

Leveraging deep learning model to visually detect changes in depression tendency of social media users

Yukun Yang

School of Management, Beijing Institute of Technology, Beijing, China

Wanxin Qiao

School of Economics and Management, Beijing University of Technology, Beijing, China and

National Science and Technology Innovation Hub and Research Center, Beijing, China, and

Yuting Lin and Rui Liu

School of Management, Beijing Institute of Technology, Beijing, China

Received 14 March 2025
Revised 27 August 2025
Accepted 14 September 2025

Abstract

Purpose – Leveraging deep learning models to detect the depressive tendencies of social media users and visualizing the trajectories of their emotional changes. Addressing the issues of subjectivity bias and high implementation cost existing in traditional depression detection approaches.

Design/methodology/approach – A new multi-instance Long short-term memory (LSTM) learning model is put forward. It employs an autoencoder to extract time-series features from the text, constructs a binary classifier to sift depression-related tweets, determines the depression tendencies of users and generates visual graphs based on the distribution of the time series data to exhibit the dynamic variations of users' emotions.

Findings – Our model is capable of outputting the probabilities of depressive tendencies for each of the user's tweets, and its performance is stronger than that of the fundamental methods.

Practical implications – Our study achieves the dynamic and visual monitoring of depressive tendencies, breaking through the constraints of traditional single-time-point detection. Utilizing social media text data, a low-cost and scalable large-scale depression screening solution is proposed.

Originality/value – Our study achieves the dynamic and visual monitoring of depressive tendencies, breaking through the constraints of traditional single-time-point detection. Utilizing social media text data, a low-cost and scalable large-scale depression screening solution is proposed.

Keywords Deep learning, Depressive tendency detection, LSTM autoencoder

Paper type Research article

1. Introduction

Depression is a severe global public health concern, ranking ninth among global diseases (Yu *et al.*, 2020). Its symptoms encompass persistent feelings of sadness, sleep disturbances, inability to concentrate and loss of interest in various matters (American Psychiatric Association, 2013). In grave circumstances, it may result in physical symptoms. Numerous factors contribute to the development of depression, such as social stress, mental health conditions, historical traumas and physical illnesses. From 2005 to 2015, the number of global depression sufferers increased by 18.4%, with over 332 million individuals diagnosed with depression (World Health Organization, 2017). Along with the escalating number of depression patients, the suicide rate has also increased. According to data released by the World Health Organization (2023a), over 700,000 people worldwide die by suicide each year and suicide has emerged as one of the principal causes of death. The number of deaths attributed to suicide exceeds those from war, homicide and malaria. In 2019, approximately



1.3% of all deaths were caused by suicide. Moreover, suicide has become the second leading cause of death among individuals aged 15–29 (World Health Organization, 2019), with adolescents developing suicidal ideations due to depression. Hence, the notion of early identification of depressive symptoms is of paramount importance to mitigate the risks of suicide and depression.

Traditional depression detection approaches depend on standardized scale measurements, patients' subjective reports and clinical diagnoses by attending physicians. Nevertheless, these methods have deficiencies. For example, participants' responses to standardized scales might be affected by the circumstances, mood and mental state at that time (Liu *et al.*, 2022). Simultaneously, the relationship between doctors and patients, as well as the sense of shame about the disease, can lead to underreporting of symptoms, causing false negative results for the scales. Scale measurements are also challenging to implement over extended periods or on a large scale, lacking temporal granularity (Kumar *et al.*, 2015). With the advancement of social media platforms such as Weibo and Twitter, individuals can express their genuine emotions and viewpoints therein. Social media, owing to its anonymity, real-time feature and the abundance of user-generated data, allows researchers to collect and analyze the content of posts (Biradar and Totad, 2019), offering novel perspectives for mental health detection. Particularly, young people are more inclined to express their personal thoughts and negative feelings when using social media, for its anonymous environment (Ma *et al.*, 2016). According to the data, 20% of those who attempted suicide and 50% of those who died by suicide left behind last words (Dejong *et al.*, 2010). At the same time, given the limitations of professional-scale-based assessments, such as high costs and their inability to support large-scale screening of individuals with depression, an increasing number of researchers are exploring the use of social media data for depression detection. By analyzing users' posts, these studies aim to enable more scalable screening and facilitate early intervention.

Our research aims to implement a multi-instance LSTM model and employ the public text dataset of Weibo to identify depressed users and visualize the variations in depression tendencies. We utilize this text data from the following aspects: (1) Construct word embedding vectors based on Weibo data to obtain the initial vectors for input into the LSTM autoencoder and acquire the output vectors after passing through the LSTM layer. (2) Place the output vectors into a pre-trained binary classifier. Herein, we compare three baseline models and obtain the depression probability corresponding to each individual post under the optimal binary classifier. (3) Adopt the concept of multi-instance learning (MIL), set a depression threshold for posts and determine whether a user is depressed. (4) Based on the depression probability derived from the binary classification, create a distribution map of users' depression tendencies to visualize the dynamic trends.

2. Literature review

2.1 Traditional approaches for recognizing depression

It is estimated that approximately 3.8% of the global population is afflicted with depression, among which 5% are adults (4% males and 6% females) and 5.7% are elderly individuals aged over 60 (World Health Organization, 2023b). Depression has emerged as a key research focus in the domains of psychology and medicine. Traditional methods for depression recognition mainly rely on scales, psychological counseling and clinical diagnosis.

Depression detection scales, as conventional methods for diagnosing depression, encompass both self-report and other-report varieties. Self-report scales include PHQ-9 (Kroenke *et al.*, 2001), Social Phobia Scale (SPS), DASS-21 (Crawford and Henry, 2003), BDI (Beck, 1961) and others; other-report scales comprise Hamilton Depression Rating Scale (HAMD/HDRS) (Hamilton, 1960), Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery and Åsberg, 1979) and suicidal affect behavior-cognition scale (Harris *et al.*, 2015), etc. Typically, self-report scales require patients to respond based on their own emotions and behavioral habits, while other-report scales need to be administered by

professional psychologists or psychiatrists. The advantages of using standardized scales for depression detection include simplicity, speed and quantitative evaluation. Nevertheless, scales might have certain deficiencies. For instance, an individual's subjective assessment of their own emotions and behaviors could be affected by memory bias, mood fluctuations and social expectancy bias. Besides, some research indicates that conducting suicide assessments might bring adverse effects on individuals with depression (Harris and Goh, 2017).

Psychological counseling is also a prevalent approach for identifying depression. Through face-to-face interviews, psychologists are capable of assessing an individual's emotional status, living habits and behavioral patterns. This method typically exhibits a relatively high level of accuracy. However, it is highly reliant on the professional competence of the diagnosticians, and an individual's emotional expression might be constrained, thereby resulting in relatively poor universality.

Clinical diagnosis is typically based on authoritative manuals such as the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) (American Psychiatric Association, 2013), and is usually carried out by assessing the patient's symptoms and life history. This modality holds high authority and professionalism. However, it is inefficient to implement, costly and limited in terms of accessibility and widespread adoption.

