IMPERIAL

# CLASSIFYING SOCIAL MEDIA USERS BEFORE AND AFTER DEPRESSION DIAGNOSIS VIA THEIR LANGUAGE USAGE: A DATASET AND STUDY

**Falwah Alhamed** Imperial College London
**Julia Ive** Queen Mary University of London
**Lucia Specia** Imperial College London
{f.alhamed20,l.specia}@imperial.ac.uk
j.ive@qmul.ac.uk

## INTRODUCTION

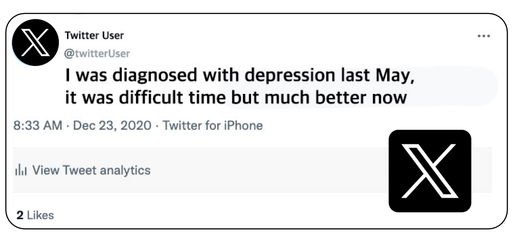
- According to the NHS, one in four people experiences mental health problems
- Mental illness can significantly impact individuals' quality of life
- Users on social media often share updates about their daily lives, including moods and feelings
- Most studies focus on the classification of users suffering from depression versus healthy users, or on the detection of suicidal thoughts

## CONTRIBUTION

- The first English dataset of textual posts **by the same users** before and after reportedly being diagnosed with depression
- A lexicon for finding posts with symptoms of depression
- Empirical work comparing multiple predictive models (based on SVM, Random Forests, BERT, RoBERTa, GPT-3, GPT-3.5, Bard, and Alpaca) built using our dataset for the task of classifying user posts as before and after depression.

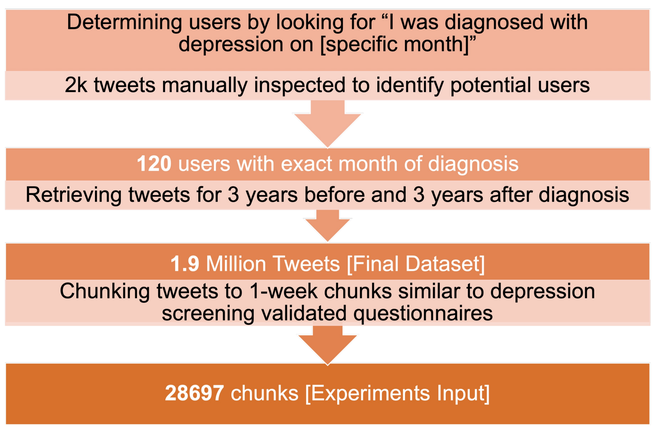
## DATA COLLECTION

**Official X (Twitter) API was used to search and save posts based on keywords.**




*Example of a post used to find users diagnosed with depression on X*

### Data Collection Process




## DATA ANALYSIS

Some analysis of the collected tweets was done to compare users' tweets before and after diagnosis.

### Posting Frequency

- **Original**: Posts written by the user in his/her timeline.
- **Quoted**: Posts are originally written by someone else, and the user quotes this tweet and adds a comment then posts it in his/her timeline.
- **Replied**: Posts are posted by someone, and the user replies to this specific tweet.
- **Retweeted**: Posts are originally written by someone else, and the user reposts it in his/her timeline.

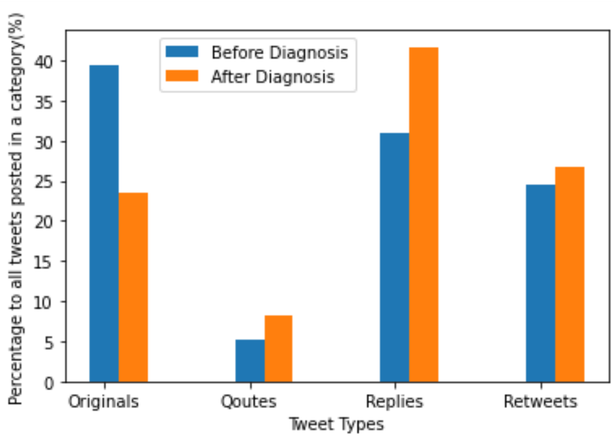



***Figure 1:*** *Posting Frequency comparison between posts types before and after diagnosing with depression*

### Linguistic Inquiry and Word Count LIWC

Users' posts are analyzed using LIWC tool to understand the style of writing for users before and after being diagnosed with depression.

