IDIAP Submission@LT-EDI-ACL2022: Detecting Signs of Depression from Social Media Text

Muskaan Singh, Petr Motlicek
IDIAP Research Institute,
Martigny, Swizerland
(msingh,petr.motlicek)@idiap.ch

Abstract

Depression is a common illness involving sadness and lack of interest in all day-to-day activities. It is important to detect depression at an early stage as it is treated at an early stage to avoid consequences. In this paper, we present our system submission of ARGUABLY for DepSign-LT-EDI@ACL-2022. We aim to detect the signs of depression of a person from their social media postings wherein people share their feelings and emotions. The proposed system is an ensembled voting model with fine-tuned BERT, RoBERTa, and XLNet. Given social media postings in English, the submitted system classify the signs of depression into three labels, namely "not depressed," "moderately depressed," and "severely depressed." Our best model is ranked 3^{rd} position with 0.54% accuracy . We make our codebase accessible here[1].

1 Introduction

Depression is a common mental illness that involves sadness and lack of interest in all day-to-day activities[2]. Detecting depression is essential as it has to be observed and treated at an early stage to avoid severe consequences (Evans-Lacko et al., 2018; Losada et al., 2017). Depression implies mental disorder which may cause disability (Organization et al., 2012; Whiteford et al., 2015; Vigo et al., 2016), very few people are able to receive treatment (Wang et al., 2007). It is far more difficult for the people with low socioeconomic status or people living in low economic conditions (Steele et al., 2007; Ormel et al., 2008) , even adjusting for disorder severity (Mojtabai and Olfson, 2010; Andrade et al., 2014) . Consequently, there is a need to detect these signs of depression early in time to avoid further repercussions. In this work, we detect the signs of depression, namely in *"not depressed," "moderately depressed," and "severely depressed"* from person's social media postings where people share their feelings and emotions.

There are dataset available for detecting depression task from social media platform such as Twitter (Leis et al., 2019; Arora and Arora, 2019; Yazdavar et al., 2020; de Jesús Titla-Tlatelpa et al., 2021; Chiong et al., 2021; Safa et al., 2021), Reddit (de Jesús Titla-Tlatelpa et al., 2021; Ríssola et al., 2019; Tadesse et al., 2019; Burdisso et al., 2019; Martínez-Castaño et al., 2020), Facebook (Chiong et al., 2021; Wongkoblap et al., 2019; Wu et al., 2020; Yang et al., 2020), Instagram (Mann et al., 2020; Ricard et al., 2018), Weibo (Li et al., 2018; Yu et al., 2021) and NHANES, K-NHANES (Oh et al., 2019). The linguistic feature extraction methods used for detecting depression signs on social media such as Word embedding (Mandelbaum and Shalev, 2016), N-grams (Cavnar et al., 1994), Tokenization (Webster and Kit, 1992), Bag of words (Zhang et al., 2010; Aho and Ullman, 1972), Stemming (Jivani et al., 2011), Emotion analysis (Leis et al., 2019; Shen et al., 2017; Chen et al., 2018), Part-of-Speech (POS) tagging (Chiong et al., 2021; Wu et al., 2020), Behavior features(Wu et al., 2020) and Sentiment polarity (Leis et al., 2019; Ríssola et al., 2019).

2 Related Work

There have been several attempts to use machine learning algorithms as SVM (Ríssola et al., 2019; Arora and Arora, 2019; Burdisso et al., 2019; Yang et al., 2020), Logistic regression (Ríssola et al., 2019; Chen et al., 2018; Tadesse et al., 2019; Yang et al., 2019), Neural networks (Wu et al., 2020; Liu et al., 2019), Random forests (Yang et al., 2020; Chiong et al., 2021), Bayesian statistics (Yang et al., 2020; Chen et al., 2018), Decision trees (Yang et al.,

[1] https://github.com/Muskaan-Singh/Depression-Detection.git
[2] http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b

Label	Train	Dev	Total
Not depressed	3801	1830	5631
Moderately depressed	8325	2306	10631
severely depressed	1261	360	1621

Table 1: Data distribution for the DepSign-LT-EDI dataset.

