



Pre- and Post-Depressive Detection using Deep Learning and Textual-based Features

Wei-Long Tey¹, Hui-Ngo Goh^{2*}, Amy Hui-Lan Lim³, Cheng-Kar Phang⁴

^{1,2,3}*Faculty of Computing and Informatics, Multimedia University, Cyberjaya, 63100, Malaysia*

⁴*Behavioral Health Centre, Sunway Medical Centre, Bandar Sunway, 47500, Malaysia*

Abstract. The rise of mental disorders, specifically depression, has shown an upward trend, especially during the COVID-19 pandemic. Previous studies have suggested that it is feasible to learn the textual and behavioral features of a user to identify depression on social media. This paper highlights three new contributions, which are, firstly, to introduce Patient Health Questionnaire-9 (PHQ-9) as the survey-based method to complement the TWINT API to collect Twitter data. Secondly, it is to propose a Bidirectional Encoder Representations from Transformers (BERT)-based model, along with emoji decoding and PHQ-9-based lexicon features for predicting the likelihood that a user will exhibit depressive symptoms. The results are promising, achieving an F1 score of 0.98 on a baseline dataset and an F1 score of 0.90 on a benchmark dataset, outperforming previous researcher's work of achieving a F1 score of 0.85 using solely textual features. Thirdly, previous researcher's work focuses on differentiating between depressed and non-depressed users only, while this paper further separates the users in the depressive class into before (pre-) and after (post-) self-reported diagnosis, which can potentially be used to detect early symptoms of depression. It was found that the top TF-IDF scores of the post-depressive class contain more frequently negatively implied words compared to the pre-depressive class.

Keywords: Deep learning; Depressive symptoms; PHQ-9; Predictive modelling; Sentiment analysis

1. Introduction

Researchers have highlighted that the cases of mental health disorders have increased during the COVID-19 pandemic (Fofana *et al.*, 2020; Berawi, 2020). However, the continuous improvement in research on depression has helped to increase mental health literacy and change the public's perception of the once highly stigmatized illness. People with symptoms related to depression are more willing to seek professional help when needed (Schomerus *et al.*, 2012).

Existing research on depression seeks to understand the relationship between depression and lifestyles such as smoking (Park and Romer, 2007), neurobiological markers in patients with depressive symptoms (Fu *et al.*, 2008), and gender differences among people with depressive symptoms (Piccinelli and Wilkinson, 2000).

The advancement of big data analytics in healthcare has motivated many researchers to gather insights that can contribute to mental health literacy (Murdoch and Detsky, 2013). The study by Lin *et al.* (2016) shows that there is a positive linear relationship between

*Corresponding author's email: hngoh@mmu.edu.my, Tel.: +603-83125223; Fax: +603-83125264
doi: [10.14716/ijtech.v14i6.6648](https://doi.org/10.14716/ijtech.v14i6.6648)

depression and social media usage. The study suggests that as depression causes lower self-esteem, it encourages one to turn to social media for validation instead. Another interesting perspective is that people might get depression due to the feeling of guilt for spending too much time on social media rather than doing meaningful things, which leads to the linear trend shown in the study. Other related research on identifying predictive markers of depression has been done on various social media platforms due to user-generated content (Surjandari *et al.*, 2019), including Instagram (Reece and Danforth, 2017), Facebook (De-Choudhury *et al.*, 2014) and Twitter (Tsugawa *et al.*, 2015).

In this paper, the terms predicting “depressive symptoms” and “predicting depression” will be used interchangeably. However, one must note that a machine learning model will not give a formal diagnosis, nor can the prediction replace a physical visit to a certified medical professional. This paper covers research motivation, contributions, related studies, the proposed framework, results, discussions, and a conclusion. Statements, declarations, and references can be found at the end of the paper.

2. Research Motivation and Contribution

The research motivations and contributions are described here. Firstly, for the clinical survey-based method, existing work relies on the Center for Epidemiologic Studies Depression Scale (CES-D) (Reece and Danforth, 2017; Tsugawa *et al.*, 2015; De-Choudhury *et al.*, 2014) rather than PHQ-9. Work by Milette *et al.* (2010) mentions that PHQ-9 produces similar reliability as compared to CES-D, albeit having only half the length of the CES-D. Therefore, in our study, we propose the use of PHQ-9 as the survey-based method to collect data other than the scraped method.