Current psychological diagnosis and treatment mostly rely on traditional methods for identifying depression, but they have some shortcomings. Firstly, some patients with depression may not seek medical help proactively, which can lead to missing the best treatment time. Secondly, when using scales or undergoing mental health counseling, patients' subjective factors may affect the accuracy of the scales. Thirdly, most of the identification methods mentioned in this section provide static point-in-time evaluations and lack the temporal granularity needed to monitor the fluctuations in depression symptoms. At the same time, in many regions, access to mental health resources remains relatively limited. These shortcomings highlight the urgency of developing scalable, low-cost and temporally sensitive approaches to depression detection—particularly those leveraging the vast, real-time behavioral data available on social media platforms.

2.2 *The rise of social media and the expression of depressive emotions*

In recent years, the rapid development of social media has exerted a profound influence on emotional expression and mental health. People are more prone to sharing their genuine feelings and living conditions on social media, presenting new possibilities for depression detection.

According to statistical data from Statista (2025), as of February 2025, the global number of Internet users had reached 5.56 billion, accounting for 67.9% of the world's total population. Among them, the number of social media users has reached 5.24 billion, representing 63.9% of the world's total population. Social media platforms such as Facebook, Twitter, Instagram, Weibo and Reddit have become an indispensable part of people's daily lives. Research indicates that the data from social media can encompass negative signals posted by patients with depression. This furnishes an important data source for depression recognition (De Choudhury et al., 2013). For example, users on Facebook and Twitter may imply the existence of their depressive symptoms by posting expressions of negative emotions such as loneliness and anxiety (Guntuku et al., 2017). Furthermore, social platforms provide multi-dimensional channels for emotional expression through text, images and videos, offering more information for the identification of depression.

2.3 *Methods of depression recognition based on social media data*

With the advancement of machine learning and deep learning technologies, researchers have embarked on attempts to identify depressive emotions via automated approaches. These methods primarily hinge on feature extraction techniques such as text analysis, image recognition, or audio-video analysis and employ algorithms for emotion recognition.

[Wei et al. \(2023\)](#) proposed a sub-attention-based multimodal fusion method that integrates audio, visual and textual modalities. Evaluated on the Distress Analysis Interview Corpus - Wizard of Oz (DAIC-WOZ) dataset for major depression detection, it achieved a precision of 0.89 and an F1 score of 0.70. [Yan et al. \(2025\)](#) proposed a deep learning model named “Depressive Emotion-Context Enhanced Network (Depressive Emotion-Context Enhanced Network)” that incorporates a depressive emotion recognition module and a context-aware representation mechanism to improve depression detection from social media content. [Roy et al. \(2020\)](#) put forward an algorithm named “Suicide Artificial Intelligence Prediction Heuristic (SAIPH)”, inputting Twitter data into a series of neural networks and using the output to train a random forest model, attaining an area under curve (AUC) of 0.88. [Mann et al. \(2021\)](#) formulated the depression recognition task as a multiple-instance learning problem and leveraged multimodal social media data for detection. Tapotosh [Ghosh et al. \(2023\)](#) proposed a model based on attention-based Bilstm-CNN to detect the depressive state of Bengali social media text. [Guo et al. \(2023\)](#) focused on Chinese social media and proposed a domain knowledge-enhanced depression detection approach. They constructed a depression-related lexicon tailored to Chinese language characteristics and extracted features such as depression-related term frequency and sentiment polarity. These features are fused using correlation-based weighting and fed into multiple machine learning models. [Peng et al. \(2019\)](#) presented a method based on multi-kernel support vector machines (SVM), extracting three types of features from user social media data and constructing an emotion dictionary; employing Term Frequency-Inverse Document Frequency (TF-IDF) to extract the word frequency statistics of Weibo text, all features are combined to classify whether users are depressed. [Zhang et al. \(2024\)](#) introduced a knowledge-aware deep learning model named DKDD (Deep Knowledge-aware Depression Detection). The model incorporates entity recognition, medical ontology alignment and attention mechanisms to extract clinically relevant entities and their temporal distribution from users’ digital traces on social media. [Tadesse et al. \(2019\)](#) detected the relevant factors of depressive emotions of online users based on Reddit data and utilized Linguistic Inquiry and Word Count (LIWC), Latent Dirichlet Allocation (LDA), N-gram and other classic machine learning algorithms in their model. [Deshpande and Rao \(2017\)](#), based on the Twitter platform, classified tweets as neutral or negative to detect depressive tendencies and employed Naive Bayes (NB) and Support Vector Machine as comparison classifiers. [Yoon et al. \(2022\)](#) constructed a multimodal depression dataset consisting of 951 video blogs (vlogs) and employed an audio–video cross-modal attention mechanism to capture inter-modal feature relationships. [Kumar and Venkatram \(2024\)](#), based on the Twitter dataset, used features including age, gender, the number of followers and online duration to predict suicidal behavior. This study proposed a rule-based classification algorithm, determining the optimal split point through the quicksort method and constructing a decision tree. [Islam et al. \(2018\)](#) utilized Facebook data, processed text data with LIWC and introduced K-Nearest Neighbors (KNN) for the depression classification of text. [Chen et al. \(2025\)](#) introduced a text-guided multimodal depression detection framework based on cross-modal feature reconstruction and disentanglement. Their model separates shared and latent representations from joint text–audio embeddings and employs a bidirectional cross attention module for enhanced cross-modal interaction. [Uddin et al. \(2019\)](#) adopted a gate recurrent unit (GRU) based method, creating several combinations of GRU and Dense layers and ultimately identifying a 5-layer GRU with 512 neurons as the optimal model. [Cai et al. \(2023\)](#) proposed a depression detection method based on multivariate time series features of user depressive symptoms. By constructing users’ emotional behavior sequences from social media posts, the method extracts dynamic features across multiple dimensions, including posting frequency, emotional fluctuation and depression-related keywords. These features are then input into classifiers for modeling. Experimental results demonstrate that this approach is effective in capturing temporal changes in users’ depressive states.

Although substantial progress has been made in leveraging social media data for depression detection, some limitations persist in current research. Firstly, many approaches rely on static

feature aggregation, lacking the capacity to capture users' emotional dynamics over time, which limits their ability to reflect the temporal progression of depressive symptoms. Secondly, although certain multimodal models are capable of capturing rich features, they often depend on costly or hard-to-obtain data types, such as user-generated audio and video, or users' facial expressions while browsing social media, which limits their applicability in large-scale or longitudinal studies. Besides, although domain knowledge has been incorporated into some models, its integration may be limited to rule-based lexicons or simple weighting schemes, lacking depth in semantic understanding and generalizability. Furthermore, some existing studies focus solely on final classification performance, overlooking the interpretability of users' emotional changes. These gaps call for a more temporally sensitive and visually interpretable framework for depression detection based on social media.

2.4 Novelty of this study

To address the aforementioned limitations, this study proposes a novel framework for depression detection on social media, integrating temporal modeling, weak supervision and visual interpretability. Specifically, we introduce a Multi-instance LSTM Learning Model that leverages an LSTM-based autoencoder to extract temporal features from users' social media posts without requiring post-level labels. By applying a weakly supervised learning strategy, we use user-level labels to guide the training of a binary classifier, which outputs the probability of depression for each individual post. This allows for fine-grained assessment of emotional tendencies over time. Furthermore, a MIL mechanism is employed to aggregate post-level predictions and determine the depression status of users, while a time series visualization module is designed to display the dynamic trajectory of users' depressive tendencies.