2020; Chiong et al., 2021), K-Nearest Neighbor (Yang et al., 2020; Burdisso et al., 2019), Linear regression (Yu et al., 2021; Ricard et al., 2018), Ensemble classifiers (Leiva and Freire, 2017; Oh et al., 2019), Multilayer Perceptron (Chiong et al., 2021; Safa et al., 2021), Boosting (Tadesse et al., 2019), K-Means (Ma et al., 2017). (Wu et al., 2020), proposed a recurrent neural network for prediction of depression from content-based, behavioral and environmental data. Further, LSTM is used for post generation for each user from the social media dataset. The public dataset available were merged with this generated dataset and fed into a deep learning classifier. (Srimadhur and Lalitha, 2020), proposed an end-to-end CNN model for detection and assessment of depression levels using speech. (de Souza Filho et al., 2021), presents best performing ML models (Random forest, K-nearest neighbors, XG Boost) for detecting depressed patients from clinical and laboratory patients of sociodemographic.

3 Shared Task Description

The shared task urges to detect the signs of depression of a person from the social media post where people share their feeling and emotions. Its aim is to detect speech for Equality, Diversity, and Inclusion (DepSign-LT-EDI@ACL-2022)(**??**). The goal was to classify the sign of depression into three labels, namely, *"not depressed," "moderately depressed", and "severely depressed"* for a given social media posting. The dataset (Sampath et al., 2022) contains 8891, 4496, 3245 comments for training, development, and test set, respectively, annotated with three different labels for the English language. The detailed distribution of the dataset based on labels can be seen in Table 1, and some instances for not depressed, moderately depressed, and severely depressed are presented in Table 2. The organizers have provided a baseline code using state-of-art machine learning techniques along with the dataset.

Comment	Label
Happy New Years Everyone : We made it another year	not depressed
Sat in the dark and cried myself going into the new year. Great start to 2020 :	moderately depressed
Words can't describe how bad I feel right now : I just want to fall asleep forever.	severely depressed

Table 2: Examples for Not depressed, Moderately depressed and severely depressed DepSign EDI dataset.

4 Methodology

Firstly, we pre-process the social media tweets with the basic NLTK library (Loper and Bird, 2002) for stop words removal, emojis removal, and punctuation removal. Secondly, we extract the features by tokenizing all the sentences and mapping those tokens with the word IDs. For every sentence in the dataset, we follow a series of steps (i) tokenize the sentences (ii) prepend the [CLS] token to the start (iii) append the [SEP] token to the end (iv) map the token to their IDs (v) pad or truncate the sentenced to max length (vi) mapping of attention masks for [PAD] tokens. We padded and truncated the max_length=30. The generated sequence sentences are passed for encoding with its attention mask (simply differentiating padding from non-padding). Finally, we predicted the labels using ensembles voting model for BERT (Devlin et al., 2018), XLNET (Liu et al., 2019) and RoBERTa model(Liu et al., 2019). BERT is Bi-directional Encoder Representation from Transformers (BERT), involving pre-training Bi-directional transformers for language understanding from an unlabelled text by jointly conditioning left t- right context for all layers. Fine-tuning of a pre-trained BERT model can be easily done with just one additional output layer for developing a state-of-art model for a wide range of NLP tasks without substantial task-specific architecture modifications. Robustly Optimized BERT approach has emphasized data being used for pre-training and the number of passes for training. The BERT model is optimized with dynamic masking, more extended training with big batches over more data, removing the next prediction objective, and dynamically changing masking patterns for training data. The model achieved