Secondly, we propose the use of a BERT-based model, along with emoji decoding and PHQ-9-based lexicon features, as the analytics method to predict the likelihood that a user exhibits depressive symptoms via Twitter. The lexicon features are symptom-wise terms that are derived from collected PHQ-9 survey feedback during the data collection.

Thirdly, we have further converted the dataset from a two-class dataset to a three-class dataset where the users in the depressive class are further divided into before (pre-) and after (post-) self-reported diagnosis.

3. Related Studies

3.1. Data Collection Methods

We can collect tweets using scraping tools (Poldi, 2019; Shrestha, 2018), clinical surveys (Guntuku *et al.*, 2017), or just use the publicly available dataset. Saxena (2018) uses a symptom-wise lexicon as a list of keywords for extracting depressive-indicative tweets. However, the likelihood of false positives tends to be high. Another method is to use a phrase for Twitter API, such as “I’m/I was/I am/I’ve been diagnosed with depression” anywhere in their tweets (Shen *et al.*, 2017; Coppersmith *et al.*, 2015a). The way to construct the non-depressed dataset is simply to scrape data that does not contain the depression-indicative term at all in their tweets.

Analysis by Smarr and Keefar (2011) reveals that the PHQ-9, CES-D, Beck’s Depression Inventory-II (BDI-II), Hospital Anxiety and Depression Scale (HADS), and Geriatric Depression Scale (GDS) are good enough for measuring depressive symptoms of an adult. Although it can help to reach a more reliable audience, the number of responses collected is limited, and utilizing crowdsourcing platforms might incur additional costs to compensate participants. The Computational Linguistics and Clinical Psychology (CLPsych)

is an example of a publicly available dataset consisting of tweets related to depression that is collected using Twitter API (Coppersmith *et al.*, 2015b).

3.2. Word Embedding Methods

Existing works (Saxena, 2018; Pedersen, 2015; De-Choudhury *et al.*, 2014; Schwartz *et al.*, 2014) use an n-gram approach to understanding patterns between words. Coppersmith *et al.* (2015a) use a single-word language model, in addition to using a character language model (CLM) in classifying mental illnesses in capturing semantics. It trains the three million tweets using Word2Vec and uses PHQ-9 to determine if a tweet is depression-indicative. Another popular word embedding method would be BERT, which it uses a bi-directional Long Short-Term Memory (LSTM) to generate word embeddings based on the contexts before and after each word. BERT uses masks. In each input, 15% of the words are masked (hidden), and the model is trained to predict the missing word(s).

In the work by De-Choudhury *et al.* (2014), a lexicon is built directly by scraping keywords from the “mental health” category from Yahoo! Answers and Wikipedia. The authors report a significantly higher usage of such lexicons when compared to non-depressed users (89% higher, $p < 0.0001$).

Other than textual features, Tsugawa *et al.* (2015) select features like the number of followers, number of followees, overall mention rate, tweet frequency, number of words, retweet rate and so on. Perhaps due to geographical differences, the posting frequency, number of followers, and number of followees are significantly different between depressives and non-depressives (De-Choudhury *et al.*, 2014). However, the retweet rate and ratio of tweets containing URL appear to be significant in differentiating both groups of users.

3.3. Machine Learning Techniques to Predict Depressive Symptoms

Alharahsheh and Abdullah (2021) have applied a few machine learning techniques with hyperparameter tuning, namely Support Vector Machine (SVM), Logistic Regression, Decision Tree (DT), Random Forest (RF), and ensemble methods on tweets that are collected from a survey conducted by Busara Center in Kenya. The results reveal that RF, Ada Boosting, and Voting-Ensemble models with the highest F1 score (0.78) and accuracy (85%) are better techniques that can be used to predict users with depressive symptoms.

Bhargava (2021) combines two data sources, namely the Sentiment140 dataset and depressive tweets, that are collected using the TWINT API. The author has compared a Convolutional Neural Network (CNN) and a hybrid of CNN with LSTM. The results reveal that hybrid CNN-LSTM performs better than CNN. There are also works that reveal that DT provides the highest accuracy and lesser completion time compared to other chosen techniques (Tiwari *et al.*, 2021; AlSagri and Ykhlef, 2020).