Compared to prior approaches, our framework not only captures longitudinal emotional patterns but also improves scalability and interpretability by avoiding reliance on expensive modalities or manual annotations at the post level. By combining deep temporal feature extraction, weak supervision and visual analytics, this study offers a cost-effective, scalable and insightful solution for large-scale depression screening based on social media data.

3. Methodology

Compared with the existing supervised learning models, which heavily rely on post-level annotations, our work proposes a novel depression detection framework, the Multi-instance LSTM Learning Model. Our model firstly extracts time series features from users' posts to assess the likelihood of depression at the post level. Then, following the principles of MIL, it predicts whether a user is at risk of depression at the user level and visualizes the temporal evolution of the user's depressive tendencies.

Unlike the previous methods, the most significant feature of our paper lies in its reliance only on coarse-grained user-level annotations, without the need for costly and highly subjective post-level annotations, which are also difficult to obtain in actual scenarios. Based on the core idea of MIL, all posts from a single user are treated as multiple instances, while the user is treated as a bag. The bag is labeled as depressed or normal, while the instances remain unlabeled. The model is trained completely on user-level labels, enabling depression detection under weak supervision.

Specifically, the proposed method comprises three key steps:

- (1) An unsupervised LSTM autoencoder is used to model each post and extract temporal features. This module retains the semantic and emotional information of the posts, enhancing the model's ability to capture behavioral traits associated with depression.
- (2) The post-level representations output by the LSTM are fed into a weakly supervised binary classifier to predict the probability of depression for each post. These probabilities are then aggregated under the MIL framework to infer the user's overall psychological state.

- (3) A visualization module is designed to plot the temporal dynamics of a user’s post-level depression probabilities, providing an intuitive representation of their mental state fluctuations over time.

Overall, our method achieves a balance between low supervision requirements, strong representational capacity and high interpretability, outperforming several existing baseline models in the comparative experiments.

3.1 Overall structure of our model

Firstly, let’s define our research problem: Suppose there are $N = N_d + N_u$ labeled tweets data, encompassing N_d tweet data $\mathbf{TD} = \{td_1, td_2, \dots, td_{N_d}\}$ and label data $\mathbf{YD} = \{yd_1, yd_2, \dots, yd_{N_d}\}$ from P_d users with depressive tendencies, as well as N_u tweet data $\mathbf{TU} = \{tu_1, tu_2, \dots, tu_{N_u}\}$ and label data $\mathbf{YU} = \{yu_1, yu_2, \dots, yu_{N_u}\}$ from P_u users without depressive tendencies, where $\mathbf{T} = \mathbf{TD} \cup \mathbf{TU} = \{t_1, t_2, \dots, t_N\}$, $\mathbf{Y} = \mathbf{YD} \cup \mathbf{YU} = \{y_1, y_2, \dots, y_N\}$. It is anticipated to extract time series features $\mathbf{SD} = \{sd_1, sd_2, \dots, sd_{N_d}\}$ and $\mathbf{SU} = \{su_1, su_2, \dots, su_{N_u}\}$ from \mathbf{TD} and \mathbf{TU} respectively by employing unsupervised approaches, where $\mathbf{S} = \mathbf{SD} \cup \mathbf{SU} = \{s_1, s_2, \dots, s_N\}$, and train a binary classifier F with these data for the classification and detection of depression in other relevant individuals in the future.

The structure of our model is illustrated in Figure 1. First, based on the timeline tweets \mathbf{T} of a given user, it is tokenized into available tweet encodings \mathbf{X} and then, using unsupervised LSTM, encoding each user’s tweets as time series features \mathbf{S} ; Then, the trained binary classifier is used to classify time series features. Then, multiple instance learning is used to detect whether a specific tweet of a given user is depressed, and each tweet is traversed. Finally, we output a plot of the user’s tendency per tweet over time.

3.2 Features extraction

To train the MIL model, time series features are used as input, which are based on the timeline tweets of each user. Time series data refers to a sequence of data points indexed in time order, so unsupervised LSTM can be used as a method to extract time series features from each tweet. In this case, to represent the time series information in the tweets of the user’s timeline, a set of vectors s_i is extracted from each tweet using an unsupervised LSTM (also known as an LSTM autoencoder). The autoencoder is a reconstructive neural network that learns a vectorized representation of each tweet from the user in an unsupervised manner.

The LSTM autoencoder requires at least two LSTM layers. Taking two LSTM layer models as an example, the first layer can be regarded as the encoder and the second layer can be

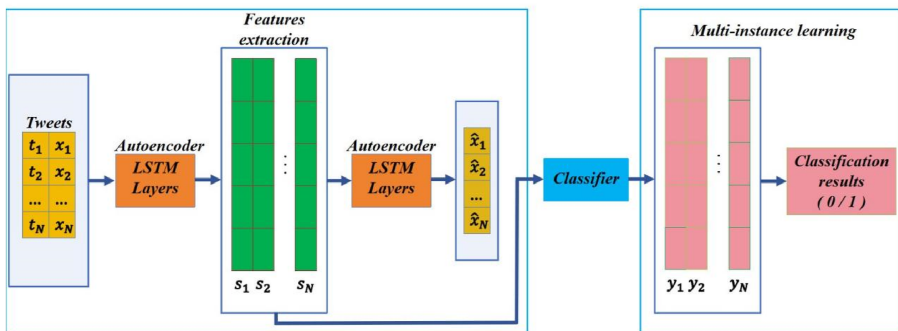


Figure 1. Overall structure of our model. Source(s): Authors’ own creation/work

regarded as the decoder. Thus, the LSTM autoencoder compresses the tokenized tweets x_i from the input layer into a well-formed code s_i , which is then decompressed into the form of vector \hat{x}_i .

The steps of the LSTM autoencoder are as follows:

- (1) Tokenization of tweets. Since posted tweets are a kind of string, the first step of the autoencoder is to tokenize the tweet before entering it. To ensure that all tweets have the same dimensions, a word embedding method is used. We first calculate the maximum number of words θ of a tweet and then pad $\mathbf{0}$ in front of the tweet vector. In this way, the tweet $T = \{t_1, t_2, \dots, t_N\}$ can be labeled as $X = \{x_1, x_2, \dots, x_N\}$.
- (2) LSTM autoencoder. The input layer sequentially receives a tokenized text sequence x_i , which is passed to the LSTM layer for processing and encoded into a sequence representation s_i . Then, s_i is decoded into \hat{x}_i by the next LSTM layer. Among them, the LSTM autoencoder layer aims to generate a matrix that has the same shape as the input tokenized tweets. Therefore, the difference between the input and output can be defined as a loss function, as follows:

$$loss = \sum_{i=1}^N \|x_i - \hat{x}_i\|_2 \quad (1)$$

In order to make the input similar to the output, this autoencoder strives to minimize the loss function.

3.3 Training of the classifier

Alternatively, depression detection is typically formulated as a binary classification task, in which models are trained to estimate the likelihood that a user is depressed. In this study, we employ two commonly used binary classification algorithms—support vector machines (SVM) and Logistic Regression (LR) to train depression detection models.

