

# Deep Learning-based Pre-diagnosis and Analysis of Psychological Disorders

Yuxin Bai\*

Southwest Minzu University, Chengdu, 610041, China

---

**Abstract:** At present, due to the lack of sufficient treatment institutions as well as professional psychotherapists, numerous patients with mental illnesses do not receive help from professional psychiatrists, thus worsening their conditions. In order to determine the condition of mental illness patients at an early stage, this paper applies natural language processing technology to the psychological field and proposes a text-based and deep learning model for mental illness recognition. The BERT (Bidirectional Encoder Representations from Transformers) pre-trained language model is used to complete the sentence-level feature vector representation of mental health text data, and the obtained feature vectors are subsequently targeted and input to a classifier for classification, which can effectively identify depression, anxiety, post-traumatic stress, and unmet mental illness multiple classifications, breaking the previous common depression identification, improving from the original simple depression dichotomous classification to a mental illness multiple classification task, and focusing on a few focal points, which not only saves human and material resources, but also can achieve twice the result with half the effort. Finally, the algorithm is validated using mental health text dataset, and the experimental results show that the lowest F1 value of the trained model on the test set is 0.77, which can achieve fast screening of text content with the tendency of mental illness, reduce the expert labeling workload, improve the labeling efficiency, and provide a new idea for the recognition of mental illness.

**Keywords:** Mental illness, Natural language processing techniques, Deep learning, Pre-diagnosis.

---

## 1. Introduction

According to data published by the World Health Organization: 6 billion people in the world, about 17% have psychosomatic disorders; 76-85% of patients, need the help of professional psychiatrists. However, there is a lack of adequate treatment institutions as well as professional psychotherapists, and according to statistics as of 2020, nearly 1 billion people worldwide suffer from mental illness[1]. 2020 after the outbreak of the new crown epidemic, the incidence of depression and anxiety increased by 25%, and mental health problems are increasingly acute conflicts[2-3]. The impairment of mental illness on the patient's ability to live socially and work productively is enormous, and the mortality rate of people with mental disorders is higher compared to the general population, but most patients do not die from mental disorders; instead they usually die from suicide or other chronic illnesses[4]. Patients with mental disorders have a high prevalence of poor health behaviors, including smoking, substance use, lack of physical activity, and poor dietary habits. In turn, these behaviors contribute to the high prevalence of chronic illness in people with mental disorders.

The commonly used method for detecting mental disorders is based on psychological scale assessment, and although this method is a good predictor of whether a user has a mental disorder, there are many problems, for example, psychologists can only passively wait for the person to actively seek help, which may miss people with depressive tendencies, and cannot achieve real-time monitoring of the mental health status of a large scale population, and the feedback time of identification results lags; in addition, patients with mental disorders In addition, patients with mental illness may choose positive alternatives that do not fit the description of their state or show cognition and behavior that do not fit their state.

With the gradual popularity of social media and the continuous development of computer technology, the number of users joining the social network is increasing continuously. More and more netizens are beginning to express their opinions and emotional dynamics in social media, and more and more personal posting information can be collected on the network. Social media has gradually become an important way for people to share the latest emotional information and discuss hot spots of public opinion, as well as an effective way to get information. Therefore, this paper presents a text-based and BERT-based mental disease recognition model, which not only abandons the inefficient mental scale evaluation method, but also classifies mental disorders into multiple categories so that psychologists can conduct targeted medical testing.

## 2. Methods

### 2.1. Data Collection Preprocessing

This paper uses the well-labeled Pate Psychology dataset, which is the first open question and answer corpus in counseling, including 20,000 counseling data, and the largest publicly available Chinese counseling conversation corpus[5]. The dataset was built with the participation of psychology professionals from Stanford University, UCLA, and Fu Jen Catholic University in Taiwan, as well as the collaboration of Chatopera and many volunteers. The dataset is rich in content, not only with multiple rounds of conversation content, but also with information such as classification. The dataset contains three dimensions: 'S1': type of worry; 'S2': mental illness; and 'S3': suicidal tendency. In order to determine the condition of patients with psychological disorders at an early stage, to conduct more professional medical tests, and to prevent patients from having poor health behaviors as well as suicidal behaviors, only the 'S2' psychological disorders dimension is studied in this paper, and the data set used is shown in Table 1.