Chen and Sokolova (2021) focus on identifying depression posts from Reddit data, specifically text posts from ‘r/depression’. The performance of NB, SVM, XLNet, and BERT are compared, and it shows that BERT achieves the highest accuracy of 72%. Dinkel, Wu, and Yu (2019) propose a text-based multi-task BGRU network to detect depression from text transcripts consisting of clinical interviews to support the treatment for mental illness. The high F1 score of 0.84 indicates the viability of using a learning approach to detect depression.

Apriliani and Maharani (2023) have crawled the tweets belonging to 159 respondents to the Depression Anxiety Stress Scales (DASS-42) questionnaire. The scores from DASS-42 are calculated and used to label the respondents’ tweets. XL-Net and its hyper-parameter tuning effect on the tweets are analyzed. The average accuracy presented is 93.33%. Nurfadhila and Girsang (2023) collected 1424 tweets from Indonesia in the Indonesian

language between August and September 2021. Multinomial NB and SVM are two traditional machine learning techniques that are selected to be compared with CNN. The results reveal that both Multinomial NB and SVM are comparatively good choices of traditional machine learning algorithms but lose out to CNN, which produces the highest accuracy of 91.23%.

Work done by [Vasha et al. \(2023\)](#) focuses on analysing 10000 posts and comments from Facebook and YouTube. Six machine learning algorithms are chosen to be compared, namely RF, Logistic Regression (LR), DT, SVM, K Nearest Neighbour (KNN), and multinomial NB. Based on the precision rate and F1 score, SVM is identified to be the most suitable machine learning algorithm to build the best predictive model.

4. Proposed Framework

4.1. Baseline (Two-Class) Data Collection

A hybrid method of survey and self-scraped data is proposed for data collection to build the two-class dataset, which we label as a baseline dataset. PHQ-9 surveys are distributed among a crowd via a crowdsourcing platform. The participants are told that they must fulfill the following criteria:

- Twitter account must be active.
- They claim to have / have not suffered from depression (depending on samples).
- They are willing to share their data for this research.
- They have not participated in this research before.
- Their account should contain only English tweets.

For scraping public tweets, users are labeled as a depressive class if they follow the strict pattern of "(I'm/ I am/ I was/ I've been/ I have been) diagnosed (with) (clinical/severe) depression." For the non-depressive class, we adopted the negatively labeled dataset from [Shen et al. \(2017\)](#) and re-scraped 250 randomly selected users' tweets using TWINT API to get their latest status. Table 1 shows the landscape of the baseline dataset. It contains only tweets and its class (which is the label).

Table 1 Landscape of the Baseline Dataset

Method	Class	Total Users	Total Tweets
Survey (PHQ-9)	Depressive	15	>160000
Scraped	Depressive	235	>2.5 million
Scraped (from benchmark)	Non-depressive	250	>1 million

4.2. Data Pre-processing

4.2.1. Data pre-processing on baseline dataset

The baseline dataset is manually reviewed to remove accounts that are irrelevant to research, and Twitter accounts with fewer than 5 tweets. Participants with a PHQ-9 score of < 10 will be labeled as non-depressive, and those with ≥ 10 will be labeled otherwise. For the depressive class, the user ID of each qualified user will be used to scrape all tweets of the user and merge them with the scraped data of the depressive class. All the collected tweets undergo the following data pre-processing steps of emoji decoding and spellchecking.

4.2.2. Transforming baseline dataset to three-class dataset

The baseline dataset from users in the depressive class is further separated into before (pre-) and after (post-) self-reported diagnosis, namely pre-depressive and post-

depressive, by scanning the tweets for the earliest self-reported diagnosis that fulfills the keywords that are used for scraping. No changes are made to the non-depressive data.

4.3. Exploratory Data Analysis

In this step, analysis such as lexicon analysis, TF-IDF, syntactical structure, and n-grams analysis are performed to find patterns among users within the baseline dataset in the hope of finding features to build the model for prediction.