- (1) Support Vector Machine (SVM). In this study, two SVM models with different kernel functions are trained. According to the Representer theorem, the parameters in SVM can be written as a linear combination of the training data, so the final classifier can be expressed as follows:

$$f(s_i) = \sum_{j=1}^N \alpha_j \gamma_j \varphi(s_j)^T \varphi(s_i) + b \quad (2)$$

Here, $\varphi(s_j)^T \varphi(s_i)$ is the kernel function, which can also be expressed as $K(s_i, s_j)$.

Linear kernel (Linear-Support Vector Machine (L-SVM)) and RBF kernel (Radial Basis Function-Support Vector Machine (R-SVM)) are employed. Here, the kernel function of the L-SVM can be expressed as follows:

$$K(s_i, s_j) = s_i^T s_j \quad (3)$$

In addition, the kernel function of R-SVM can be expressed as follows:

$$K(s_i, s_j) = \exp\left(-\frac{\|s_i - s_j\|_2}{2\sigma^2}\right) \quad (4)$$

(2) LR. The LR classifier $f(s_i)$ can be defined as follows:

$$f(s_i) = \frac{1}{1 + e^{-(\mathbf{w}^T s_i + b)}} \quad (5)$$

Here, \mathbf{W} and \mathbf{b} are the parameters to be learned.

To measure the learning loss of the LR-based binary classifier, we adopt the cross-entropy loss function, which is defined as follows:

$$\text{loss}(f(s_i), y_i) = \frac{1}{N} \sum_{i=1}^N [y_i \lg f(s_i) + (1 - y_i) \lg (1 - f(s_i))] \quad (6)$$

Here, y_i is the true label corresponding to s_i .

To train the classifier, we use the gradient descent algorithm to minimize the loss function.

3.4 Multi-instance learning

After training the model to detect depressive tweets, the next step is to predict whether a Weibo user is experiencing depression. To address the issue of dataset imbalance, we adopt a MIL framework as the user-level detection model. The structure of the proposed algorithm is described as follows:

- (1) After self-encoding, the time series features of user u_i are extracted as $\mathbf{S}_i = \{s_{i1}, s_{i2}, \dots, s_{ij}, \dots, s_{in}\}$, where n is the number of tweets of the user. The selected $f(s_i)$ is used to test each feature, and the output result is represented as $\hat{\mathbf{X}}_i = \{\hat{x}_{i1}, \hat{x}_{i2}, \dots, \hat{x}_{ij}, \dots, \hat{x}_{in}\}$, $j = 1, 2, 3, \dots, n$.
- (2) Next, traversing the elements of $\hat{\mathbf{X}}_i = \{\hat{x}_{i1}, \hat{x}_{i2}, \dots, \hat{x}_{ij}, \dots, \hat{x}_{in}\}$. Define a function $T(\hat{x}_{ij})$ and introduce an external threshold parameter p . If a specific $\hat{x}_{ij} > p$, then $T(\hat{x}_{ij})$ equals 1; otherwise, if $\hat{x}_{ij} \leq p$, then $T(\hat{x}_{ij})$ equals 0. Finally, sum up the n numbers of $T(\hat{x}_{ij})$ values as l .

$$T(\hat{x}_{ij}) = \begin{cases} 1, & \hat{x}_{ij} > p \\ 0, & \hat{x}_{ij} \leq p \end{cases}, l = \sum_{j=1}^n T(\hat{x}_{ij}) \quad (7)$$

- (3) Finally, classifying the user u_i as depressed or non-depressed. An exogenous weight parameter ω is then added. If the weight of the predicted posts by the user that are deemed depressive exceeds the weight of all posts across all time axes, i.e. $l > \omega \times n$, the user is classified as depressed. Conversely, if $l \leq \omega \times n$, the user is classified as non-depressed. Thus, the final result of the algorithm can be defined as follows:

$$\hat{\mathbf{Y}}_i = \begin{cases} 1, & l > \omega \times n \\ 0, & l \leq \omega \times n \end{cases} \quad (8)$$

Above $\hat{\mathbf{Y}}_i = 1$ means that the user u_i is classified as depressed; $\hat{\mathbf{Y}}_i = 0$ means that the user u_i is classified as non-depressed.

4. Experimental results

4.1 Description of dataset

To validate the proposed method, we utilize a dataset named WU3D (Wang et al., 2020), which is publicly available. WU3D collects and labels a large number of Weibo posts from both depressed and normal users. All the users have their user IDs hidden, achieving anonymity.

Each user sample includes their nickname, posts, posting time, gender and other details. The overall statistics are presented in [Table 1](#). We select a subset of 2000 users for the experiment, including 400 depressed users and 1,600 normal users. Each time, they are divided into a training set and a test set according to an 8:2 ratio randomly.

4.2 Visual results of this model

In MIL, we extract the time series features of each user’s tweets and detect each feature. The time series feature of the user u_i is $S_i = \{s_{i1}, s_{i2}, \dots, s_{ij}, \dots, s_{in}\}$, and the output is $\hat{X}_i = \{\hat{x}_{i1}, \hat{x}_{i2}, \dots, \hat{x}_{ij}, \dots, \hat{x}_{in}\}$. Let’s assume there are m users in total. After completing the detection of each user in turn, we find out $\max(\hat{x}_{ij})$ and $\min(\hat{x}_{ij})$, $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ in them. Next, we normalize each \hat{x}_{ij} to $\widetilde{\hat{x}}_{ij}$, as shown in [equation \(9\)](#).

$$\widetilde{\hat{x}}_{ij} = \frac{\hat{x}_{ij} - \min(\hat{x}_{ij})}{\max(\hat{x}_{ij}) - \min(\hat{x}_{ij})} \quad (9)$$

Then we obtain the normalized vector. The normalized vector of the user u_i is expressed as $\widetilde{\hat{X}}_i = \{\widetilde{\hat{x}}_{i1}, \widetilde{\hat{x}}_{i2}, \dots, \widetilde{\hat{x}}_{ij}, \dots, \widetilde{\hat{x}}_{in}\}$, $j = 1, 2, \dots, n$. $\forall \widetilde{\hat{x}}_{ij} \in (0, 1)$. Finally, the values of each element of the output vector are printed in the images.

We conduct a visualization-based analysis and comparison of depression probability scores over time for a large number of users labeled as either depressed or non-depressed. The results reveal that depressed users exhibit greater fluctuations in predicted probabilities, indicating more unstable depressive symptoms over time. In contrast, the depression probabilities of normal users remain relatively stable, with most values falling below 0.3. These findings are consistent with previous research by [Cai et al. \(2023\)](#) and [Seabrook et al. \(2018\)](#).

To facilitate demonstration, we randomly select one depressed user and one non-depressed user from the dataset and visualize their depression probability scores over time. The specific tweet contents and their posting timestamps are shown in [Tables 2](#) and [3](#), respectively, while the visualization results are presented in [Figures 2](#) and [3](#).

For the depressed user, there are 10 tweet records spanning from May 8, 2019 to June 12, 2019. The depression probability scores range from 0.22 to 0.97, resulting in a difference of 0.75, which indicates a high degree of fluctuation and a sharper changing trend over time. For the non-depressed user, there are 18 tweet records spanning from September 18, 2019 to March 21, 2020, with most predicted scores remaining below 0.2. The overall variation is smaller, and the probability trend appears more stable and consistent.