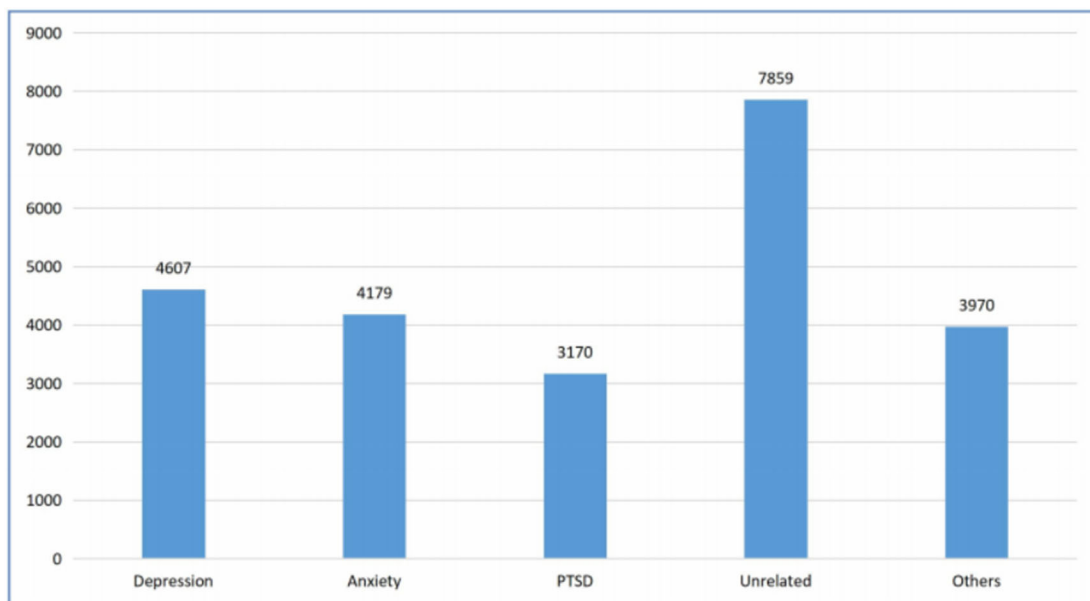
**Table 1.** Data set introduction

Data set	Number of data	Symptoms
Depression	906	Prolonged and persistent depressive mood which is significantly more than necessary, lack of self-confidence, avoidance of people
Anxiety	716	long-lasting anxiety, without a clear objective object but still nervous and worried, fidgeting
Bipolar disorder	94	Also known as "bipolar disorder". Manic phase: Feeling alive, energetic and emotionally charged or easily provoked
Post traumatic stress disorder	236	The first step is to experience trauma: physical or psychological abuse as a child; mental or physical discomfort and tension when exposed to something relevant, and the traumatic scenario will be replayed in your mind over and over again.
Panic disorder	69	Also known as acute anxiety disorder, it is a recurring panic attack. Panic attacks are sudden short-term intense fears (near-death feelings)
Anorexia and bulimia	33	Anorexia: eating too little resulting in low weight; bulimia: eating a lot and then trying to throw up.
Unrelated	17795	Not yet serious enough for mental illness
Others	156	A condition that has seriously affected life and work, or even life and work cannot be performed, but it is not possible to confirm which type of disease it is

Obviously the original dataset is extremely unbalanced, and if the unbalanced ratio exceeds 4:1, the data imbalance will lead to model learning bias, and the model will tend to learn the features of the data with a high ratio, and only learn few features for the data with a low ratio, which leads to poor or even unpredictable prediction performance. To effectively address the problems caused by unbalanced datasets, common treatments include expanding the dataset, resampling the dataset, and artificial data, with the aim of enhancing the overall quantity and quality of the sample, and also adding noise to improve model performance and prevent model overfitting. In the field of NLP, data augmentation has been less studied, and the two more stable and commonly used

methods are back translation and EDA (Easiest Data Augmentation). In this paper, the EDA method is used for training set data augmentation because, compared to back translation, EDA is a deterministic transformation that can be arbitrarily specified for the design of a specific scenario.

Since the data samples of bipolar disorder, panic disorder, anorexia nervosa and bulimia nervosa remain small after data augmentation, this paper readjusts the data samples of these three categories and classifies them uniformly as other psychological disorders, and the number of data samples after augmentation is shown in Figure 1. The ratio of training data test data to validation data is 8:1:1.