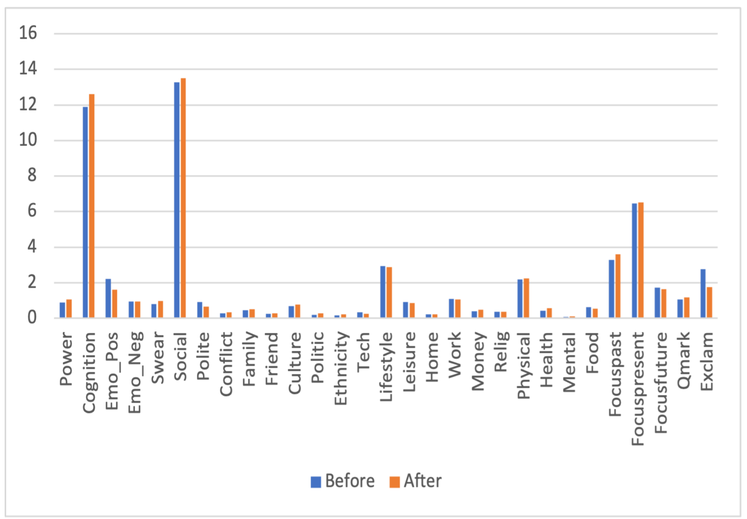



***Figure 2:*** *LIWC comparison between tweets before and after diagnosing with depression*

## BUILDING DEPRESSION LEXICON

To obtain representative posts from the dataset we aimed to create and develop a depressive lexicon that can be used to screen texts for depression symptoms. The main goal of this lexicon is to filter out posts that do not contain any symptoms of depression and will act as noise for annotators and models. The lexicon was created by using words related to depression symptoms in the validated questionnaire CES-D which can be categorized into: poor appetite and eating disturbance, feeling down and depressed, concentration problems, feeling tired or having little energy, sleep disturbance, loss of interest, self-blame and shame, loneliness, and suicidal thoughts. For each category, we created a list of relevant words.

## MODELS

The classification unit is a **chunk** of posts and the same user will have some chunks labelled as positive (after depression) and negative (before depression).

- **Classical ML Models**
  - SVM Word2Vec embeddings
  - RF Word2Vec embeddings
- **Transformer-based Models**
  - BERT Hugging Face "bert-base-uncased" model card.
  - RoBERTa Hugging Face "roberta-base" model card.
  - MantalBERT Hugging Face "mental-bert- base-uncased "
- **Large Language Models (LLMs) [zero-shot]**
  - GPT-3 "text-curie-001"
  - GPT-3.5 "text-davinci-003"
  - Google Bard (Gemini)
  - Alpaca

### Prompts

| |
|---|
| Do you think the person who wrote this text is depressed? |
| Do you think the person who wrote this text is depressed? answer yes or no |
| Do you think the person who wrote this text is depressed? return a probability percentage |
| To what extent do you think the person who wrote this text is depressed? |
| To what extent do you think the person who wrote this text is depressed? answer with one word only |
| Classify if the person who wrote this text is depressed |
| Classify if the person who wrote this text is depressed, reply with one word only |

***Table 1:*** *Prompts used for LLMs answers*

### T-SNE Visualization of BERT

To assess whether BERT is learning meaningful patterns rather than overfitting the data, we explored the learned representations in a lower-dimensional space using T-SNE. T-SNE can be used to visualize word embeddings or document embeddings to gain insights into the semantic relationships between classes. The visualization in Figure 3 illustrates that BERT was able to capture meaningful patterns between the posts before and after diagnosis and was able to split data based on this understanding.

