

# Text Analysis for Depression Detection: Mental Health Digital Transformation

Samaneh MADANIAN<sup>a,1</sup> and Yuan Gao<sup>a</sup>

<sup>a</sup>*Department of Data Science and Artificial Intelligence, AUT*

ORCID ID: Samaneh MADANIAN <https://orcid.org/0000-0001-6430-9611>

**Abstract.** Depression is a pervasive mental health disorder affecting millions globally. The rise of social networks and their digital footprint provides a unique avenue to leverage AI for early identification of users who may be suffering. We built upon BERT for feature extraction from individual user posts, followed by a Convolutional Neural Network for classification. Since the pre-trained BERT model may not effectively capture social media language, we propose an approach to pre-train BERT on Reddit data before integrating it into the BERT+CNN architecture.

**Keywords.** Digital Health, Digital Mental Health, Depression, AI, NLP.

## 1. Introduction and Background of the Research

Depression is a serious mental health disorder affecting an estimated 3.8% of the global population, including 5.0% of adults and 5.7% of older adults World Health Organization [1]. It often goes undiagnosed due to its subtle presentation [2]. To enhance detection, researchers use objective measures, such as analyzing Social Media content (SM), which can reveal depressive symptoms before individuals seek help [3]. Recent studies employed crowdsourcing to gather gold standard labels and developed metrics to identify users with depressive symptoms [4]. They found consistent linguistic patterns related to mental health issues across different cultures, e.g., in English and Japanese tweets [5]. This cross-cultural consistency indicates that certain linguistic indicators of mental health issues might have universal applicability. However, a key limitation is adapting BERT's pre-training data, which is formal, to the informal language of social media. Our research aims to enhance detection accuracy by pre-training BERT models on Reddit data, thereby better capturing the language patterns associated with depression.

## 2. Research Method

We utilized the SMHD dataset [2], which comprises the original depression data (1316 users) and the data for the control group (also 1316 users), both of which were randomly sampled. We split the dataset into training, validation, and testing. We used the original depression data, and the data for the control group were randomly sampled, with the same number of users as the depression group. Inspired by [6], we adopted BERT+CNN pipeline (Figure 1). After processing through BERT, we obtained user-level features with dimensions of  $768 \times n_k$ , where 768 is the length of the sentence-level embedding output by the BERT model for a single post. The CNN model processes an input matrix of shape  $(N, 768)$ , where N is set to 512. It features two 1D convolutional blocks, each with a

---

<sup>1</sup> Corresponding Author: Samaneh Madanian, email: [sam.madanian@aut.ac.nz](mailto:sam.madanian@aut.ac.nz).

convolutional layer (first:  $40 \times 768$ , 32 filters; second:  $40 \times 32$ , 16 filters), followed by max-pooling and a 20% dropout rate. ReLU activations are used for non-linearity.

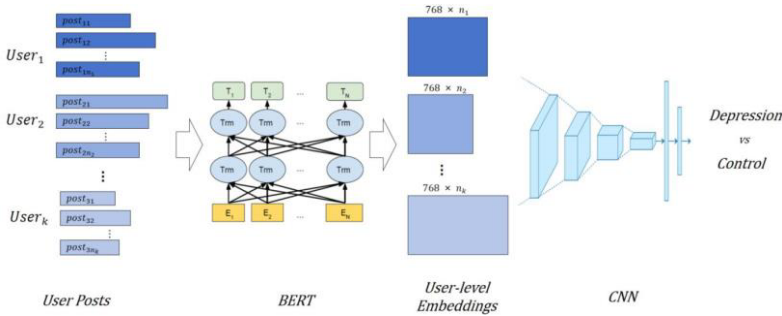


Figure 1. BERT+CNN pipeline.

### 3. Results and Discussion

Training with the original BERT model achieved the highest F1 score, 0.813, which increased to 0.834 after pretraining the BERT model. The improvements in depression detection validate our hypothesis that adapting language models to SM contexts can improve depression detection accuracy. This improvement likely stems from the model's better understanding of informal language patterns and context-specific expressions commonly used in SM discussions about mental health. From a clinical perspective, this improved detection accuracy represents an advancement in our ability to identify individuals who may be experiencing depression through their SM activity. This could enable earlier intervention and support, particularly for individuals who might not otherwise seek professional help. Our findings suggest the potential for developing scalable, cost-effective screening tools that can complement existing mental health services, particularly in resource-limited settings and for public mental health monitoring.

### References

- [1] World Health Organization, Depressive disorder (depression), in, WHO, Online, 2023.
- [2] E.I. Fried and R.M. Nesse, Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR\*D study, *Journal of Affective Disorders* **172** (2015), 96-102.
- [3] S. Madanian, H. Rasoulipannah, and J. Yu, Stress Detection on Social Network: Public Mental Health Surveillance: Public mental health surveillance, in: *Proceedings of the 2023 Australasian Computer Science Week*, Association for Computing Machinery, Melbourne, VIC, Australia, 2023, pp. 170–175.
- [4] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, Predicting Depression via Social Media, *Proceedings of the International AAAI Conference on Web and Social Media* **7** (2021), 128-137.
- [5] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, Recognizing Depression from Twitter Activity, in: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Seoul, Republic of Korea, 2015, pp. 3187–3196.
- [6] Z. Chen, R. Yang, S. Fu, N. Zong, H. Liu, and M. Huang, Detecting Reddit Users with Depression Using a Hybrid Neural Network SBERT-CNN, in: *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, 2023, pp. 193-199.