

A Dataset for Research on Modelling Depression Severity in Online Forum Data

Isuri A. Nanomi Arachchige, Ruvan Weerasinghe

University of Colombo School of Computing,
Sri Lanka.

isurianuradha96@gmail.com
arw@ucsc.cmb.ac.lk

Vihangi H. Jayasuriya

University of Kelaniya, Sri Lanka.
vihahimaya0717@gmail.com

Abstract

People utilize online forums to either look for information or to contribute it. Because of their growing popularity, certain online forums have been created specifically to provide support, assistance, and opinions for people suffering from mental illness. Depression is one of the most frequent psychological illnesses worldwide. People communicate more with online forums to find answers for their psychological disease. However, there is no mechanism to measure the severity of depression in each post and give higher importance to those who are diagnosed more severely depressed.

Despite the fact that numerous researches based on online forum data and the identification of depression have been conducted, the severity of depression is rarely explored. In addition, the absence of datasets will stymie the development of novel diagnostic procedures for practitioners. From this study, we offer a dataset to support research on depression severity evaluation. The computational approach to measure an automatic process, identified severity of depression here is quite novel approach. Nonetheless, this elaborate measuring severity of depression in online forum posts is needed to ensure the measurement scales used in our research meets the expected norms of scientific research.

1 Introduction

With technological advancement, online support forums (OSF) are popular among society and

attempt to seek solutions for their physical, social and psychological problems (Dosani, Harding, and Wilson 2014). As subgroups of OSF, online mental support forums (OMSF) are designed primarily for psychological expertise and experienced peer groups to provide support and guidance to the psychologically impaired community.

Early identifying and analysing the severity of depression is crucial for: (a) providing suitable medications and therapeutic processes immediately after diagnosing (b) preventing fatal accidents such as deliberate self-harms or suicides. Depression is greater than definitely feeling unhappy or bored stiff for a few days. The majority undergo durations of feeling down, but whilst you're depressed you experience persistently unhappy for weeks or months, in preference to only some days. The best information is that with the proper treatment and support, the majority with depression could make a complete recuperation (MacKin et al. 2014).

To prevent from those tragic situations, in clinical psychology several psychological assessments such as DSM -V, BDI were defined to measure the severity of depression disorder. Among those assessments, DSM-V criteria are widely used to diagnose clinical depression disorder and have been classified into different severity levels according to their number of symptoms, duration and cause for depressive episodes. Table 1 elaborates examples of forum posts categorised into four severity classes of depression according to different psychological criteria: (a) no-depression; (b) Mild depression; (c) Moderate depression; (d) Severe depression.

Recent work on modelling social media data in Natural language processing (NLP) has focused

on distinguishing depressive posts from non-depressive posts (Islam et al. 2018; Jeri-Yabar et al. 2019; Shrestha and Spezzano 2019). Moreover, limited datasets are available due to the difficulties such as complex nature, time consumption, etc, that exist in the process of data analysing. However, there are no previous studies to access the severity of depression based on OMSF data.

Severity Scales	DSM-V criteria	Beck's Inventory
No Depression	0-7	<=10
Mild Depression	8-16	10-18
Moderate Depression	17-23	19-29
Severe Depression	>24	>= 30

Table 1: Psychological Criteria used for measuring severity of depression. (The DSM-V outlines the following main criterion to make a diagnosis of depression severity. Recurrent thoughts of deaths, recurrent suicidal ideation without a specific plan for committing suicide. To receive a diagnosis of depression these main criteria must cause the individual clinically significant distress.)

This paper presents a systematic study on analysing the severity of depression disorder from forum posts with computational methods and the following are our main contributions:

- Preparing dataset with four severity levels of depression using automatic labelling method.
- Create a new classification task for identifying different severity levels of depression from forum posts.
- Evaluate deep learning algorithms combined with linguistic information.

The rest of this paper is structured as follows: Section 2 gives a brief overview of existing and similar approaches used to monitor the severity of depression; Section 3 presents our aim and motivation to conduct the research; Section 4 outlines the proposing data and tasks designed for automatic labelling process and the results obtained; Finally, the paper is concluded with Section 5.