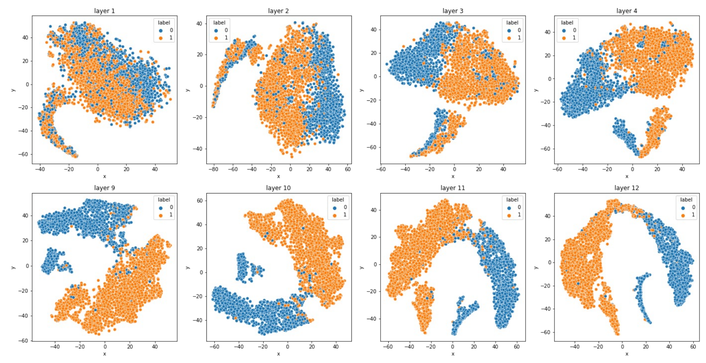



***Figure 3:*** *T-SNE visualization of BERT classification*

## RESULTS

| Model | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| SVM | 0.49 | 0.49 | 0.49 | 0.44 |
| SVM-Filtered | 0.60 | 0.50 | 0.60 | 0.48 |
| RF | 0.50 | 0.50 | 0.50 | 0.50 |
| RF-Filtered | 0.60 | 0.63 | 0.60 | 0.45 |
| BERT | 0.90 | 0.91 | 0.90 | 0.90 |
| BERT-Filtered | **0.98** | **0.98** | **0.98** | **0.98** |
| RoBERTa | 0.97 | 0.97 | 0.97 | 0.97 |
| RoBERTa-Filtered | **0.98** | **0.98** | **0.98** | **0.98** |
| MentalBERT | 0.96 | 0.96 | 0.96 | 0.96 |
| MentalBERT-Filtered | **0.98** | **0.98** | **0.98** | **0.98** |

***Table 2:*** *Results for classical and transformers-based models on classifying data on full dataset and lexicon-filtered dataset.*

| Model | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| SVM-Filtered | 0.60 | 0.50 | 0.60 | 0.44 |
| RF-Filtered | 0.60 | 0.63 | 0.60 | 0.45 |
| BERT-Filtered | **0.97** | **0.97** | **0.97** | **0.97** |
| RoBERTa-Filtered | 0.95 | 0.96 | 0.95 | 0.95 |
| GPT-Filtered "Text-Curie-001" | 0.51 | 0.51 | 0.247 | 0.32 |
| BARD-Filtered | 0.46 | 0.5 | 0.285 | 0.36 |
| Alpaca-Filtered | Hallucinating | | | |

***Table 3:*** *Comparison between results for all models with chunk length trimmed to 2049 tokens.*

## DISCUSSION

### Lexicon Filtering

Our results in Table 2 show that using our lexicon to filter the dataset improved results according to all metrics, especially in precision and recall.

### Chunk Length

Our experiment shows that chunking data rows into one-week chunks that align with the clinical periodicity of depression diagnosis validated questionnaires yields significantly improved results compared to previous studies that applied the same model for post- level classification

### LLMS

While LLMs are powerful language models known for their NLP capabilities, LLMs have faced challenges in certain mental health classification tasks where they have not performed as effectively as transformer-based models .

**Dataset and lexicon are available on:**
**https://github.com/falwah-alhamed/Depression_Tweets**

## References

[1] Marcus, M., Yasamy, M. T., van Ommeren, M., and Chisholm, D. (2012). Depression, a global public health concern. *WHO De- partment of Mental Health and Substance Abuse*, pages 1–8.

[2] Satinsky, E., Fuhr, D. C., Woodward, A., Sondorp, E., and Roberts, B. (2019). Mental health care utilisation and access among refugees and asylum seekers in Europe: A systematic review. Health Policy, 123(9):851– 863.

[3] Onan, A., Korukog˘lu, S., and Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. Expert Systems with Applications, 57:232–247.

[4] Pranckevivius, T. and Marcinkevicius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classifica- tion. Balt. J. Mod. Comput., 5.

[5] Al-Garadi, M. A., Yang, Y.-C., Cai, H., Ruan, Y., O'Connor, K., Graciela, G.-H., Perrone, J., and Sarker, A. (2021). Text classification models for the automatic detection of nonmedical prescription medication use from social media. BMC Medical Informatics and Decision Making, 21(1):27.

[6] Acheampong, F. A., Nunoo-Mensah, H., and Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review.

[7] Matero, M., Idnani, A., Son, Y., Giorgi, S., Vu, H., Zamani, M., Limbachiya, P., Guntuku, S. C., and Schwartz, H. A. (2019). Suicide Risk Assessment with Multi-level Dual-Context Language and. pages 39–44.