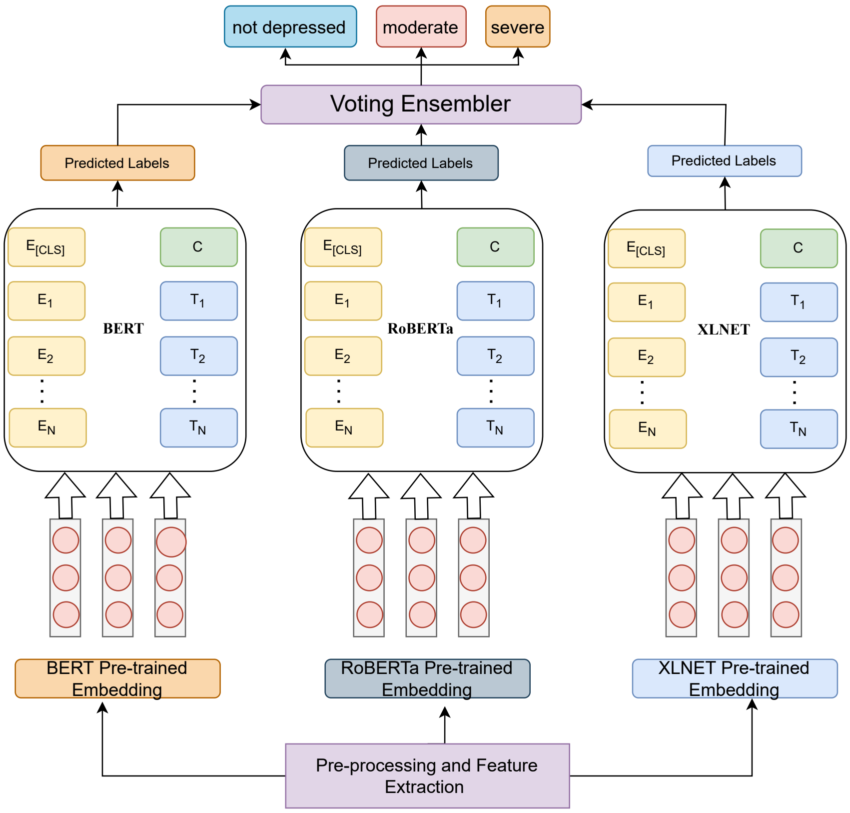

Figure 1: We predicted the labels using fine-tuned BERT, XLNET, and RoBERTa models, respectively, then we applied an ensemble voting classifier. Each model gives a label to the sentence, highest vote is chosen as the final label.

state-of-art results on GLUE, RACE, and SQuAD without multi-task finetuning for GLUE or additional data for SQuAD. ERNIE 2.0 is another continual pre-training framework that efficiently supports customized training tasks in multi-task learning incrementally. The pre-trained model is fine-tuned to adapt to various language understanding tasks. The framework has demonstrated significant improvement over BERT and XLNET on approximately 16 tasks, including GLUE. We take each label to the sentence and number of labels with the highest vote if chosen as the final label in ensemble voting(Dimitriadou et al., 2001).

4.0.1 Experimental Setup

We use V1 100 GPU with 53GB RAM alongside 8 CPU cores for the experimental setup. We divide the entire dataset into a 90:10 training and validation split of 8 batches, with a learning rate (1e-5) and Adam optimizer(Kingma and Ba, 2014) with epsilon (1e-8). We feed a seed_val of 42. For calculating the training loss over all the batches, we use gradient descents (Andrychowicz et al., 2016) with clipping the norm to 1.0 to avoid exploding gradient problem.

5 Results

We evaluate our model quantitatively and qualitatively for the DepSign-LT-EDI dataset. The classification report for our proposed model with average and best submission among all the teams is reported in Table 3. The proposed model has shown progressive results with the 3^{rd} position on the leaderboard https://competitions.codalab.org/competitions/36393#learn_the_details-result. Analysing our quantitative results, 0.53, 0.57, 0.54 are the reported precision, Recall, and F1-score, which is relatively 0.06, 0.07, 0.06 more than the average and 0.05, 0.02, 0.04 less for best-performing submission, respectively. Qualitative analysis of the predicted labels by the proposed methodology can be seen in Table 4. The first, third, and fifth comments were not depressed, moderately depressed, and severely depressed. They are correctly classified instances indicating our model has efficiently identified the phrases with a negative sentiment, such as "depressed," "anxious," "I