2 Related Works

2.1 Psychological Assessments

Depression diagnosis requires five or more symptoms according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V)

(Medicaidmentalhealth.org 2018). One of them must be either Depressed mood or Anhedonia, named main criteria. Although the secondary symptoms can be divided into somatic and non-somatic clusters, the DSM-5 identify depression in all or none fashion. In contrast, depression severity is a continuous variable (American Psychiatric and Association 2013).

The Beck Depression Inventory (BDI) is widely used to screen for depression and to measure behavioral manifestations and severity of depression. The BDI can be used for ages 13 to 80. The inventory contains 21 self-report items which individuals complete using multiple choice response formats. The BDI takes approximately 10 minutes to complete. Validity and reliability of the BDI has been tested across populations, worldwide (Beck et al. 1961).

The Center for Epidemiologic Studies Depression Scale (CES-D) was designed for use in the general population and is now used as a screener for depression in primary care settings. It includes 20 self-report items, scored on a 4-point scale, which measure major dimensions such as restless sleep, poor apatite and feeling lonely of depression experienced in the past week. The CES-D can be used for children as young as 6 and through older adulthood. It has been tested across gender and cultural populations and maintains consistent validity and reliability. The scale takes about 20 minutes to administer, including scoring (Radloff 1977). The CES-D Scale: A self-report depression scale for research in the general population (Wright 1985). The Hamilton Rating Scale for Depression (Hamilton 1960), abbreviated HDRS, HRSD or HAM-D, measures depression in individuals before, during and after treatment. The scale is administered by a health care professional and contains 21 items, but is scored based on the first 17 items, which are measured either on 5-point or 3-point scales. It takes 15 to 20 minutes to complete and score.

The Children's Depression Inventory is a modification of the BDI for adults. The CDI is now on its second edition. It assesses depression severity in children and adolescents 7 to 17 years old. Two scales measure emotional problems and functional problems. Three separate rater forms are available: one for parents (17 items), one for teachers (12 items) and a self-report (28 items). Administration time is between five and 15 minutes (Kovacs 1981).

2.2 Linguistic Features in Depression

One of the functions of language is to communicate our inner monologue to others. Depressed persons, predictably, utilize more negative emotion words and fewer positive emotion words than nondepressed people. Depressed college students, for example, used more negatively valenced phrases when writing about their "deepest thoughts and feelings about being in university". Following is an example for language characteristic of people diagnosed with severely depressed.

"In May I was admitted into hospital with Severe Depression. This was my first ever admission and was discharged after a long hard 9 weeks. Since being home, I seem to have some really good days, then suddenly start getting feelings of sadness, insomnia, and feeling worthless. Workcover is now looking at a return-to-work project and I am so worried at what will happen when one day I am coping and the next plotting my death. Do others experience these highs and lows, and how do manage day to day"

Language style refers to how we express ourselves rather than the content we express. "Absolutist words" – words that convey absolute magnitudes or probabilities, such as "always", "nothing", or "completely" – were discovered to be better markers for mental health forums than pronouns or negative emotion words (Holtzman 2017).

2.3 Depression and Emotion words

People suffering from depression may not be able to identify it by name; they rarely used the word "depressed" when describing their feelings, instead opting for words like "down," "stressed," or "upset". So, the key is to spot a pattern over time many words from depressed children and adults, such as bad, sad, helpless, hopeless, aching, lost, worthless, useless, stupid, stuck, untethered, adrift, hurting, alone, afraid, unsure, insecure, despair, black and blue, to name a few," These words are frequently used when describing themselves or their personal feelings (Houben, Van Den Noortgate, and Kuppens 2015; Palmier-Claus et al. 2012).

2.4 Forum Data & Depression Identification

A significant number of experiments have been developed to analyse the depressive content posted by depressed users on social networks. Most studies focused on social network data such as Facebook (De Choudhury et al. 2014; Islam et al. 2018), Twitter (Ahmed Hussein Orabi, Prasadith Buddhitha, Mahmoud Hussein Orabi 2018; Gaikar et al. 2019; Sasso et al. 2019; Tsugawa et al. 2015; Yates, Cohan, and Goharian 2017) compared to online forum data for analysing depressive posts or identifying depressed individuals. Online forums like Reddit (Tadesse et al. 2019a, 2019b; Yates et al. 2017; Yoo, Lee, and Ha 2019; Zirikly et al. 2019), ReachOut.com (Cohan, Young, and Goharian 2016; Shrestha and Spezzano 2019; Yates et al. 2017), MedHelp (Zhang, Cho and Zhai, 2015; Vydiswaran and Reddy, 2019) used to mine textual posts related to depression. They used different feature extraction techniques to derive domain-specific features related to depression. Recent works adapted LIWC (Ahmed Hussein Orabi, Prasadith Buddhitha, Mahmoud Hussein Orabi 2018; Coppersmith et al. 2015; Yoo et al. 2019), Psychological scales (DASS-21) (Havigerová et al. 2019) and emotional lexicons (Liu 2015) with NLP techniques.