4.4. Neural Modeling

BERT is chosen as a predictive modeling technique. The rationale for choosing BERT stems from works on predicting depression on Reddit (Chen and Sokolova, 2021) and on an interview text transcript (Dinkel, Wu, and Yu, 2019). Both works produce remarkable results from using BERT, and this research further shows that BERT is outstanding for this Natural Language Processing (NLP) task.

Once the BERT architecture has been set up, an ablation study is conducted where BERT, BERT with emoji decoding, and BERT with emoji decoding, and the lexicon frequencies are passed as an additional input to the decoder for the BERT architecture are applied to three types of datasets which are the baseline dataset, three-class dataset and a benchmark dataset that is available from another research (Shen et al., 2017).

5. Results and Discussions

Exploratory data analysis and neural model have been carried out to examine how the writing styles can be used in identifying depressive symptoms.

5.1. PHQ-9 Lexicon Analysis

Based on the ten questions in PHQ-9, we extracted words indicating depressive symptoms, such as 'hopeless,' 'tired,' and 'failure.' These seed words were then augmented by identifying their synonyms through Thesaurus.com. In total, we constructed a lexicon comprising 86 words.

Each matched lexicon from a tweet is called a "hit". It is observable that the depressive-labeled users include depressive terms in 0.6% of the tweets, whereas non-depressive labeled users include them in only 0.4%. The results of the analysis suggest that using a lexicon constructed from PHQ-9 may be an important feature in detecting depression.

5.2. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF scores show the importance of the words within documents. All tweets are parsed through Porter Stemmer prior to the TF-IDF measurement. Stemming is a process of reducing a word's linguistics morphology into its root form. Table 2 illustrates the list of words with the highest TF-IDF scores in the respective group for the two-class dataset.

Table 2 Words with highest TF-IDF scores for two-class dataset

Group	Words
Depressive	keep, bad, even, touch, thank, sever, wrong, mostli, hate, free, beauti, mood, blue, style, jesu, poor, tire, hair, cours, choic, amen, straight, stuck, option, chicago, ugh, fast, aww, yup, omg, gross, curiousca
Non-depressive	Manag, update, said, good, mr, date, singl, true, everybodi, mouth, light, help, sorri, china, chill favorit, present, absolut, okay, forgot

Non-depressive group does not display clear and significant cliques. However, it is obvious in the depressive group such as words indicating negative emotions, such as "bad", "severe", "wrong", "hate", "mood", "blue", "poor", "tire", "stuck", "gross", and religion indicative terms such as "jesus" and "amen".

Table 3 illustrates the list of words with the highest TF-IDF scores in depressive groups for three-class dataset. The list of words with the highest TF-IDF scores in non-depressive groups is the same as in Table 2.

Table 3 Words with highest TF-IDF scores for three-class dataset

Group	Words
Pre-depressive	really, petxpe, fc3gt, 15, ku, stupid, ask, sweet, run, girls, moofmurphy, christmas, gender, eye, choice, gun, pizza, status, must
Post-depressive	Christmas, curiouscat, reading, hell, petxpe, scared, cis, lilyluchesi, evil, voice, sleep, lost, deserve, series, lots, coming, bill, fan, rock, type

It is observable that significantly fewer negatively implied words are used in the pre-depressive group as compared to the post-depressive group. Depressed patients are more likely to have frequent melancholic mode, explaining why they include more depressed terms in their tweets.

5.3. Syntactical Structure

All tweets are pre-processed (lowercase normalization, emoji replacement, and spelling corrections) prior to performing the part-of-speech (POS) tagging. Table 4 summarizes the most used words in the mentioned THREE tags of NOUN, VERB, and ADJECTIVE in respective groups.

Table 4 Top usage of POS tag by group

POS Tag	Group	Words
NOUN	Depressive	time, day, life, today, someone, depression, way, lt, thing, everyone, something, lol, year, video, night
	Non-Depressive	time, video, gt, day, year, life, man, lol, way, today, thing, lt, shit, pa, love, something, girl
VERB	Depressive	be, get, do, have, go, know, see, make, love, take, follow, help, feel, say, want, give, let, think, tell
	Non-Depressive	<i>be, get, do, have, go, know, make, see, take, let, say, love, give, tell, think, stop, want, keep, feel</i>
ADJECTIVE	Depressive	good, new, much, u, i, happy, last, bad, other, great, same, first, many, little, real, old, mental
	Non-Depressive	good, i, u, much, happy, last, other, same, bad, first, real, great, many, ur, little, next, sure, old, own

The most significant difference is the NOUN tag. The depressive group uses more arbitrary references and determiners like “someone”, “everyone,” or “something”. Possibly because users with depressive symptoms tend to report poorer concentration and memory and, therefore, cannot recall the subjects precisely (Zuckerman et al., 2018).