Table 3: Classification system's performance measured in terms of macro averaged Precision, macro averaged Recall and macro averaged F-Score across all the classes. Sklearn classification report was utilized to generate the reports by all the submission teams

	Accuracy	Recall	Precision	Weighted F1- score	Macro F1-score
Average of all teams	0.5988	0.5058	0.4782	0.6012	0.4821
Best of all teams	0.6709	0.5912	0.586	0.666	0.583
Our submission	**0.6253**	**0.572**	**0.5303**	**0.6333**	**0.5467**

Text_data	Label
Sometimes people can be either too oblivious or choose not to care and they may not intend to harm us but it does hurt : [removed]	not depressed
TMS : My doctor wants me to do TMS for my depression. Has anyone done TMS or is doing it? I was just want to know it is worth it.	not depressed
Depressed : I have nothing to look forward to, I wake up feeling so down and depressed , anxious about everything, I look at myself in the mirror and i feel and look so ugly , I shouldn't be allowed out in public being so disgusting looking...:(	moderate
Uncertain : I would like to die, but I'm scared of the repercussions. More specifically, I have to attend a birthday party and a gathering to say goodbye to a friend who will be moving in the next few days and I don't want to ruin their celebrations	moderate
my whole life has fallen apart : everyone hates me. all my friends hate me. my moms hates me and my dads too busy for me. i don't talk to my family. the only person i have is my boyfriend who will probably leave me soon because of how i am. i eat lunch in the bathroom. no one in my classes talks to me. i got my boyfriend and his friend accidentally suspended for an incident they jokingly started that ended in me almost getting beat up (they meant no harm). i cried all day and i had to leave school early. i can't eat. my head is pounding. there's no hope. there's no point in living and no one cares. everyone just hates me. and i'm not a bad or mean person i don't think, but now that's all i am to everyone. i want to end it, but if i fail i get readmitted to the psych ward and i promised myself if i ever went back there, i would kill myself. i don't know what to do anymore.	severe
Antidepressants : Do antidepressants help if your not depressed? I started taking them to get through a rough patch and they have helped me - does this mean I technically have depression because I read online that antidepressants don't help if your not depressed?	severe

Table 4: Qualitative Results for not depression, moderate, severe

would kill myself," and so on. Since the first comment barely had any negative phrases, the model classified it as not depressed. However, in the case of the second instance, the comment is labeled as not depressed when in reality, it is a case of severe depression. The probable reason for this misclassification is that the model cannot identify medical terms like "TMS," and overall, the second comment barely has any negative words or expressions.

The fourth instance is labeled as moderate; however, the person claims that they want to die; this indicates that this comment is instead a case of severe depression. The probable reason for this misclassification is that the model focuses more on the phrases like "party," "celebration," "die." rather than the entire sentences. Since this statement has a mix of positive and negative phrases, the model assumes it to be a moderate case. Lastly, the sixth instance is classified as severe; it seems like the case of mild depression.

6 Conclusion

In this paper we present our system paper submission for DepSign-LT-EDI@ACL-2022. We aim to detect the signs of depression of a person from their social media postings wherein people share their feelings and emotions. The proposed system is an ensembled voting model with fine-tuned BERT, RoBERTa, and XLNet. Given social media postings in English, the submitted system classify the signs of depression into three labels, namely "not depressed," "moderately depressed," and "severely depressed." Our best model is ranked 3^{rd} position with 0.54% accuracy . The system performs quite well to recognize the comments for depression comments; In the future, we intend to work on a multitask learning framework to handle all kinds of depression or illness and even the severity of depression. We also aim to detect multilingual depression speech in the code-mixing scenarios.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 833635 (project ROXANNE: Real-time network, text, and speaker analytics for combating organized crime, 2019-2022).