Furthermore, an examination of the relationship between tweet content and the model’s predicted depression probability reveals a clear correspondence. Tweets containing explicit expressions of negative emotions or psychological distress tend to receive higher prediction scores. For example, the tweet “最后最后 我还是选择了药物治疗 (In the end, I chose medication)” is assigned a score of 0.97, suggesting a strong association with depressive states. In contrast, tweets that describe daily routines, emotionally neutral, or positive experiences generally receive lower depression probabilities. For instance, the tweet “今日份晚餐

Table 1. Description of dataset

Labels	Number of users	Total tweets	Average number of tweets per user
Normal	22,245	1,564,349	70.32
Depression	10,325	372,377	36.07
Total	32,570	1,936,726	59.46

Source(s): Authors’ own creation/work

Table 2. The contents of a user labeled as depressed

Order number	The time the tweet was posted	Content of the tweet (The original text is in Chinese, and the English translation is in parentheses)	Normalized predicted values \hat{x}
1	2019-06-12 12:25:43	狗咬你一口 你会选择怎么办? (What would you do if a dog bit you?)	0.22
2	2019-06-09 00:45:04	最后最后 我还是选择了药物治疗 (In the end, I chose medication)	0.97
3	2019-06-04 03:16:27	本以为回家会好的 可是在家里越来越严重 我真的不知道是自己心里固执的抗争 还是身体的毛病 是真的整夜整夜的失眠啊 失眠打卡四十三天 中卫 (I thought I would be better at home, but it got worse and worse. I really don't know whether it is my stubborn resistance or my physical problem. I really can't sleep all night. I have been insomnia for 43 days in Zhongwei)	0.90
4	2019-05-21 19:36:30	小人得势 他呢就悲贱的像条狗 你不知道他跳起来摇尾巴的时候有多乖 兄弟 我劝你做个人 生而为人 劝你善良 (When a man is in power, he is as miserable as a dog. You don't know how good he is when he jumps up and wag his tail. Brothers, I advise you to be a man. Being human, I advise you to be kind)	0.35
5	2019-05-21 18:50:26	又是一夜无眠 还不如早起看太阳 (Another sleepless night. Better get up and see the sun)	0.71
6	2019-05-15 23:50:41	其实没有人影响你的情绪是你自己放不过你自己 (In fact, the only person who can't affect your mood is yourself)	0.76
7	2019-05-09 01:24:53	十多岁的孩子承受了那么多 (Teenagers have endured so much)	0.63
8	2019-05-08 01:01:56	不是因为患有抑郁症才会导致精力很弱 是因为成天失眠 抑郁症患者的精力很弱吗 (It's not because of depression that you have low energy. It's because you're losing sleep all day. Do people with depression have low energy)	0.94
9	2019-05-08 01:01:03	让自己忙起来 抑郁症如何走出心理阴影 (Get yourself busy. How to get over the psychological shadow of depression)	0.96
10	2019-05-08 00:55:23	我真的希望在与抑郁症抗争的日子里 看得到那种从心底里散发出来的阳光 * 而不再是所有的阴暗 同时不再与失眠继续斗争 不再惧怕黑夜的来临 (I really hope that in the days of fighting against depression I can see the sunshine from the bottom of my heart * instead of all the darkness and no longer struggle with insomnia and fear the coming of night)	0.93

Source(s): Authors' own creation/work

(Today's dinner)" from a non-depressed user is assigned a score of only 0.08. This suggests that the model is capable of capturing emotional cues embedded in the language, thereby associating specific linguistic features with the likelihood of depression.

5. Evaluation of the model

In order to evaluate the effect of the proposed model, we compare the proposed model with the traditional method. And four evaluation metrics, including accuracy, precision, recall and F1-score, are set.

5.1 Contrast algorithms

To verify the performance of the proposed deep learning method, three other methods are compared. They are described as follows.

- (1) NB. In machine learning, NB is a simple probabilistic classifier algorithm based on Bayes' theorem. Its idea is based on classifying the given items, solving the

Table 3. The contents of a user labeled as non-depressed

Order number	The time the tweet was posted	Content of the tweet (The original text is in Chinese, and the English translation is in parentheses)	Normalized predicted values \hat{x}
1	2019-9-18 22:45	最喜欢医院每天下午的加餐了 (I love the extra meals at the hospital every afternoon the most)	0.64
2	2019-9-24 22:49	中午的饭忘了拍 不是很满意今天的饭 点了两顿外卖 但是南瓜汤真的是好好喝 明天正好要来 不过关系不大本来打算看阅兵 提前煮好了明日份奶茶 加点龟苓膏就可以愉快的吃吃喝喝啦 (I forgot to take a picture of lunch. I'm not very satisfied with today's meal. I ordered two takeaway meals. But pumpkin soup was really good. It was just coming tomorrow, but it didn't matter much. I was going to watch the parade. I cooked milk tea for tomorrow and added Gulinggao. Then I could eat and drink happily)	0.23
3	2019-9-25 23:11	这是昨晚的加餐 (This is an additional meal from last night)	0.05
4	2019-9-27 8:00	早餐照样是这么简单 午餐点了老汤烧鸭面 幸好自己又准备了份水煮菜 不然今日份蔬菜都没吃够 现在晚上天天都顶胃 晚上就梨口奶茶解决了 奶茶里加了冰糖 (Breakfast is still as simple as that. Lunch was ordered with old soup and duck noodles. Fortunately, I prepared a boiled dish myself. Otherwise, I wouldn't have eaten enough vegetables today. Now, every night it's overwhelming, and at night it's just pear □ Milk tea solved the problem by adding rock sugar to it)	0.08
5	2019-9-28 23:19	月子中心的晚餐和加餐 完全有理由怀疑会回到生前体重 (The dinner and additional meals at the confinement center have every reason to suspect that they will return to their previous weight)	0.11
6	2019-9-29 22:51	以前和我爸只有有事的时候才聊天 有了小汤圆之后就 哈哈 (I used to chat with my father only when I had something to do. But after I had my baby xiaotangyuan, it changed hahaha)	0.09
7	2019-9-30 22:08	先记录下昨天 昨天太累了 准备下午空腹四小时抽血 所以早上吃早点 抽完血吃了两个贝果 晚上点的牛肉炒饭料很多 可是好吃是需要付出代价的 然后就躺下睡了 (First, I recorded yesterday that I was too tired. I was ready to take blood on an empty stomach for four hours in the afternoon. So I ate well in the morning, took blood, ate two bagels, and ordered a lot of beef Fried Rice in the evening. But the food was delicious and there was a price to pay. Then I lay down and slept)	0.44
8	2019-10-1 19:59	无痛真的是人类之光 (Pain free is truly the light of humanity)	0.25
9	2019-10-2 13:46	头发洗了真舒服 (It is really comfortable to wash my hair)	0.06
10	2019-10-3 18:04	酸汤肥牛底下还有你家的豆皮 (Under the sour soup and fat cow, there is still tofu skin)	0.09
11	2019-10-4 20:17	老王买的橙子真是酸倒我了 还不如贝果甜 贝果还是喜欢放平底锅煎一下表面脆脆的 午餐晚餐都超满足的 晚饭的乌冬面只放了一半蔬菜 剩下的可以明天吃 上午去剪了头发 不忍心剪太多 这样过年还可以烫一下 (The oranges that Lao Wang bought are so sour that they're not as sweet as bagels. Berries prefer to be pan fried and have a crispy surface, which satisfies both lunch and dinner. We only put half of the vegetables in the udon noodles for dinner, and the rest can be eaten tomorrow morning. I went to cut my hair and couldn't bear to cut too much. This way, during the Chinese New Year, I can also perm them)	0.21