It is often supported by utilising various machine learning approaches to different NLP techniques (Tadesse et al. 2019a, 2019b; Zhang et al. 2015). Moreover, some studies employed classification algorithms like Decision Trees (Islam et al. 2018), Naive Bayes (Biyani et al. 2014), and Random Forest (Cacheda et al. 2019; Shrestha and Spezzano 2019). With recent advances in neural network models in NLP, a novel contribution for identification of depression has arisen with the implementation of advanced deep learning architectures to outperform more traditional machine learning methods.

When assessing the severity of depression, there were fewer studies compared to depression identification. (Fatima et al. 2018) developed a framework to classify the degree of depression in social communities based on the mood tags (ex: Rejected, surprised, lazy). The framework used the ANEW lexicon (Margaret M. Bradley and Peter Lang 1999) and valence value to calculate the severity of depression. However, in real life, clinical experts use psychological scales and more

UserId	Headline	Text	No. of Replies
mapz	I [selfharmd] myself again	I have been feeling down the whole week but doing everything I can to fight it. Last night, I felt suffocated, my chest was heavy and it's hard to breathe. I knew it by then I had to do something. I looked for my blade and locked myself in the shower. I cut my skin until I saw blood running with the water. It was satisfying.It's like my immediate cure right now.	5
proactiveBanana5575	Another depressed person	Hey everybody or anybody, honestly, I wouldn't be making this if it wasn't anonymous and anybody here knew me IRL. It feels humiliating and almost illegal to talk about this with people I know. I feel like most people in my life view depression as laziness or a way to garner sympathy as if I would throw away my life just to make people feel sorry for me. I honestly find it very hard to explain.	6

Table 2: Example for extracted forum posts

factors to judge the severity of depression and those specific terms not included in the ANEW lexicon. Also, another research has conducted to measure severity of major depression based on self-reports (R.C. et al. 2016).

We first described the psychological scales and their linguistic patterns used for the severity assessment of depression. Then we mentioned studies that use different psychological assessments to capture depression in the online forum posts. We also discuss the approaches used for depression identification related to taxonomies, symptom associations with linguistic styles, and machine learning techniques. According to our analysis, most of the research works focused only to identify the depressive contents expressed by depressed individuals. However, in this paper, we propose a model for predicting the severity of depression as an important factor in determining the treatment type and urgency.

3 Motivation

In previous studies, forum data was annotated as depressive or non-depressive posts and but with no information regarding the severity level of depression in each post based on the individual's thought process.

We aim to build a level of severity for each forum post, annotated dataset for depression severity assessment using an automatic annotation approach which enables us to extract evidence

related to depression. Moreover, to predict the level of depression severity from posts available in OMSF.

4 Methodology

4.1 Tasks & Data

We selected four different OMSF to extract posts tagged under the category of depression. Following standards were considered in the selection process of forum posts.

- Post should consist minimum of 2 sentences.
- Post should post under the category of depression.
- Forum thread should have minimum 3 posts posted by poster.
- Substance abuse, OCD, Anxiety etc other psychological illness were not considered.

We have defined two tasks to carry out in the process. The first task is designed for automatic labelling of OMSF data into different severity levels. The second task is defined as a multi-label classification task.

4.2 Depression Severity Categories

In automatic depression severity annotation, we adapt the criteria defined by DSM-V and Beck's Inventory. Table 1 illustrates the scores assigned according to the symptoms of depression.

Note that the severity levels of depression were classified according to the symptoms that exist on the forum post.

4.3 Text Processing

We used four OMSF to extract the posts and all the posts were posted under the depression category.

- Beyond Blue Online Forum
- Understand Forum
- 7 Cup Forum
- Depression Forums

A specific web scraper is developed from scratch with python libraries to extract online forum data. After the extraction of following features, data were stored in a relational database.

