# Trend of Depression in Online Mutual Help Group for Major Depressive Disorder by a Two-step NLP Method

*Keywords: Major depressive disorder, Mutual help group, BiLSTM, Dictionary method, Chinese*

## Extended Abstract

Major depressive disorder (MDD) has become a common problem in recent years, affecting about 280 million people in 2021 (Institute of Health Metrics and Evaluation, 2021). Meanwhile, with the development of the social media, people increasingly share their depressive feelings on these platforms. Besides posting monologues, individuals look for interactions and engagements to communicate, find support, and pursue a sense of belonging (Martínez-Pérez et al., 2014). Online mutual help groups provide such a channel. The purpose of this study is to quantify the degree of depression by a comprehensive two-step Natural Language Processing (NLP) method, and then analyze the trend of depression through within-group interaction in online mutual help groups for MDD.

We collected data from a MDD mutual help group on the Chinese social platform Douban. The group provides a channel for MDD patients to share feelings with each other and emphasizes self-healing ability. In this group, members express depressive feelings to seek suggestions or resonance, share the happiness of recovery, and discuss philosophy or literary work. The sample period is from 08/26/2020 to 09/07/2022. Focusing on the period allowed us to capture the fast growth of interaction, as well as the complete post and reply information. Our final data consisted of 6,403 posts and 105,327 replies offered by 16,380 users.

In order to quantify the degree of depression reflected from the posts and replies, we constructed the depressive index (DI) by a comprehensive two-step NLP method. The traditional dictionary method shows limitation in detecting overall sentiment when the content includes both positive and negative keywords (Hardeniya & Borikar, 2016). As a result, we supplemented a Bidirectional Long Short-Term Memory (BiLSTM) algorithm to exclude confusing context before applying the dictionary method. We trained the BiLSTM model by the largest published Chinese psychological counseling dialogue corpus, which included 4,467 labeled data (Wang et al., 2020). Using the Word2Vec model, each post was vectorized into a 1 x 300 vector as the input. The BiLSTM model contained four layers and used the Adam optimization algorithm. The loss function was Categorical Cross Entropy. The model classified all posts into two categories, the depressive category and the non-depressive category. The depressive category contained posts that reflected depressive sentiments, while the non-depressive category contained posts that discussed philosophical therapy or reflected positive sentiments. The classification result is shown in Figure.1. From the figure, the model achieved the best result with an accuracy of 92.3% (precision = 95.5%, recall = 90.4%, F-1 score = 92.9%) and a loss of 0.205 after 13 epochs. A total of 1,161 (18.1%) posts were classified into the non-depressive category, and 5,242 (81.9%) posts were classified into the depressive category.

Secondly, we built three dictionaries using keywords from six official depression assessment sheets and diagnostic manuals (CES-D, BDI, ICD-10, DSM-5, PHQ-9, and HAMD) to identify a depression index (DI) for each post. The three dictionaries separately contained keywords for negative sentiments, MDD symptoms, and self-harm/suicide. In the non-depressive category, the DI of posts was defined as 0; for posts in the depressive category and replies, the DI was calculated as a linear combination of the keyword ratios. To validate our dictionaries, we human-coded 1,500 samples into five levels from most positive to most negative sentiment. Figure.2 compares the normalized DI with human-coded labels, suggesting a positive association between our DI and human-coded labels.

There are three main findings in this study. First, we find that most replies show relatively low DI. Distributions of DI for posts and replies are shown in Figure.3. For the posts, the mean DI is 0.155 ($SD$ = 0.144), with about 26.0% DI of 0; for replies, the distribution is more right-tailed, with mean DI of 0.0617 ($SD$ = 0.0856) and 51.6% DI of 0. In general, replies are more positive compared with posts, which indicates the dominance of comfort and encouragement in the replies. Second, posts with no reply tend to have a lower DI compared with posts with replies ($p < 0.05$), which demonstrates members' tendency for replying to more depressive posts. Third, DI in general decreases as the ordinal number of the post increases. The trend of DI along time is shown in Figure.4. The mean DI of the 20$^{th}$ posts is significantly lower than DI of the 1$^{st}$ posts ($p < 0.05$). The trend indicates that, through the within group interaction, posters become less depressive as they post more in the group.

To conclude, we quantify the trend of depression degree reflected from the content in online mutual help groups for MDD by introducing a comprehensive two-step NLP method. Data demonstrates that DI of posts decreases as the posters post more times in the group. The possible reason is that members tend to reply positively, especially when the posts show severe depression, and thus posters show less depression after receiving this comfort and encouragement. This study result is important in the field of psychology and MDD. Based on the result, psychologists can try to offer related treatment, such as encouraging communication groups with comfort from other patients, which may have additional advantages beyond the help from specialists.