As for the VERB and ADJECTIVE tags, depressives tend to express themselves more. A significant difference is observed in the words “help” and “mental” as patients with major depression are more likely to open their self on the Internet (Ybarra, Alexander, and Mitchell, 2005).

5.4. N-grams

We have experimented using different N values i.e.: $N = \{1, 2, 3\}$, and bigram with $N = 2$ shows a significant difference between the two classes. Table 5 shows the most used bigram.

Table 5 Bigram distribution

Depressive Group		Non-depressive group	
Bigram	Normalized Frequency	Bigram	Normalized Frequency
I wa	0.12	I want	0.09
I think	0.10	I love	0.08
I love	0.08	I wa	0.07
I know	0.07	Werewolf germani	0.06

Users in the depressive group tend to use words that relate to their feelings, like “I think”, whereas users in the non-depressive group tend to use words that relate to expressing their opinions, like “I want”. This gives an indicator to further investigate feelings-related words/phrases for the depressive group.

5.5. Neural Modelling Results

As described in Section 4.4, an ablation study is conducted, and the performance of each architecture model on datasets is studied, as shown in Table 6.

Table 6 Results on different architecture models on baseline (Two-class) and Three-class datasets

No	Dataset	Architecture	F1 score	Accuracy
1	Baseline (Two-Class dataset)	BERT	0.98	0.99
2	Baseline (Two-Class dataset)	BERT + emoji	0.99	0.99
3	Baseline (Two-Class dataset)	BERT + emoji + lexicon	0.98	0.99
4	3-class dataset	BERT	0.73	0.72
5	3-class dataset	BERT + emoji	0.75	0.76
6	3-class dataset	BERT + emoji + lexicon	0.76	0.77

Having additional lexicon features has not recorded an improvement in F1 score for the baseline dataset. However, it improves when applied to a three-class dataset. Overall, this proves that the ability of a model to capture detailed data representation allows the model to learn and thus perform better.

To verify the effectiveness of the architecture models, we have also applied the architecture models to a dataset that is used in work by [Shen et al. \(2017\)](#), which we labeled as a benchmark, which is a two-class dataset. We adopt and customize the benchmark dataset to suit our ablation, as in Table 7.

Table 7 Results on different architectures

No	Dataset	Architecture	F1 score	Accuracy
1	Benchmark	BERT	0.87	0.84
2	Benchmark	BERT + emoji	0.88	0.84
3	Benchmark	BERT + emoji + lexicon	0.90	0.87

BERT, with emoji decoding and lexicon features, has proven to be better, with an F1 score of 0.90 and an accuracy score of 0.87. After experimenting with various token lengths, we found that a length of 256 tokens works best in this case.

In general, BERT-based architecture models have reported high F1 scores when these models were applied to baseline, three-class, and benchmark datasets. By pre-training BERT on masked language modeling and next sentence prediction, BERT has a much deeper understanding of the context of each token, which makes it more powerful in recognizing the linguistic features of tweets when compared to older embedding methods like Word2Vec or GloVe.

6. Conclusions

In this paper, a method to predict users with depressive symptoms on Twitter that use

a BERT-based model with emoji decoding along with our lexicon-based handcrafted feature is proposed. We have demonstrated that by using only textual features, we can achieve outstanding results and outperform a model that is built using linguistic features. We also briefly describe a plan to convert a two-class dataset to a three-class dataset. Our study shows that it is indeed possible to distinguish between “pre-depressive” and “post-depressive” groups. However, the process of finding the differences is significantly harder because we rely on self-reported diagnosis. This means a patient might have already been diagnosed before or after they posted the tweets. Future researchers should take note of the methodologies used, especially when they collect data, to ensure that the cut-off line is more accurate. As for application, the trained model can be integrated as part of social media platforms/tools whereby the model can analyze user inputs and provide an indication of potential depression based on learned patterns and characteristics. Future work may include non-textual features in this deep learning model for possibly better performance.