1. **User ID** identifier for each user who posted the personal information
2. **Headline** related each post.
3. **Text** body of the depression post
4. **Number of Replies** count the number of forum posts in individual thread posted by the user.

After the collection of dataset, unwanted texts such as formatting, HTML links and quotations of earlier comments were removed (Petz et al. 2012). Table 2 illustrates examples of forum posts. For automatic annotation, individual forum posts (P) were divided into sequence of individual sentences (S). Each sentence was considered as a token and forum post consist of sequence of tokens.

$$P = \{S1 + S2 + \dots + Sn\}$$

4.4 Depression Severity Annotation

Though we collected posts related to depression from different OMSF, some forums posts were irrelevant to depression disorder. Therefore, the dataset has been annotated manually by two psychological experts and one computational linguistic expert. The severity annotation process followed a similar procedure to that defined by Moreno (Moreno et al. 2011). Three annotators were asked to determine which posts can be considered as a reference to depression following the DSM-V criteria and to remove non-depression posts from the dataset.

The three annotators achieved a pairwise Cohen's Kappa score ranging between 0.62687

and 0.7192. But achieving a high inter-rater agreement can be a difficult task due complex nature of the posts. Also, 215 non-depressive forum posts were removed, and 157 posts were selected based on majority voting method between the annotators. Manually annotated dataset were used mainly for validating purpose.

The dataset consists of 2140 data points that remained after data filtration and were subjected to automatic data labelling. According to the defined criteria of Table 1, we used domain-specific keywords and sentences (DSM-V and Beck's inventory) for the process of automatic labelling (AL). Two different computational approaches were designed and implemented.

- Based on the taxonomical similarity and n-gram models (AL₁).
- Based on the sentence similarity and pre-trained transformer models (AL₂).

Lexical similarity and n-gram models (AL₁)

In the first approach, we created a lexicon which based on the symptoms of depression defined in DSM-V criteria. In order to build the lexicon, previous taxonomies associated with depression, gathered from online vocabularies were taken into consideration. Moreover, all the pronouns and the negations were included on this lexicon with the help of POS tags generated by Stanford Parser. Furthermore, all the possible sets of words and synonyms relative to symptoms of depression were extracted from Wordnet and added into the lexicon. The lexicon includes all the words relevant to the symptoms which belong to different severity levels. To be more accurate, we validated our lexicon with the aid of clinical psychologists (co-authors of the paper).

N Gram Analysis: We considered lexicon as a unigram model to calculate the word level similarity. Based on the severity levels in depression, we have divided the lexicon into three categories and assigned labels for each category (1, 2 and 3). We use cosine similarity and extract string matching between each post and each category of depression severity lexicon. The goal was to obtain a ranked list of the posts according to the depression severity with the lexicon. Thus, the higher ranked forum posts were those with the highest severity level of depression.

Sentence Similarity and Pre-trained Transformer models. (AL₂)

In the second approach we have selected Beck's Depression Inventory to measure the degree of severity. The Beck Depression Inventory (BDI) is a 21-item test presented in multiple-choice format, which measures presence and degree of depression in adolescents and adults consistent with the DSM-IV. It is not intended as a diagnostic instrument. The BDI evaluates 21 symptoms of depression, 15 of which cover emotions, four cover behavioral changes, and six somatic symptoms.

Forum posts were split into sentences and similarities were checked according to BDI score using Universal sentence encoder model.

Universal Encoder Model: The Universal Sentence Encoder encodes text into high dimensional vectors to measure semantic similarity. We trained our model using the sentences defined in Becks Inventory and assign a score for each sentence. And calculate an individual score using cosine similarity for the score. Figure 1 and Figure 2 illustrated the comparison of results obtained from the two-domain specific automatic labelling models and the manually annotated validation dataset. Following Figures demonstrate the distribution of posts relevant different severity levels.