## References

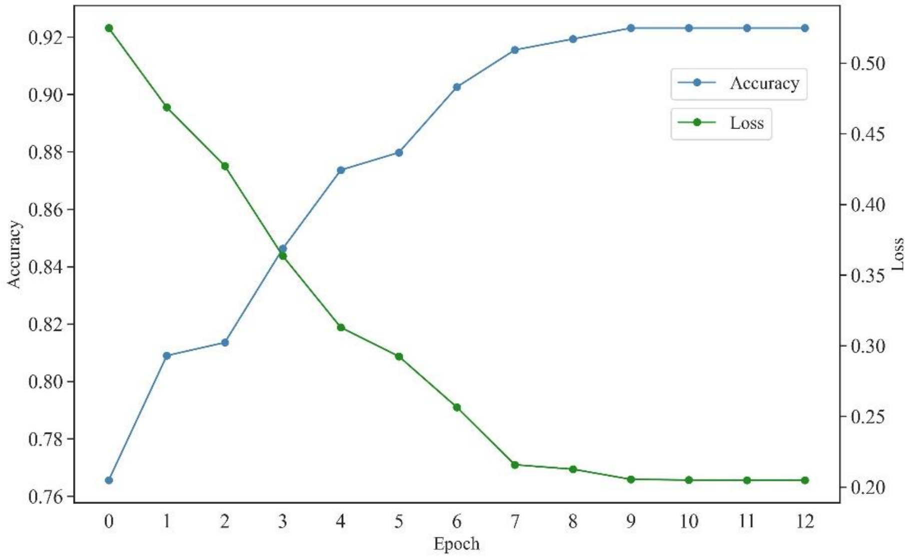
Institute of Health Metrics and Evaluation. (2021, May 1). *Global Health Data Exchange.* GHDx. Retrieved February 12, 2023, from https://ghdx.healthdata.org/

Martínez-Pérez, B., de la Torre-Díez, I., Bargiela-Flórez, B., López-Coronado, M., & Rodrigues, J. J. P. C. (2014). Content analysis of neurodegenerative and mental diseases social groups. *Health Informatics Journal*, *21(4)*, 267 – 283. https://doi.org/10.1177/1460458214525615

Hardeniya, T., & Borikar, D. A. (2016). Dictionary based approach to sentiment analysis-a review. *International Journal of Advanced Engineering, Management and Science*, 2(5), 239438.

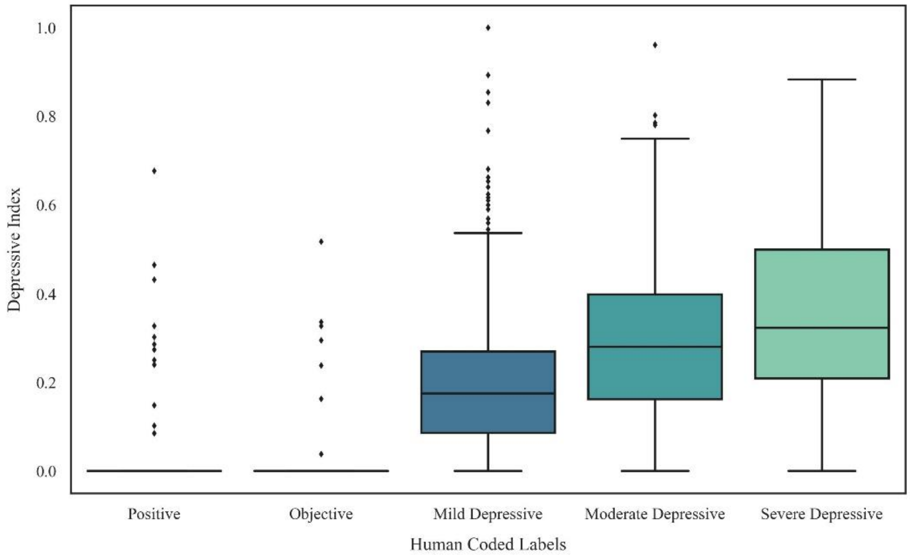
Wang, H., Wu, Z., & Lang, J. (2020, April 22). *Emotional First Aid Dataset*. GitHub. Retrieved February 12, 2023, from https://github.com/chatopera/efaqa-corpus-zh