**Figure 1.** Sample size of the enhanced data

## 2.2. Data Analysis

In this paper, we counted the high-frequency words and characteristics of the mental illness group. High-frequency words are generally words that appear more frequently in

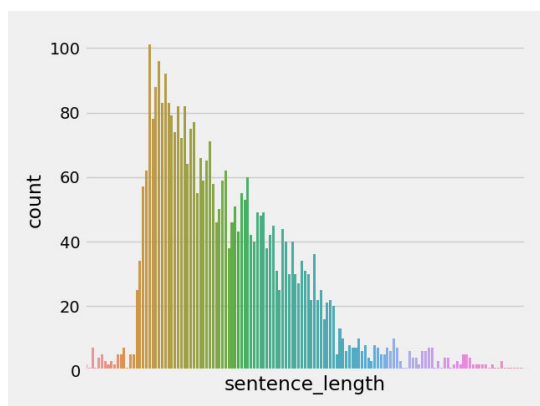
documents and are not useless, and they represent the focus of attention of the mental illness group to some extent. The top 15 high-frequency words of the mental illness group are listed here, as shown in Table 2.

**Table 2.** High Frequency Words in Psychological Disease Groups

Groups	High Frequency Words
Mental illness groups	Feelings, Depressive disorder, Feel fear, Always, Mind, Don't want, Depression, Anxiety, Insomnia, Others, Now, Daily, Mental bad

The analysis of the high-frequency words in Table 2 shows the characteristics of perception, thinking, attention, emotion, consciousness and action behavior of the mental illness group. Firstly, the frequency of using the verb "feel" in the text of the mental illness group is the highest, which indicates that the mental illness group has a stronger sense of self-subjectivity, is more physically sensitive, and lives in its own world, and has more sensory verbs in the high-frequency words, which shows that the true self-presentation is slightly higher than the positive self-presentation. Secondly, people with mental illness sometimes feel that they have a tendency to mental illness, such as "depression", so they often talk about related topics around the theme of "depression", such as "depression", "anxiety", "insomnia", and so on. In addition, depressed people are almost always depressed, sad, irritable, uninterested in everything and have difficulty experiencing any pleasure, and have significant sleep disturbances.

The next analysis of the mental illness dataset was performed by sentence length distribution, which allowed to obtain texts with the longest characters around 100, as shown in Figure 2. Usually, texts with no more than 160 characters are considered as short texts, so this paper focuses on short texts.

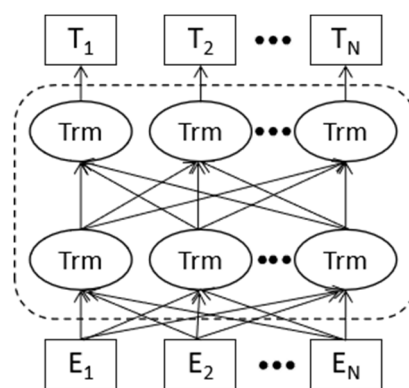


**Figure 2.** Sentence length distribution

### 2.3. BERT Model

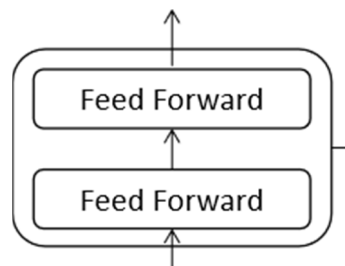
BERT model is the most popular pre-training model based on language model, the token and sentence in BERT series model are generally modeled together, which is also an advantage of Transformer model, the representation of each token in this layer is the result of using attention for all tokens in the previous layer, so it makes getting[6]. The representation of the sentence becomes exceptionally simple by using only a special token like [CLS]. In contrast, the traditional word2vec model does not have a particularly good way to obtain the sentence-level representation directly, and the general practice is to average all word vectors, however, this is not equivalent to the displayed modeled sentence representation. the BERT model is a combination of the pre-training model and the downstream task model, which means that the BERT model is still used when doing the downstream task, and naturally supports the text classification tasks, and

no modifications to the model are required when doing text classification tasks. bert model consists of a multi-layer bidirectional transformer encoder with a Transformer encoding unit consisting of 6 Encode and 6 Decode stacked together. the structure diagram of BERT is shown in Figure 3, which consists of an input layer, an encoding layer, and an output layer.  $e\{E_1, e_2, \dots, E_N\}$  denotes the input of text, i.e., the input layer; the encoding layer consists of multiple Transformers stacked together;  $T\{T_1, T_2, \dots, E_i\}$  has a one-to-one correspondence with  $T_i$ , and  $i$  is the  $i$ -th word of the input text sequence[7].