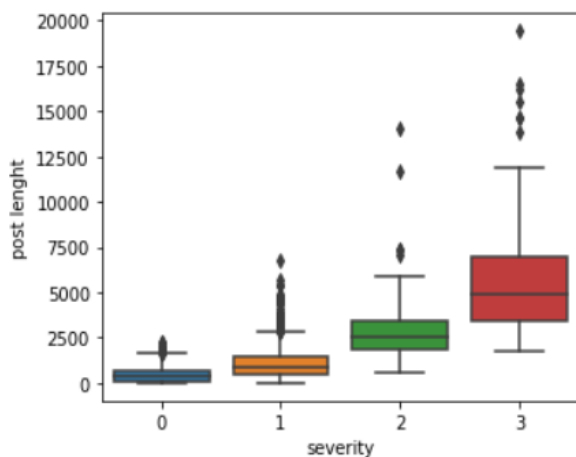


Figure 1: Distribution of the words in posts belongs to different severity level (AL₁)

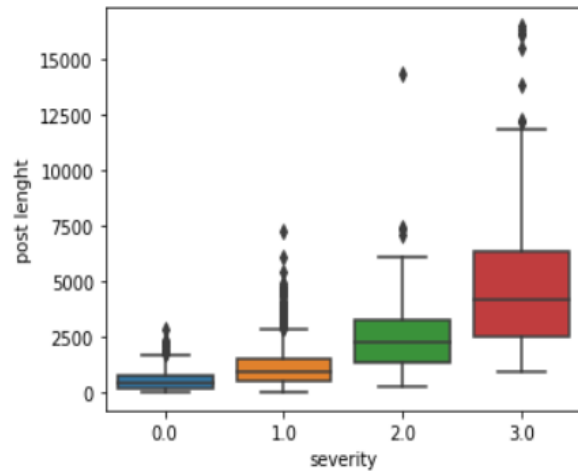


Figure 2: Distribution of the words in posts belongs to different severity level (AL₂)

4.5 Multiclass Classification Model

Since depression severity prediction is a new task, our goal is to classify and identify the forum posts that show some evidence of different severity levels. The effectiveness is evaluated using classic metrics such as Precision, Recall and *F1*.

4.6 Experiments and Results

We have selected different classification algorithms and common textual features to perform multiclass classification. Also, the features are the following: (1) Unigrams: represents the raw frequency count of the words extracted from posts; (2) Unigram-TFIDF: list all the relevant words extract from posts; (3) Word2Vec Embedding: captures all the words appear in similar posts; (4) Global Vectors (GloVe) Embedding: focuses on words co-occurrences over the forum posts; (5) Bidirectional Encoder Representation from Transformer (BERT) Embeddings: defines the contextual language representations derived from the posts.

In general, supervised model, input features are not abstracted in high-level and try to predict output by mathematical transformation on a subset of input features. But in deep learning all the abstraction of the input features are formed automatically and internally. Therefore, we have employed both supervised and deep learning classification algorithms for multiclass prediction based on the annotated dataset. Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Neural

AL ₁	SVM(%)	MNB(%)	LR(%)	NN(%)	CNN(%)	LSTM(%)
Unigram-count	75.6521	75.21739	81.30431	80.86956	77.36452	78.260868
Unigram-TFIDF	80.8695	65.65217	78.26084	77.39130	79.00123	79.130434
Word2Vec Embed	66.3043	72.28262	64.13043	67.39130	63.82624	67.391305
GloVe Embed	67.4305	67.39132	59.00037	62.07066	69.82653	77.173913
BERT Embed	66.5429	71.63469	60.43581	68.75397	70.25342	75.986076

Table 3: Accuracy of selected classification models (AL₁)

AL ₂	SVM(%)	MNB(%)	LR(%)	NN(%)	CNN(%)	LSTM(%)
Unigram-count	76.6738	71.4578	77.5647	82.1537	70.7480	77.8947
Unigram-TFIDF	79.7383	63.7483	74.8363	78.7332	71.3678	75.3561
Word2Vec Embed	62.4647	63.6758	66.6452	69.7364	60.5467	66.4356
GloVe Embed	69.3574	61.2453	55.2435	60.4369	65.3467	72.6789
BERT Embed	67.9862	72.6728	56.2453	71.5363	72.5242	74.7355

Table 4: Accuracy of selected classification models (AL₂)