**Figure 1**
*Accuracy and Loss of BiLSTM Model*



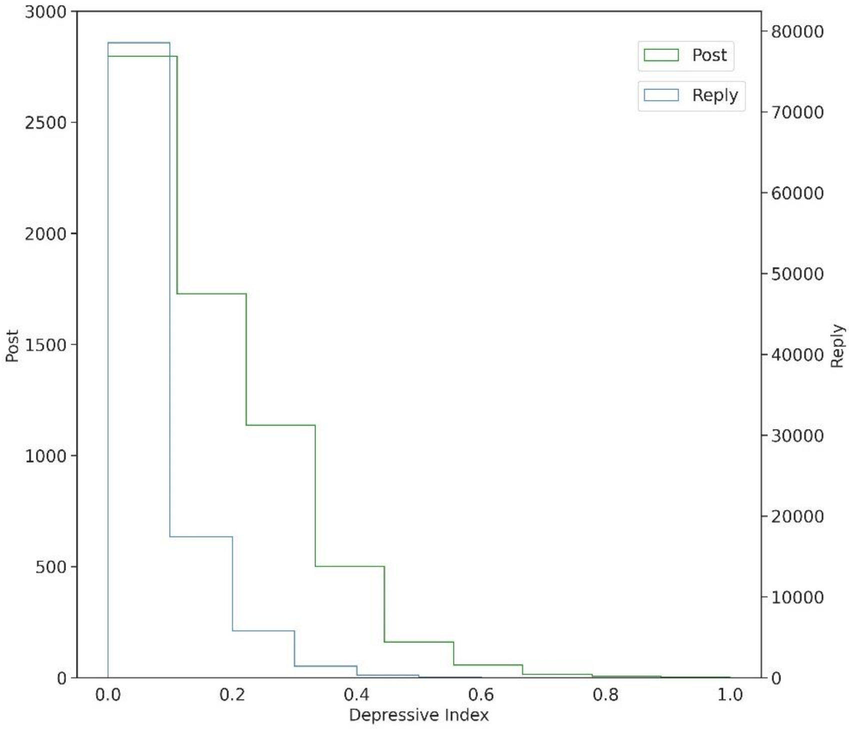

*Note.* The blue curve represents the trend of accuracy through training for 12 epochs with y-axis on the left. The green curve represents the trend of loss with y-axis on the right.

**Figure 2**
*Validation Result*



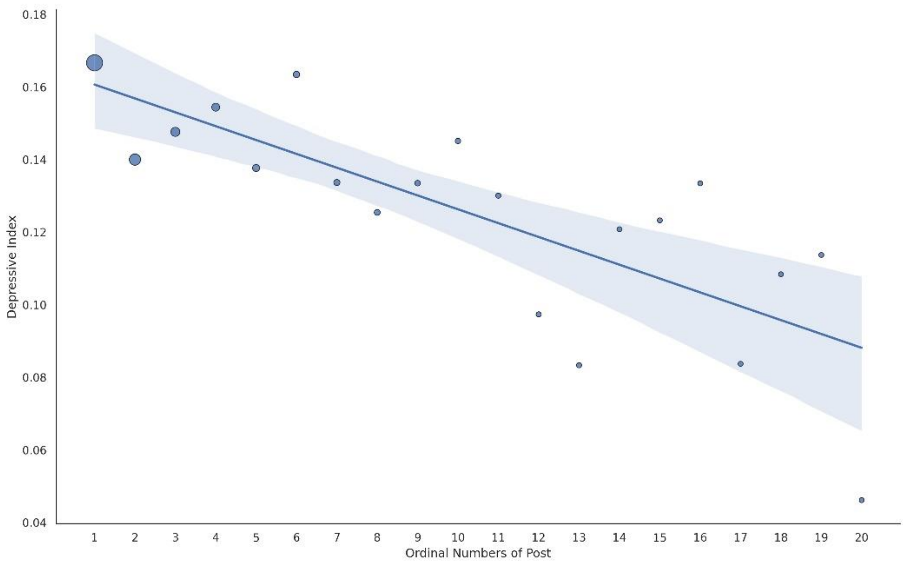

*Note.* Each box represents the distribution of Depressive Index of posts in the same human-coded depression level. The severity of human-coded depression level increases from left to right.

**Figure 3**
*Distributions of Depressive Index for Posts and Replies*




*Note.* The green bar represents the count of posts in each interval of Depressive Index with y-axis on the left. The blue bar represents the count of replies in each interval of Depressive Index with y-axis on the right.

**Figure 4**
*Trend of Depressive Index along Time*




*Note.* Each point represents the mean value of Depressive Index of posts with the same ordinal number. The line represents the fitted trend. The shadow represents the shaded error bands. The size of the point represents the number of posts for the same ordinal number.