References

- Alharahsheh, Y.E., Abdullah, M.A., 2021. Predicting Individuals Mental Health Status in Kenya using Machine Learning Methods. *In: 2021 12th International Conference on Information and Communication Systems (ICICS)*, Institute of Electrical and Electronics Engineers (IEEE), pp. 94–98
- AlSagri, H.S., Ykhlef, M., 2020. Machine Learning-based Approach for Depression Detection in Twitter using Content and Activity Features. *Institute of Electronics, Information and Communication Engineers (IEICE) Transactions on Information and Systems*, Volume 103(8), pp. 1825–1832
- Apriliani, F., Maharani, W., 2023. Depression Detection on Social Media Twitter using XLNet Method. *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika (Scientific Journal of Informatics Research and Learning)*, Volume 8(1), pp. 172–180
- Berawi, M.A., 2020. Empowering Healthcare, Economic, and Social Resilience During Global Pandemic COVID-19. *International Journal of Technology*, Volume 11(3), pp. 436–439
- Bhargava, C., 2021. Depression Detection using Sentiment Analysis of Tweets. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, Volume 12(11), pp. 5411–5418
- Chen, Z., Sokolova, M., 2021. Sentiment Analysis of the COVID-related r/Depression Posts. *arXiv preprint arXiv:2108.06215*
- Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., 2015. From Attention Deficit Hyperactivity Disorder (ADHD) to Seasonal Affective Disorder (SAD): Analyzing the Language of Mental Health on Twitter Through Self-reported Diagnoses. *In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 1–10
- Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., Mitchell, M., 2015. Computational Linguistics and Clinical Psychology (CLPsych) 2015 Shared Task: Depression and Post Traumatic Stress Disorder (PTSD) on Twitter. *In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 31–39
- De-Choudhury, M., Counts, S., Horvitz, E.J., Hoff, A., 2014. Characterizing and predicting postpartum depression from shared facebook data. *In: Proceedings of the 17th Association for Computing Machinery (ACM) conference on Computer supported cooperative work & social computing*, pp. 626–638

- Dinkel, H., Wu, M., Yu, K., 2019. Text-based Depression Detection on Sparse Data. *arXiv preprint arXiv:1904.05154*
- Fofana, N.K., Latif, F., Sarfraz, S., Bashir, M.F., Komal, B., 2020. Fear and Agony of the Pandemic Leading to Stress and Mental Illness: An Emerging Crisis in the Novel Coronavirus (COVID-19) Outbreak. *Psychiatry Research*, Volume 291, p. 113230
- Fu, C.H., Mourao-Miranda, J., Costafreda, S.G., Khanna, A., Marquand, A.F., Williams, S.C., Brammer, M.J., 2008. Pattern Classification of Sad Facial Processing: Toward the Development of Neurobiological Markers in Depression. *Biological psychiatry*, Volume 63(7), pp. 656–662
- Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H., Eichstaedt, J.C., 2017. Detecting Depression and Mental Illness on Social Media: an integrative review. *Current Opinion in Behavioral Sciences*, Volume 18, pp. 43–49
- Lin, L.Y., Sidani, J.E., Shensa, A., Radovic, A., Miller, E., Colditz, J.B., Hoffman, B.L., Giles, L.M., Primack, B.A., 2016. Association Between Social Media use and Depression Among US Young Adults. *Depression and anxiety*, Volume 33(4), pp. 323–331
- Milette, K., Hudson, M., Baron, M., Thombs, B.D., Canadian Scleroderma Research Group, 2010. Comparison of the Patient Health Questionnaire Depression Scale (PHQ-9) and Center for Epidemiologic Studies Depression Scale (CES-D) Depression Scales in Systemic Sclerosis: Internal Consistency Reliability, Convergent Validity and Clinical Correlates. *Rheumatology*, Volume 49(4), pp. 789–796
- Murdoch, T.B., Detsky, A.S., 2013. The Inevitable Application of Big Data to Health Care. *Jama*, Volume 309(13), pp. 1351–1352
- Nurfadhila, B., Girsang, A.S., 2023. Identifying Indication of Depression of Twitter User in Indonesia Using Text Mining. *International Journal of Intelligent Systems and Applications in Engineering*, Volume 11(2), pp. 523–530
- Park, S., Romer, D., 2007. Associations Between Smoking and Depression in Adolescence: an Integrative Review. *Journal of Korean Academy of Nursing*, Volume 37(2), pp. 227–241
- Pedersen, T., 2015. Screening Twitter users for Depression and Post Traumatic Stress Disorder (PTSD) with Lexical Decision Lists. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pp. 46–53
- Piccinelli, M., Wilkinson, G., 2000. Gender Differences in Depression: Critical Review. *The British Journal of Psychiatry*, Volume 177(6), pp. 486–492
- Poldi, F., 2019, Twint—Twitter Intelligence Tool. Available online at <https://github.com/twintproject/twint/wiki>, Accessed on May 21, 2020
- Reece, A.G., Danforth, C.M., 2017. Instagram Photos Reveal Predictive Markers of Depression. *European Physical Journal (EPJ) Data Science*, Volume 6(1), p. 15
- Saxena, A., 2018. A Semantically Enhanced Approach to Identify Depression-Indicative Symptoms Using Twitter Data.
- Schomerus, G., Schwahn, C., Holzinger, A., Corrigan, P.W., Grabe, H.J., Carta, M.G., Angermeyer, M.C., 2012. Evolution of Public Attitudes About Mental Illness: A Systematic Review and Met - Analysis. *Acta Psychiatrica Scandinavica*, Volume 125(6), pp. 440–452
- Schwartz, H.A., Eichstaedt, J., Kern, M., Park, G., Sap, M., Stillwell, D., Kosinski, M., Ungar, L., 2014. Towards Assessing Changes in Degree of Depression Through Facebook. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pp. 118–125

- Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.S., Zhu, W., 2017. Depression Detection Via Harvesting Social Media: A Multimodal Dictionary Learning Solution. *In: International Joint Conferences on Artificial Intelligence (IJCAI)*, pp. 3838–3844
- Shrestha, K., 2018. Machine Learning for Depression Diagnosis using Twitter Data. *International Journal of Computer Engineering in Research Trends*, 5(2).
- Smarr, K.L., Keefer, A.L., 2011. Measures of Depression and Depressive Symptoms: Beck Depression Inventory - II (BDI - II), Center for Epidemiologic Studies Depression Scale (CES - D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire - 9 (PHQ - 9). *Arthritis Care & Research*, Volume 63(S11), pp. S454–S466
- Surjandari, I., Wayasti, R.A., Laoh, E., Rus, A.M.M., Prawiradinata, I., 2019. Mining Public Opinion on Ride-hailing Service Providers using Aspect-based Sentiment Analysis. *International of Journal Technology*, Volume 10, pp. 818–828
- Tiwari, P.K., Sharma, M., Garg, P., Jain, T., Verma, V.K., Hussain, A., 2021. A Study on Sentiment Analysis of Mental Illness using Machine Learning Techniques. *In: International Operating Procedure (IOP) Conference Series: Materials Science and Engineering*, Volume 1099(1), p. 012043
- Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., Ohsaki, H., 2015. Recognizing Depression from Twitter Activity. *In: Proceedings of the 33rd Annual Association for Computing Machinery (ACM) Conference on Human Factors in Computing Systems*, pp. 3187–3196
- Vasha, Z.N., Sharma, B., Esha, I.J., Al Nahian, J., Polin, J.A., 2023. Depression Detection in Social Media Comments Data using Machine Learning Algorithms. *Bulletin of Electrical Engineering and Informatics*, Volume 12(2), pp. 987–996
- Ybarra, M.L., Alexander, C., Mitchell, K.J., 2005. Depressive Symptomatology, Youth Internet use, and Online Interactions: A National Survey. *Journal of Adolescent Health*, Volume 36(1), pp. 9–18
- Zuckerman, H., Pan, Z., Park, C., Brietzke, E., Musial, N., Shariq, A.S., Iacobucci, M., Yim, S.J., Lui, L.M., Rong, C., McIntyre, R.S., 2018. Recognition and Treatment of Cognitive Dysfunction in Major Depressive Disorder. *Frontiers in Psychiatry*, Volume 9, p. 655