AL ₁	Measure	SVM(%)	MNB(%)	LR(%)	NN(%)	CNN(%)	LSTM(%)
Unigram-count	Precision	70.8737	70.7840	72.7851	78.6436	74.7643	74.0642
	Recall	63.2637	74.4673	75.3645	82.3564	72.3549	81.3547
	F measure	66.8565	72.5795	74.0523	80.4571	73.5398	77.5384
Unigram-TFIDF	Precision	73.8960	70.1862	73.8765	80.5438	70.8752	77.8764
	Recall	78.5349	68.8642	79.8765	80.9785	74.8753	80.8753
	F measure	76.1453	69.5191	76.7594	80.7605	72.8203	79.3475
Word2Vec Embedding	Precision	63.9863	64.8912	61.5679	76.9864	69.7641	70.4591
	Recall	69.1429	60.5684	66.8751	74.7641	72.1237	76.1278
	F measure	66.4647	62.6553	64.1118	75.8589	70.9242	73.1838
GloVe Embedding	Precision	61.7865	57.8753	52.5694	63.2306	62.9348	74.7535
	Recall	59.7632	53.6754	57.7531	66.8754	71.6539	78.6534
	F measure	60.7580	55.6962	55.0394	65.0019	67.0119	76.6538
BERT Embedding	Precision	60.2341	62.8961	57.3462	67.3428	62.7354	71.3542
	Recall	67.2319	60.3672	51.8364	71.3648	68.3452	70.6351
	F measure	63.5409	61.6057	54.4522	69.2954	65.4202	70.9928

Table 5: Evaluation of AL₁ using the validation dataset

Network (NN), Conventional Neural Network (CNN) and Long-Short Term Memory (LSTM) were used for multi classification task. Moreover, the classification models were trained on the automatically derived sets and evaluated using the validated set (manually annotated posts).

We observe from the above results, an overall higher recall (Average recall) is achieved in approach AL₁. Also, a higher precision (Average recall) was obtained in AL₂. According to the nature of the proposed approaches, higher recall is usually preferable. Subsequently, *F1* measure concludes that any proposed approaches can effectively use to automatically. A set of forum

posts and used such set to train the classification model.

5 Conclusion

There has been growing research interest towards the identification changes of the mental state through the exploitation of social data. The absence of resources, such as datasets might limit the development of cutting-edge technologies to support health practitioners. We developed a dataset and automatic process for labelling the forum posts with four severity levels of depression.

AL ₂	Measure	SVM(%)	MNB(%)	LR(%)	NN(%)	CNN(%)	LSTM(%)
Unigram-count	Precision	73.641	70.398	78.893	79.349	75.437	79.832
	Recall	77.562	68.734	79.345	80.537	71.671	81.563
	F measure	75.550	69.556	79.118	79.938	73.505	80.688
Unigram-TFIDF	Precision	78.526	71.638	80.353	82.637	71.837	81.416
	Recall	73.637	66.671	77.243	75.872	67.239	78.321
	F measure	76.002	69.065	78.767	79.101	69.461	79.838
Word2Vec Embedding	Precision	67.239	67.354	60.253	77.635	63.251	69.353
	Recall	63.826	69.683	56.271	75.297	61.872	67.353
	F measure	65.488	68.498	58.193	76.448	62.553	68.338
GloVe Embedding	Precision	68.234	62.546	53.836	64.638	67.347	78.748
	Recall	64.356	64.237	57.354	60.234	68.645	77.353
	F measure	66.238	63.380	55.539	62.358	67.989	78.044
BERT Embedding	Precision	63.345	67.463	56.637	69.354	67.834	74.363
	Recall	59.372	60.354	50.363	69.023	66.354	73.423
	F measure	61.294	63.710	53.316	69.188	67.085	73.890

Table 6: Evaluation of AL₂ using the validation dataset

Also, we modelled depression severity level prediction as a new multi-class classification task and conducted experiments using different deep learning-based algorithms combined with linguistic features. The linguistic features contribute to this study were obtained by considering characteristics like frequency, importance of words in given information. According to the results obtained deep learning algorithms performs better than general classification algorithms, when analysing unstructured forum data. Since this research is ongoing research, we have considered only quarter of our dataset for this experiment. In future, we are planning to train this model with full dataset which requires huge computation time. Also, we are planning improve the overall accuracies of the models by optimizing the classification algorithms. Nowadays, emoji plays a fundamental role in human computer-mediated communications, allowing the latter to convey body language, objects, symbols, or ideas in text messages. The relationship between language, emoji, and emotions is now being studied by several disciplines such as linguistics, psychology, natural language processing (NLP), and machine learning (ML).

Therefore, we hold the opinion that in the future, could also benefit from a more qualitative approach, such as interviews and case studies to learn about emoji use in the context of real-world communication. In the future, we plan to identify

the progression of depression severity within the forum thread in OMSF.