**Figure 3.** Bert model structure

For the Transformer encoder, an encoder contains two layers, the Self-Attention layer and the feedforward neural network, as shown in Figure 4.



**Figure 4.** Encoder

The Self-Attention layer helps the current node to not only focus on the current word but also to obtain the semantics of the context. In order to enhance the semantic representation of the target word with contextual information in a differentiated way, the Attention mechanism involves three main concepts: Query, Key and Value. The Self-Attention mechanism takes the semantic vector representations of the target word and each context word as input, and first obtains the Query vector representation of the target word, the Key vector representation of each context word, and the original Value representation of the target word and each context word by linear transformation, then calculates the similarity between the Query vector and each Key vector as weights, and weightedly fuses the Value vector of the target word and

the Value vector of each context word as the output of Attention, i.e., the enhanced semantic vector representation of the target word. Attention is calculated as follows.

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where  $d_k$  denotes the dimensionality of the query and key vectors for each word, and  $Softmax()$  is the normalized exponential function. The final Attention value is a matrix value, and each row of the matrix value represents a new vector representation of the corresponding word Attention vector in the input sentence, which contains information about the interrelationship between the word and the words in other positions in the sentence.

Transformer is the core module of BERT, and the Self-Attention mechanism is the most critical part of Transformer.

## 2.4. Model Design

The research of deep learning based preliminary screening algorithm for mental illness is a feature representation algorithm for short text, but due to the sparse nature of short text, characters or words cannot represent the complete semantics of short text, resulting in the feature representation vector of short text cannot represent the semantics of short text better, so this paper proposes a BERT based preliminary screening model for mental illness, using BERT pre-training model for dataset for feature extraction, which better ensures

the semantic information. The design of the psychological disease prediction model mainly consists of three parts: short text preprocessing, short text vectorization, and short text classification.

The first is text preprocessing, and the specific process is as follows: 1) word separation. In this paper, we use the stuttering word separation method, which combines the rule-based word separation and statistical word separation algorithms, implements efficient word map scanning based on prefix dictionaries, generates a directed acyclic graph of all possible word formation cases of Chinese words in a sentence, and uses dynamic programming to find the maximum probability path, and finds the maximum cut combination based on word frequency. 2) Cleaning. The text cleaning mainly includes the cleaning of useless information, the cleaning of special text and the processing of deactivated words. Deactivated words are words that do not contain or contain very little semantic meaning, because deactivated words and words that occur particularly infrequently are often of little use for analysis, so they are generally removed. 3) Word standardization. The text standardization mainly includes unified digit writing, word form reduction and word stemming extraction. Next, the BERT model is used for text feature extraction, and then feature vector stitching is performed to obtain a short text vector representation. Finally, the feature vectors are fed into the Softmax regression model for training and the classification results are output, and the flow chart is shown in Figure 6.

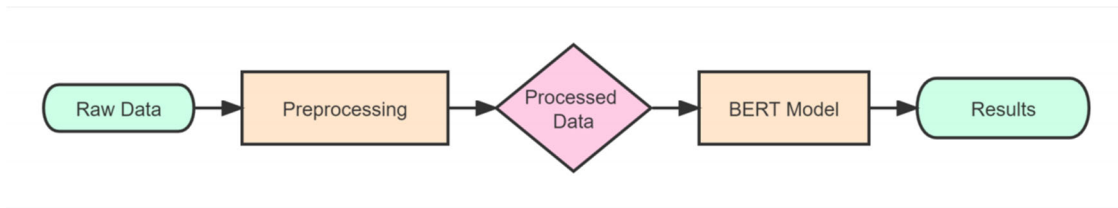


Figure 6. Flow chart of mental illness classification

## 3. Results

### 3.1. Experimental Procedure

In this paper, we use the BERT-Base pre-training model

provided by Google. The model has a 12-layer network structure with 768-dimensional hidden layers, which is efficient in execution and good in classification, and can better express the contextual information.