(continued)

Table 3. Continued

Order number	The time the tweet was posted	Content of the tweet (The original text is in Chinese, and the English translation is in parentheses)	Normalized predicted values \hat{x}
12	2019-10-5 19:55	快要十月了明天该写个总结了 每天早上最没食欲所以能吃的很清淡 早点把事情做完的感觉真好 台风又要来了 不会在温岭登陆了吧 老王担心不会在这两天生吧哈哈 anyway明天去囤点水 (It's almost October, and tomorrow I should write a summary. Every morning, I have the least appetite, so I eat very lightly. It feels great to finish things early. The typhoon is coming again, and it won't make landfall in Wenling, right? Lao Wang is worried that I will give birth to our baby in these two days, hahaha. Anyway, I'll go stock up on some water tomorrow)	0.32
13	2019-10-6 20:51	今日份晚餐 (Today's dinner)	0.08
14	2019-10-7 20:54	见红了 医生说宫缩规律开一指了 都还没做好心理准备筋膜训练该了 (Seeing the blood, the doctor said that the uterine contractions were regular. I'm not even mentally prepared yet. It's time for fascia training)	0.35
15	2019-10-9 21:23	恭喜发财鸭 (May you be happy and prosperous!)	0.10
16	2019-10-9 21:24	刚吃完最后一个贝果新的一批又到啦 午餐很满足 本来打算一个人吃的老王妹妹临时说回来 连忙加了牛筋丸和荷包蛋 猪肉就喜欢这种原始的味道 晚餐吃了两个香梨感觉刚刚好 不顶胃但是现在又饿了 吃个今天刚到的红豆碱水球 嗯还会再回购的 (Just finished the last bagel, a new batch has arrived. Lunch was very satisfying, originally planned to be eaten alone. Lao Wang's sister came back at the last minute, so I quickly added beef tendon pills and poached eggs. I just like the primitive taste of pork. I had two fragrant pears for dinner and felt just right, not too full, but now I'm hungry again. Have a red bean alkali water ball that just arrived today, and I will repurchase it again)	0.18
17	2019-10-11 20:27	感觉练瑜伽有点拖延症了 学习倒是一直在进程中 碱水结感觉一般啊 还是贝果比较yummy ps现在觉得蔬菜最好烧了 (I feel like practicing yoga is a bit procrastinating, but learning is always in progress. I feel like Pretzel is just so-so, but bagels are more yummy. PS: now I think vegetables are best cooked)	0.21
18	2020-3-21 23:04	hello 小汤圆 (hello my baby xiaotangyuan)	0.09

Source(s): Authors' own creation/work

probability of each category under the condition, considering the maximum value in the given item and its basic definition is as follows:

$$\hat{y} = \arg_k \max p(C_k) \prod_{i=1}^n p(x_i | C_k) \quad (10)$$

There are three types of NB, which are Gaussian NB, polynomial NB and Bernoulli NB. Since Bernoulli NB is suitable for the classification of short texts, this event model is chosen for comparison.

- (2) Random Decision Forest (RDF). In machine learning, a RDF is a classifier containing multiple decision trees whose output class is determined by the pattern number (classification) or mean prediction (regression) of the output class of each tree-like

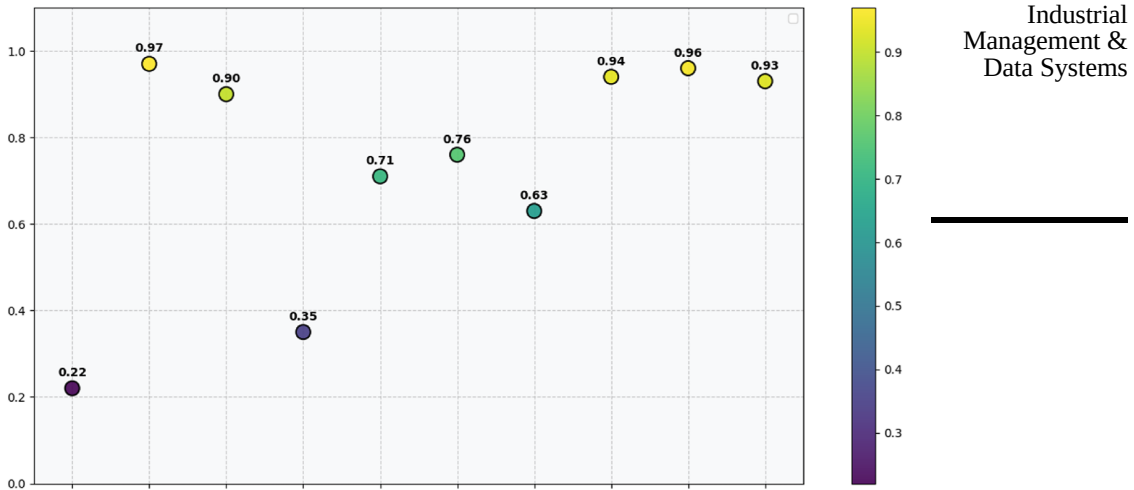


Figure 2. The visualization result of a user labeled as depressed. **Source(s):** Authors' own creation/work

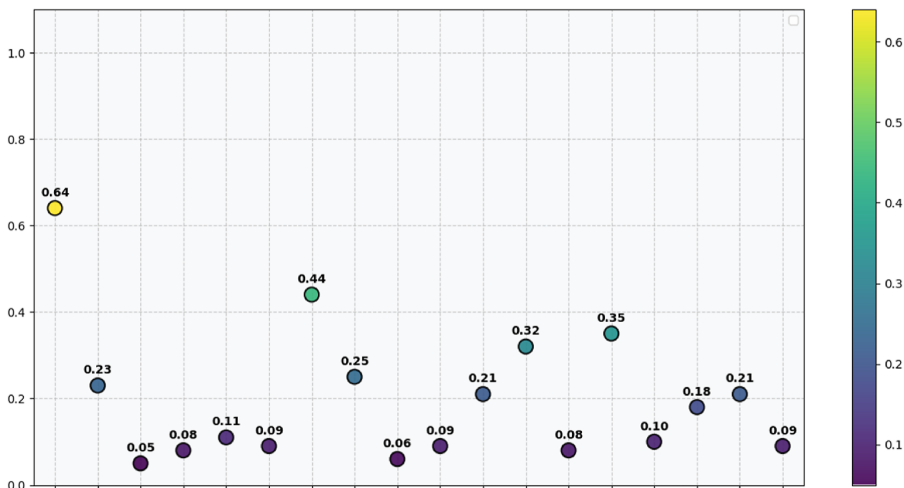


Figure 3. The visualization result of a user labeled as non-depressed. **Source(s):** Authors' own creation/work

structure. Such algorithms are widely used for classification, regression and other tasks.

- (3) Bidirectional Encoder Representations from Transformers (BERT). It's a pre-training language model based on the Transformer architecture. Self-attention mechanism is its core advantage, which allows the model to focus on all positions in the sequence, greatly improving efficiency and modeling ability. The principle of the self-attention mechanism is introduced as follows.