Acknowledgement

We acknowledge Miss. Priyadharshany Sandanapitchai from the Rutgers, The State University of New Jersey for support and guidance given on psychology domain area.

References

- Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, Diana Inkpen. 2018. [Deep Learning for Depression Detection of Twitter Users](#). In *Proceedings Of the Fifth Workshop on Computational Linguistics and Clinical Psychology*, pages 175–80.
- American Psychiatric and Association. 2013. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. In *United States*.
- Beck, A. T., C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh. 1961. An Inventory for Measuring Depression. *Archives of General Psychiatry* 4(6):561–71.
- Biyani, Prakhar, Sumit Bhatia, Cornelia Caragea, and Prasenjit Mitra. 2014. [Using Non-Lexical Features for Identifying Factual and Opinionative Threads in Online Forums](#). *Knowledge-Based Systems* 69(1):170–78.
- Cacheda, Fidel, Diego Fernandez, Francisco J. Novoa, and Victor Carneiro. 2019. [Early Detection of Depression: Social Network Analysis and Random Forest Techniques](#). *Journal of Medical Internet Research* 21(6).

- De Choudhury, Munmun, Scott Counts, Eric J. Horvitz, and Aaron Hoff. 2014. [Characterizing and Predicting Postpartum Depression from Shared Facebook Data](#). In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, pages 625–37.
- Cohan, Arman, Sydney Young, and Nazli Goharian. 2016. [Triaging Mental Health Forum Posts](#). 143–47.
- Coppersmith, Glen, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. [CLPsych 2015 Shared Task: Depression and PTSD on Twitter](#). 1–10.
- Dosani, Sabina, Claire Harding, and Simon Wilson. 2014. [Online Groups and Patient Forums](#). *Current Psychiatry Reports* 16(11).
- Fatima, Iram, Hamid Mukhtar, Hafiz Farooq Ahmad, and Kashif Rajpoot. 2018. [Analysis of User-Generated Content from Online Social Communities to Characterise and Predict Depression Degree](#). *Journal of Information Science* 44(5):683–95.
- Gaikar, Mrunal, Jayesh Chavan, Kunal Indore, and Rajashree Shedge. 2019. [Depression Detection and Prevention System by Analysing Tweets](#). *SSRN Electronic Journal* 1–6.
- Hamilton, M. 1960. [A Rating Scale for Depression](#). *Journal of Neurology, Neurosurgery, and Psychiatry* 23(1):56–62.
- Havigerová, Jana M., Jiří Haviger, Dalibor Kučera, and Petra Hoffmannová. 2019. [Text-Based Detection of the Risk of Depression](#). *Frontiers in Psychology* 10(MAR):1–11.
- Holtzman, Nicholas S. 2017. [A Meta-Analysis of Correlations between Depression and First Person Singular Pronoun Use](#). *Journal of Research in Personality* 68:63–68.
- Houben, Marlies, Wim Van Den Noortgate, and Peter Kuppens. 2015. [The Relation between Short-Term Emotion Dynamics and Psychological Well-Being: A Meta-Analysis](#). *Psychological Bulletin* 141(4):901.
- Islam, Md. Rafiqul, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, and Anwaar Ulhaq. 2018. [Depression Detection from Social Network Data Using Machine Learning Techniques](#). *Health Information Science and Systems* 6(1):0–12.
- Jeri-Yabar, Antoine, Alejandra Sanchez-Carbonel, Karen Tito, Jimena Ramirez-delCastillo, Alessandra Torres-Alcantara, Daniela Denegri, and Yhuri Carreazo. 2019. [Association between Social Media Use \(Twitter, Instagram, Facebook\) and Depressive Symptoms: Are Twitter Users at Higher Risk?](#) *International Journal of Social Psychiatry* 65(1):14–19.
- Kovacs, Maria. 1981. [Rating Scales to Assess Depression in School-Aged Children](#). *Acta Paedopsychiatrica: International Journal of Child & Adolescent Psychiatry*.
- Liu, Bing. 2015. [Sentiment Analysis: Mining Opinions, Sentiments, and Emotions](#). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (May):1–367.
- MacKin, R. Scott, J. Craig Nelson, Kevin L. Delucchi, Patrick J. Raue, Derek D. Satre, Dimitris N. Kiosses, George S. Alexopoulos, and Patricia A. Arean. 2014. [Association of Age at Depression Onset with Cognitive Functioning in Individuals with Late-Life Depression and Executive Dysfunction](#). *American Journal of Geriatric Psychiatry* 22(12):1633–41.
- Margaret M. Bradley and Peter Lang. 1999. [Affective Norms for English Words \(ANEW\): Instruction Manual and Affective Ratings](#). *Gainesville, Fla, NIMH Centre for the Study Of Emotion and Attention, University Of Florida*.
- Medicaidmentalhealth.org. 2018. [DSM-5 Criteria: Major Depressive Disorder Treatment of Major Depressive Disorder](#). *DSM-5 Criteria: Major Depressive Disorder* 24–32.
- Moreno, Megan A., Lauren A. Jelenchick, Katie G. Egan, Elizabeth Cox, Henry Young, Kerry E. Gannon, and Tara Becker. 2011. [Feeling Bad on Facebook: Depression Disclosures by College Students on a Social Networking Site](#). *Depression and Anxiety* 28(6):447–55.
- Palmier-Claus, J. E., Pamela J. Taylor, P. Gooding, G. Dunn, and S. W. Lewis. 2012. [Affective Variability Predicts Suicidal Ideation in Individuals at Ultra-High Risk of Developing Psychosis: An Experience Sampling Study](#). *British Journal of Clinical Psychology* 51(1):72–83.
- Petz, Gerald, Michał Karpowicz, Harald Fürschuß, Andreas Auinger, Stephan M. Winkler, Susanne Schaller, and Andreas Holzinger. 2012. [On Text Preprocessing for Opinion Mining Outside of Laboratory Environments](#). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7669 LNCS:618–29.
- R.C., Kessler, Van Loo H.M., Wardenaar K.J., Bossarte R.M., Brenner L.A., Cai T., Ebert D.D., Hwang I., Li J., De Jonge P., Nierenberg A.A., Petukhova M.V., Rosellini A.J., Sampson N.A., Schoevers R.A., and Wilcox M.A. 2016. [Testing a Machine-Learning Algorithm to Predict the Persistence and Severity of Major Depressive](#)