References

Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.

Laura Helena Andrade, J Alonso, Z Mneimneh, JE Wells, A Al-Hamzawi, G Borges, E Bromet, Ronny Bruffaerts, G De Girolamo, R De Graaf, et al. 2014. Barriers to mental health treatment: results from the who world mental health surveys. *Psychological medicine*, 44(6):1303–1317.

Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. 2016. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems*, 29.

Priyanka Arora and Parul Arora. 2019. Mining twitter data for depression detection. In *2019 International Conference on Signal Processing and Communication (ICSC)*, pages 186–189. IEEE.

Sergio G Burdisso, Marcelo Errecalde, and Manuel Montes-y Gómez. 2019. A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133:182–197.

William B Cavnar, John M Trenkle, et al. 1994. N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*, volume 161175. Citeseer.

Xuetong Chen, Martin D Sykora, Thomas W Jackson, and Suzanne Elayan. 2018. What about mood swings: Identifying depression on twitter with temporal measures of emotions. In *Companion Proceedings of the The Web Conference 2018*, pages 1653–1660.

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*, 135:104499.

José de Jesús Titla-Tlatelpa, Rosa María Ortega-Mendoza, Manuel Montes-y Gómez, and Luis Villaseñor-Pineda. 2021. A profile-based sentiment-aware approach for depression detection in social media. *EPJ Data Science*, 10(1):54.

Erito Marques de Souza Filho, Helena Cramer Veiga Rey, Rose Mary Frajtag, Daniela Matos Arrowsmith Cook, Lucas Nunes Dalbonio de Carvalho, Antonio Luiz Pinho Ribeiro, and Jorge Amaral. 2021. Can machine learning be useful as a screening tool for depression in primary care? *Journal of Psychiatric Research*, 132:1–6.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Evgenia Dimitriadou, Andreas Weingessel, and Kurt Hornik. 2001. Voting-merging: An ensemble method for clustering. In *International conference on artificial neural networks*, pages 217–224. Springer.

Sara Evans-Lacko, Sergio Aguilar-Gaxiola, Ali Al-Hamzawi, Jordi Alonso, Corina Benjet, Ronny Bruffaerts, WT Chiu, Silvia Florescu, Giovanni de Girolamo, Oye Gureje, et al. 2018. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the who world mental health (wmh) surveys. *Psychological medicine*, 48(9):1560–1571.

Anjali Ganesh Jivani et al. 2011. A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl*, 2(6):1930–1938.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

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

Victor Leiva and Ana Freire. 2017. Towards suicide prevention: early detection of depression on social media. In *International Conference on Internet Science*, pages 428–436. Springer.

Ang Li, Dongdong Jiao, and Tingshao Zhu. 2018. Detecting depression stigma on social media: A linguistic analysis. *Journal of affective disorders*, 232:358–362.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

David E Losada, Fabio Crestani, and Javier Parapar. 2017. erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 346–360. Springer.

Long Ma, Zhibo Wang, and Yanqing Zhang. 2017. Extracting depression symptoms from social networks and web blogs via text mining. In *International Symposium on Bioinformatics Research and Applications*, pages 325–330. Springer.

Amit Mandelbaum and Adi Shalev. 2016. Word embeddings and their use in sentence classification tasks. *arXiv preprint arXiv:1610.08229*.

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

Rodrigo Martínez-Castaño, Juan C Pichel, and David E Losada. 2020. A big data platform for real time analysis of signs of depression in social media. *International Journal of Environmental Research and Public Health*, 17(13):4752.

Ramin Mojtabai and Mark Olfson. 2010. National trends in psychotropic medication polypharmacy in office-based psychiatry. *Archives of General Psychiatry*, 67(1):26–36.

Jihoon Oh, Kyongsik Yun, Uri Maoz, Tae-Suk Kim, and Jeong-Ho Chae. 2019. Identifying depression in the national health and nutrition examination survey data using a deep learning algorithm. *Journal of affective disorders*, 257:623–631.

