with frustration, rage, and resentment, which are likely triggers of anger, are more prevalent in posts from the suicidal group. It should be noted that we observe differences in the relative precedence of the emotions *anger* and *sadness* at the word and post levels, which could potentially be influenced by factors such as the scope of features (i.e., word or post) and the type of resource or framework employed.

Table 4 presents a comparison of grammatical attributes, including subordinating conjunctions, verbs, adjectives, and prepositions, between the two groups. Mean and median values indicate that the grammatical attributes are highly similar in both types of posts, with differences typically below 5%. The Mann-Whitney U test confirms that these differences are not statistically significant. These findings suggest that the use of grammatical features alone may not be sufficient to effectively differentiate between posts representing generic depression and suicidal tendencies. Notably, the slightly higher presence of verbs in suicidal forums aligns with earlier studies by Gregory et al. (2018) and Schoene et al. (2016), who also reported a similar pattern in suicidal notes. Additionally, at the semantic level, the presence of negation words, named entities, and deontic modals exhibit similar patterns in both groups, as indicated in Table 4. However, differences are observed in the post-length statistics, as presented in Table 5, with the average length of suicidal text tending to be slightly shorter compared to depressive text, which could be a significant distinguishing factor. Nevertheless, the average sentence length remains similar in both groups.

This preliminary study analyzing a number of features followed by the significance test reveals that only a limited number of length-related and affective features demonstrate discriminative capabilities for distinguishing these two types of posts. The high similarity and lack of apparent distinction among various attributes suggest a strong connection between depression and suicidal tendencies which is very challenging to separate.

Despite predominantly negative results and findings, this study lays the groundwork for future analyses by highlighting the limitations of conventional attributes in distinguishing between these two types of posts. Moreover, these findings imply that, in addition to closely monitoring suicidal forums, it is crucial to also monitor depression forums, as users of these forums may be at risk of suicidal ideation. Future research will involve a more comprehensive analysis with additional features and larger datasets to identify better signals. Furthermore, manual intervention in the data labeling process will be incorporated to enhance the quality of the currently automatically labeled data.

5 ETHICAL STATEMENT

This study utilizes Reddit data publicly available on Kaggle[6]. The research ensures that no user identity information is collected, used, or disclosed during the analysis or afterwards.

REFERENCES

[1] Muhammad Abdul-Mageed and Lyle Ungar. 2017. EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 718–728. https://doi.org/10.18653/v1/P17-1067

[2] Mario Ezra Aragón, Adrián Pastor López Monroy, Luis Carlos González-Gurrola, and Manuel Montes. 2019. Detecting depression in social media using fine-grained emotions. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. 1481–1486.

[3] 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 (2017), 649–685.

[4] Raymond Chiong, Gregorius Satia Budhi, Sandeep Dhakal, and Fabian Chiong. 2021. A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine* 135 (2021), 104499.

[5] 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*. 31–39.

[6] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media*, Vol. 7. 128–137.

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

[8] L Eisenberg. 1975. Primary prevention and early detection in mental illness. *Bulletin of the New York Academy of Medicine* 51, 1 (1975), 118.

[9] Barbara L Fredrickson. 2001. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist* 56, 3 (2001), 218.

[6]https://www.kaggle.com/

[10] Mary Jane Friedrich. 2017. Depression is the leading cause of disability around the world. *Jama* 317, 15 (2017), 1517–1517.

[11] Amanda Hess. 2015. Please Do Not Downvote Anyone Who's Asked for Help'. *Slate, Tuesday* 3 (2015).

[12] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). To appear.

[13] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 168–177.

[14] Tianqiang Hu, Dajun Zhang, Jinliang Wang, Ritesh Mistry, Guangming Ran, and Xinqiang Wang. 2014. Relation between emotion regulation and mental health: A meta-analysis review. *Psychological reports* 114, 2 (2014), 341–362.

[15] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

[16] Katherine M Keyes, Dahsan Gary, Patrick M O'Malley, Ava Hamilton, and John Schulenberg. 2019. Recent increases in depressive symptoms among US adolescents: trends from 1991 to 2018. *Social psychiatry and psychiatric epidemiology* 54 (2019), 987–996.

[17] Gina R Kuperberg, Philip K McGuire, and Anthony S David. 1998. Reduced sensitivity to linguistic context in schizophrenic thought disorder: evidence from on-line monitoring for words in linguistically anomalous sentences. *Journal of abnormal psychology* 107, 3 (1998), 423.

[18] Lauren Levine. 2022. The Distribution of Deontic Modals in Jane Austen's Mature Novels. In *Proceedings of the 6th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. 70–74.

[19] Xueying Liu, Shiaofen Fang, George Mohler, Joan Carlson, and Yunyu Xiao. 2022. Time-to-event modeling of subreddits transitions to r/SuicideWatch. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 974–979.

[20] Daniel M Low, Kelly L Zuromski, Daniel Kessler, Satrajit S Ghosh, Matthew Nock, and Walter Dempsey. 2021. It's quality and quantity: the effect of the amount of comments on online suicidal posts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, Vol. 2021. NIH Public Access, 95.

[21] Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics* (1947), 50–60.

[22] Saif Mohammad and Peter Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*. 26–34.

[23] Daniel M Ogilvie, Philip J Stone, Edwin S Shneidman, and PJ Stone. 2009. Some characteristics of genuine versus simulated suicide notes. *The content analysis reader* (2009), 404–409.

[24] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. 2018. Deep learning for depression detection of twitter users. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*. 88–97.

[25] Minsu Park, Chiyoung Cha, and Meeyoung Cha. 2012. Depressive moods of users portrayed in Twitter. In *Proceedings of the 18th ACM International Conference on Knowledge Discovery and Data Mining, SIGKDD 2012*. 1–8.

[26] Inna Pirina and Çağrı Çöltekin. 2018. Identifying depression on reddit: The effect of training data. In *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task*. 9–12.

[27] Esteban A Ríssola, Seyed Ali Bahrainian, and Fabio Crestani. 2020. A dataset for research on depression in social media. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 338–342.

[28] Salim Sazzed. 2022. The Dynamics of Ukraine-Russian Conflict through the Lens of Demographically Diverse Twitter Data. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 6018–6024.

[29] Salim Sazzed. 2022. Stylometric and Semantic Analysis of Demographically Diverse Non-native English Review Data. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 470–476.

[30] Salim Sazzed. 2023. Discourse Mode Categorization of Bengali Social Media Health Text. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis@ACL*. 52–57.

[31] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. 2017. Depression detection via harvesting social media: A multimodal dictionary learning solution.. In *IJCAI*. 3838–3844.

[32] Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*. 25–36.

[33] Michael M Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2019. Detection of depression-related posts in reddit social media forum. *IEEE Access* 7 (2019), 44883–44893.

[34] Alina Trifan, Rui Antunes, Sérgio Matos, and Jose Luís Oliveira. 2020. Understanding depression from psycholinguistic patterns in social media texts. In *European Conference on Information Retrieval*. Springer, 402–409.

[35] Frank Wilcoxon. 1992. *Individual comparisons by ranking methods*. Springer.

[36] Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. *arXiv preprint arXiv:1709.01848* (2017).

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