- Disorder from Baseline Self-Reports. *Molecular Psychiatry* 21(10):1366–71.
- Radloff, Lenore Sawyer. 1977. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* 1(3):385–401.
- Sasso, Maria P., Annaleis K. Giovanetti, Anastasia L. Schied, Hugh H. Burke, and Gerald J. Haefel. 2019. #Sad: Twitter Content Predicts Changes in Cognitive Vulnerability and Depressive Symptoms. *Cognitive Therapy and Research* 43(4):657–65.
- Shrestha, Anu and Francesca Spezzano. 2019. [Detecting Depressed Users in Online Forums](#). In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*, pages 945–51.
- Tadesse, Michael M., Hongfei Lin, Bo Xu, and Liang Yang. 2019a. Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access* 7(c):44883–93.
- Tadesse, Michael M., Hongfei Lin, Bo Xu, and Liang Yang. 2019b. Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access* 7:44883–93.
- Tsugawa, Sho, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. 2015. [Recognizing Depression from Twitter Activity](#). *Conference on Human Factors in Computing Systems - Proceedings 2015-April*:3187–96.
- Vydiswaran, V. G. Vino. and Manoj Reddy. 2019. “Identifying Peer Experts in Online Health Forums.” *BMC Medical Informatics and Decision Making* 19(3):41–49.
- Wright, Benjamin D. 1985. Additivity in Psychological Measurement. *Measurement and Personality Assessment* 101–12.
- Yates, Andrew, Arman Cohan, and Nazli Goharian. 2017. [Depression and Self-Harm Risk Assessment in Online Forums](#). *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings* 2968–78.
- Yoo, Minjoo, Sangwon Lee, and Taehyun Ha. 2019. Semantic Network Analysis for Understanding User Experiences of Bipolar and Depressive Disorders on Reddit. *Information Processing and Management* 56(4):1565–75.
- Zhang, Thomas, Jason H. D. Cho, and Chengxiang Zhai. 2015. [Understanding User Intents in Online Health Forums](#). *IEEE Journal of Biomedical and Health Informatics* 19(4):1392–98.
- Zirikly, Ayah, Philip Resnik, “ Ozlem Uzuner, and Kristy Hollingshead. 2019. CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 24–33.