World Health Organization et al. 2012. Good health adds life to years: Global brief for world health day 2012. Technical report, World Health Organization.

Johan Ormel, Maria Petukhova, Somnath Chatterji, Sergio Aguilar-Gaxiola, Jordi Alonso, Matthias C Angermeyer, Evelyn J Bromet, Huibert Burger, Koen Demyttenaere, Giovanni De Girolamo, et al. 2008. Disability and treatment of specific mental and physical disorders across the world. *The British Journal of Psychiatry*, 192(5):368–375.

Benjamin J Ricard, Lisa A Marsch, Benjamin Crosier, and Saeed Hassanpour. 2018. Exploring the utility of community-generated social media content for detecting depression: an analytical study on instagram. *Journal of medical Internet research*, 20(12):e11817.

Esteban A Ríssola, Seyed Ali Bahrainian, and Fabio Crestani. 2019. Anticipating depression based on online social media behaviour. In *International Conference on Flexible Query Answering Systems*, pages 278–290. Springer.

Ramin Safa, Peyman Bayat, and Leila Moghtader. 2021. Automatic detection of depression symptoms in twitter using multimodal analysis. *The Journal of Supercomputing*, pages 1–36.

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 2022. Findings of the shared task on detecting signs of depression from social media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

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*, pages 3838–3844.

NS Srimadhur and S Lalitha. 2020. An end-to-end model for detection and assessment of depression levels using speech. *Procedia Computer Science*, 171:12–21.

Leah S Steele, Carolyn S Dewa, Elizabeth Lin, and Kenneth LK Lee. 2007. Education level, income level and mental health services use in canada: Associations and policy implications. *Healthcare Policy*, 3(1):96.

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

Daniel Vigo, Graham Thornicroft, and Rifat Atun. 2016. Estimating the true global burden of mental illness. *The Lancet Psychiatry*, 3(2):171–178.

Jue Wang, Maneesh Agrawala, and Michael F Cohen. 2007. Soft scissors: an interactive tool for realtime high quality matting. In *ACM SIGGRAPH 2007 papers*, pages 9–es.

Jonathan J Webster and Chunyu Kit. 1992. Tokenization as the initial phase in nlp. In *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*.

Harvey A Whiteford, Alize J Ferrari, Louisa Degenhardt, Valery Feigin, and Theo Vos. 2015. The global burden of mental, neurological and substance use disorders: an analysis from the global burden of disease study 2010. *PloS one*, 10(2):e0116820.

Akkapon Wongkoblap, Miguel A Vadillo, and Vasa Curcin. 2019. Predicting social network users with depression from simulated temporal data. In *IEEE EUROCON 2019-18th International Conference on Smart Technologies*, pages 1–6. IEEE.

Min Yen Wu, Chih-Ya Shen, En Tzu Wang, and Arbee LP Chen. 2020. A deep architecture for depression detection using posting, behavior, and living environment data. *Journal of Intelligent Information Systems*, 54(2):225–244.

Xingwei Yang, Rhonda McEwen, Liza Robee Ong, and Morteza Zihayat. 2020. A big data analytics framework for detecting user-level depression from social networks. *International Journal of Information Management*, 54:102141.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.

Amir Hossein Yazdavar, Mohammad Saeid Mahdavinejad, Goonmeet Bajaj, William Romine, Amit Sheth, Amir Hassan Monadjemi, Krishnaprasad Thirunarayan, John M Meddar, Annie Myers, Jyotishman Pathak, et al. 2020. Multimodal mental health analysis in social media. *Plos one*, 15(4):e0226248.

Lixia Yu, Wanyue Jiang, Zhihong Ren, Sheng Xu, Lin Zhang, and Xiangen Hu. 2021. Detecting changes in attitudes toward depression on chinese social media: a text analysis. *Journal of affective disorders*, 280:354–363.

Yin Zhang, Rong Jin, and Zhi-Hua Zhou. 2010. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1):43–52.