For the input sequence $X = \{x_1, x_2, \dots, x_n\}$, first calculate each element of it by linear transformation:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (11)$$

Among them, W_Q , W_K and W_V are the network parameters learned.

Next, calculate the attention score of the element x_i relative to x_j :

$$S_{ij} = \frac{Q_i K_j}{\sqrt{d_k}} \quad (12)$$

Then we get the weight of the element x_i relative to x_j :

$$A_{ij} = \text{softmax}(S_{ij}) = \frac{\exp(S_{ij})}{\sum_{k=1}^n \exp(S_{ik})} \quad (13)$$

Further output the attention score of element x_i :

$$o_i = \sum_{j=1}^n A_{ij} V_j \quad (14)$$

Finally, the attention score vector of each element in the output sequence is:

$$O_i = \{o_1, o_2, \dots, o_n\} \quad (15)$$

This mechanism enables BERT to efficiently capture the global context. For example, in the sentence “She borrowed a book from the library and went to the coffee shop for a cup of coffee”, BERT accurately understands the sequence of “her” behavior by calculating the attention weights of “she” and “library” and “coffee shop”. Other models may struggle to associate “library” with “cafe”.

5.2 Evaluation indices

In the field of binary classification, data statistics mainly adopt four indicators:

TP (True Positive), which means the actual value is depressed, and the predicted value is also depressed;

TN (True Negative), which means the actual value is normal, and the predicted value is also normal;

FP (False Positive), which means the actual value is normal, but the predicted value is depressed;

FN (False Negative), which means the actual value is depressed, but the predicted value is normal.

With the help of the above four statistical data, the recognition performance of comparative methods and comparative features is evaluated by comparing accuracy, recall, precision and F1 score and the specific description of them is as follows:

- (1) Accuracy: This is the ratio of correctly predicted values to the total value. Accuracy is almost certainly the most intuitive performance metric, and there is:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (16)$$

- (2) Recall: Also known as sensitivity, is the ratio of correctly predicted positive values to all values in the actual value, and there is:

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (17)$$

- (3) Precision: Also known as positive predictive value, it is the ratio of the positive number of correct predictions to the total number of positive predictions, and there is:

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (18)$$

- (4) F1 score: This is a weighted average of precision and recall. It is usually expressed as the harmonized average of precision and recall, and there is:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \times 100\% \quad (19)$$

5.3 Results of comparing experimental

To optimize model performance, we compare our approach with several baseline algorithms on the same dataset, evaluating performance across different classifiers, hyperparameter settings, feature dimensions and learning rates.

All experiments are conducted on a Windows 11 (64-bit) operating system using a NVIDIA GeForce RTX 3060 graphics card. The software configuration is Python 3.13, PyTorch 2.7.1, scikit-learn 1.6.0, numpy 2.2.1 and pandas 2.2.3.

Unless otherwise specified, the default experimental settings are as follows: R-SVM classifier, hyperparameters $p = 0.20$, $\omega = 0.50$, learning rate = 0.06, feature-dimension = 128 and training epochs = 100. In each comparative experiment, only one variable is modified at a time while all other parameters are held constant.

5.3.1 Comparison of different models. Firstly, we conduct a series of experiments to verify the performance of the proposed MIL model. Prior to training, all extracted features were standardized to the same dimension. The features used in these experiments include social network features, sentiment features, topic features and domain features, each of which can be treated as a distinct modality. Given the nature of the proposed method, only time-series features are used as input for depression prediction. Table 4 presents a performance comparison between the proposed method and three baseline methods. As shown, the proposed model achieves the highest accuracy and precision scores.

What accounts for the performance differences among these models? We argue that traditional classifiers, such as NB and Random Forest, perform poorly due to their reliance on surface-level lexical features and their limited ability to model contextual semantics—particularly in short, noisy social media texts. These methods often fail to capture implicit emotional expressions and dynamic user behavior. In contrast, deep learning-based models demonstrate stronger performance by learning semantic representations from data. The

Table 4. Performance of different models

Models	Accuracy/%	Precision/%	Recall/%	F1_Score/%
NB	81.9	80.1	83.2	81.6
RDF	89.0	73.7	70.0	71.8
BERT	86.5	85.0	<i>88.4</i>	<i>86.7</i>
Our model	<i>95.0</i>	<i>94.4</i>	79.5	86.3

Note(s): Italic fonts indicate that a certain model, method or setting has achieved the highest score in a specific metric compared to others. The following tables are the same

Source(s): Authors' own creation/work

proposed model further improves on this by leveraging MIL to aggregate tweet-level predictions into user-level inferences, and by incorporating temporal dynamics to reflect fluctuations in depressive symptoms over time. This design allows the model to more accurately identify depressed users, especially those whose expressions of distress are intermittent or indirect. Additionally, experimental results show that BERT achieves higher recall and F1 scores, validating its strength in extracting rich semantic features from user-generated content. However, the proposed model outperforms BERT in accuracy and precision (with comparable F1 scores), due to its ability to integrate user-level context and model temporal patterns. While BERT captures linguistic context at the individual tweet level, our model synthesizes users' tweet histories to detect consistent behavioral trends over time.

5.3.2 Comparison between different classifiers. Table 5 presents the performance comparison among three classifiers. All classifiers achieve relatively strong results. However,

Table 5. Performance with different classifiers

Classifiers	Accuracy/%	Precision/%	Recall/%	F1_Score/%
LR	94.6	94.1	79.1	86.1
L-SVM	94.8	94.2	79.3	86.1
R-SVM	95.0	94.4	79.5	86.3

Note(s): Italic fonts indicate that a certain model, method or setting has achieved the highest score in a specific metric compared to others

Source(s): Authors' own creation/work

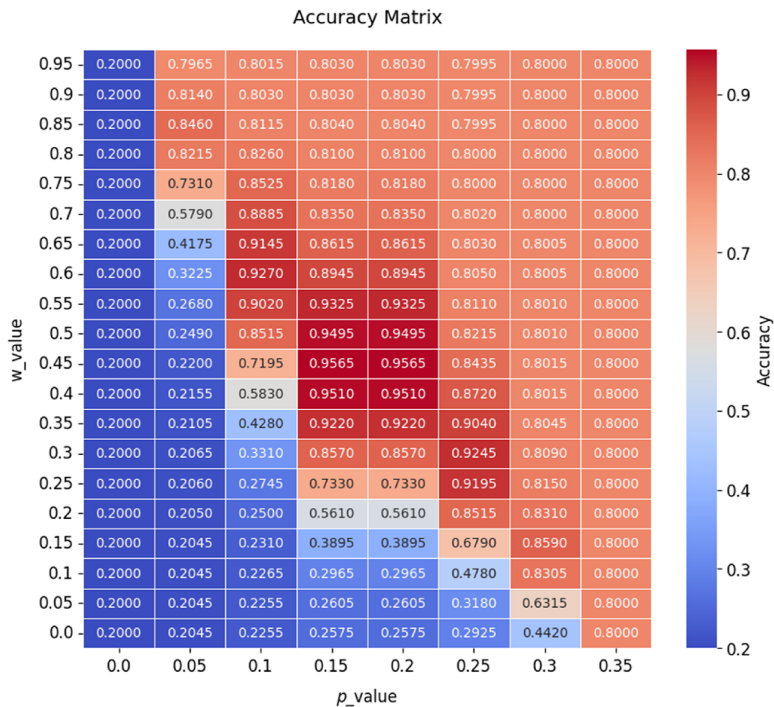


Figure 4. Accuracy scores under different combinations of hyperparameters. **Source(s):** Authors' own creation/work