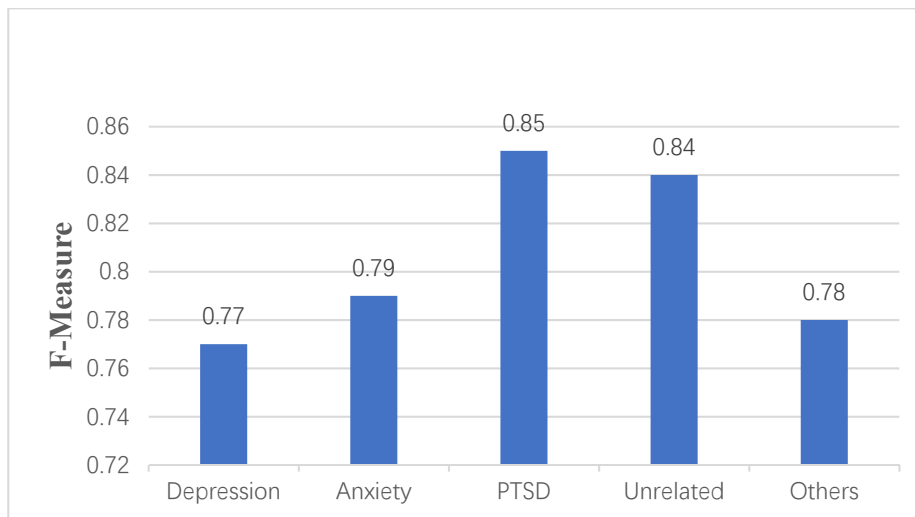
Table 3. Parameter settings

Parameter Name	Parameter Value
Number of model iterations	10
Learning rate	2e-5
Gradient pruning	5
Maximum sequence length	103
Amount of data in each batch training set	32

### 3.2. Experimental Results

Using the mental illness group dataset, this paper analyzes and summarizes the characteristics of perception, thinking, attention, emotion, consciousness and action behavior of the

mental illness group. Based on these characteristics, the preliminary screening algorithm for mental disorders was supported, and this paper did a multiple classification task for mental disorders, and the evaluation index mainly used F1 values, and the results are shown in Figure 5.



**Figure 5.** Classification results

To address the problems of traditional face-to-face conversations to detect depression and the inadequacy of detecting only depression, this study proposes a deep learning-based preliminary screening algorithm for mental illness to detect early predisposition to mental illness. The BERT model is used to extract features from the dataset to better ensure the semantic information, and a training classifier is used to implement a multi-classification task for mental disorders, and it is tested on a generic labeled dataset to verify the effectiveness and advancement of the proposed method.

#### 4. Discussion

In traditional clinical diagnosis, psychologists need to conduct face-to-face interviews with visitors to diagnose whether they are suffering from depression and the extent of the condition. However, there are many potential problems with this approach. Given that patients with depressive tendencies are more willing to confide their moods and states to social media, computer-assisted methods are used and textual information in social media is utilized. In this paper, we use the BERT model instead of the commonly used word2vec model for vector representation of short texts in the detection of users with mental illness tendencies, and propose a mental illness screening model based on the BERT model, and improve it from the original simple depression dichotomous classification to the mental illness multiclassification. The experimental results show that the algorithm has a good classification effect on multiple categories of mental illness data, however, this paper does not incorporate positional information such as expressions and punctuation marks on sentence representations to enrich the sentence vector feature representation of short texts, which causes some semantic loss; some conditions should be set on the determination of mental illness to decide when to be

classified as a mental illness group; the deep learning superior performance is mainly obtained by training with a large number of labeled datasets; currently there are fewer datasets focusing on depression recognition with textual information, the training dataset is not large enough and the knowledge that can be learned is limited, and the model effect can be improved later by expanding the dataset.

#### Acknowledgment

This work was supported by the Funds of Southwest Minzu University for Innovative research project(320022250116).

#### References

- [1] Health T L G. Mental health matters[J]. The Lancet. Global Health, 2020, 8(11): e1352.
- [2] Yohannes A M. COPD patients in a COVID-19 society: depression and anxiety[J]. Expert Review of Respiratory Medicine, 2021, 15(1): 5-7.
- [3] Li Y, Wang A, Wu Y, et al. Impact of the COVID-19 pandemic on the mental health of college students: a systematic review and meta-analysis[J]. Frontiers in psychology, 2021, 12.
- [4] Fryer D. Unemployment and mental health: Hazards and challenges of psychology in the community[M]//Health effects of the new labour market. Springer, Boston, MA, 2002: 11-23.
- [5] L.Y.L..Hai Liang Wang, Zhi Zhi Wu, Emotional first aid dataset (2020).URL <https://github.com/chatopera/efaqa-corpus-zh>
- [6] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [7] Wolf T, Debut L, Sanh V, et al. Huggingface's transformers: State-of-the-art natural language processing[J]. arXiv preprint arXiv:1910.03771, 2019.