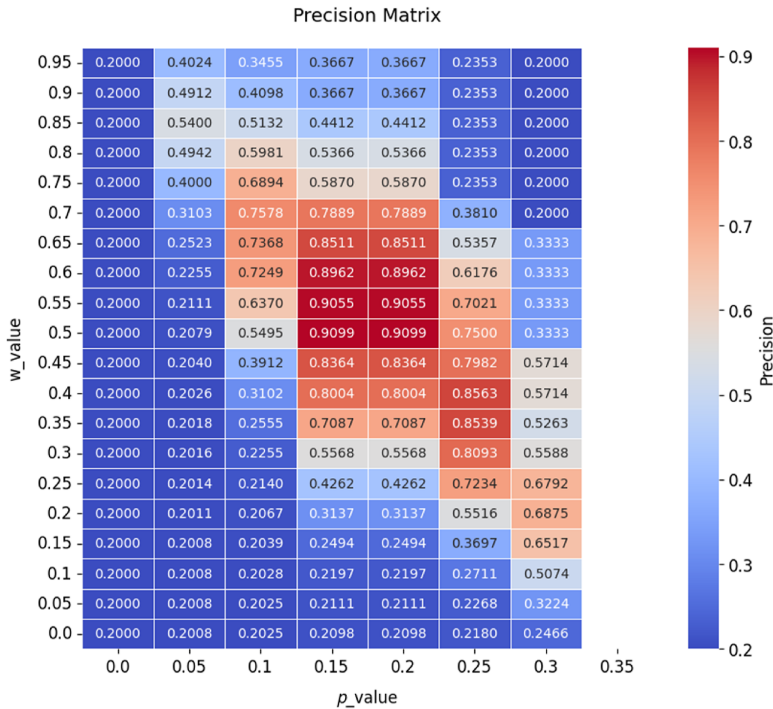


Figure 5. Precision scores under different combinations of hyperparameters. **Source(s):** Authors' own creation/work

the SVM with an RBF kernel consistently outperforms the others across all four evaluation metrics. These results indicate that unsupervised LSTM features are effective for depressive tweet detection. The superior performance of the Radial Basis Function-Support Vector Machines (RBF-SSVM) suggests that the extracted time-series features may be non-linearly separable, making the RBF kernel more suitable than a linear kernel in scenarios with moderate data sizes.

5.3.3 Comparison of different hyperparameters. The hyperparameters p and ω in the model are exogenously specified. To identify optimal values, we adopt the grid search method following the approach of [Jahnavi et al. \(2023\)](#). This method systematically explores all combinations of predefined hyperparameter values, evaluates the model performance for each combination and selects the configuration that yields the best results.

To define the search range of the hyperparameter p , we first examine the values \hat{x}_{ij} obtained from the MIL output. The observed minimum is close to zero, and the maximum is approximately 0.332. Therefore, it is reasonable to restrict the search interval for p to $[0, 0.35]$. In contrast, the optimal range for the hyperparameter ω is dependent on the value of p and cannot be determined analytically. To ensure robustness, we conduct a thorough evaluation across a broad range of ω values.

Based on the design of this model, we hypothesize that reducing the values of p and ω increases the chance of users being identified as depressed under the same conditions. This can improve recall rates but may reduce precision; the reverse may also be true. In other words, the recall rate is negatively correlated with the value of p and ω ; the accuracy is positively correlated with the value of p and ω . (The following experiments have confirmed this hypothesis.)

[Figures 4–7](#) show the four kinds of scores under different hyperparameters p and ω . Blank cells in [Figures 5 and 7](#) correspond to scenarios where an excessively high value of p leads the

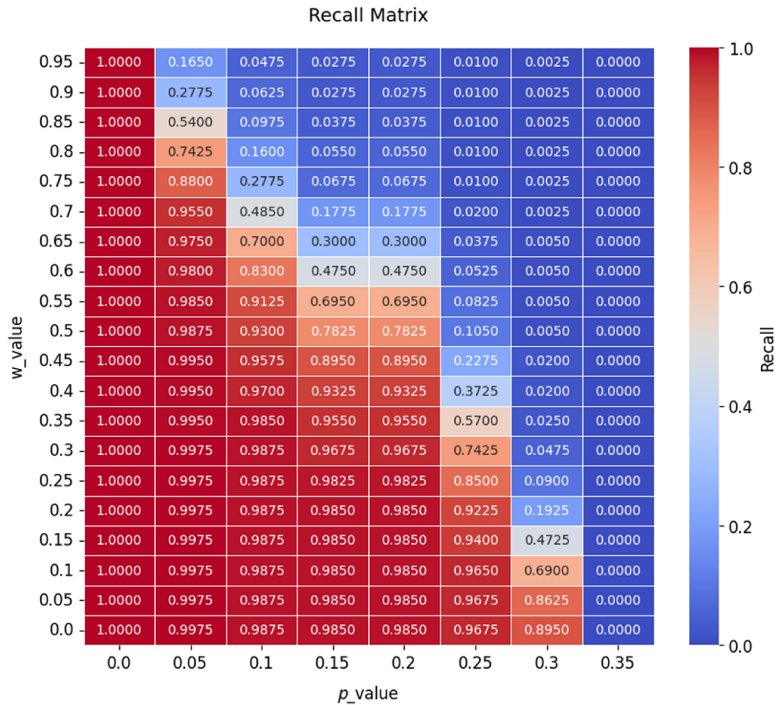


Figure 6. Recall scores under different combinations of hyperparameters. **Source(s):** Authors’ own creation/work

model to classify all users as non-depressed. In such cases, precision and F1-score cannot be computed, and recall drops to zero. Nevertheless, since 80% of users in the dataset are labeled as non-depressed, the resulting accuracy remains at 0.8, despite the model’s poor discriminative performance.

5.3.4 Comparison of different learning rates. As a critical hyperparameter in training machine learning models, the learning rate determines the step size for parameter updates during optimization. An appropriate learning rate can accelerate convergence and improve model performance, while a rate that is too high or too low may result in poor training outcomes. In the above experiments, the learning rate is set to 0.06. To further optimize model performance, we conduct additional experiments with different learning rates. [Table 6](#) summarizes the model’s performance under different learning rate settings.

5.3.5 Comparison of different feature dimensions. In previous experiments, we used a 128-dimensional LSTM hidden layer. To identify the optimal feature dimension, we conduct a series of experiments with different dimensional settings. A feature dimension that is too low may limit the amount of representational information available to the model, hindering its ability to capture important patterns and potentially resulting in underfitting. Conversely, excessively high dimensions can increase computational complexity, reduce generalization and lead to overfitting.

[Table 7](#) shows the scores of the model across four dimensions: {16, 32, 64, 128}. When increasing the dimension to 256, the model fails to converge, as indicated by stagnant loss values. The results suggest that the performance improves as the dimension increases, with the best accuracy and precision achieved at 128 dimensions.

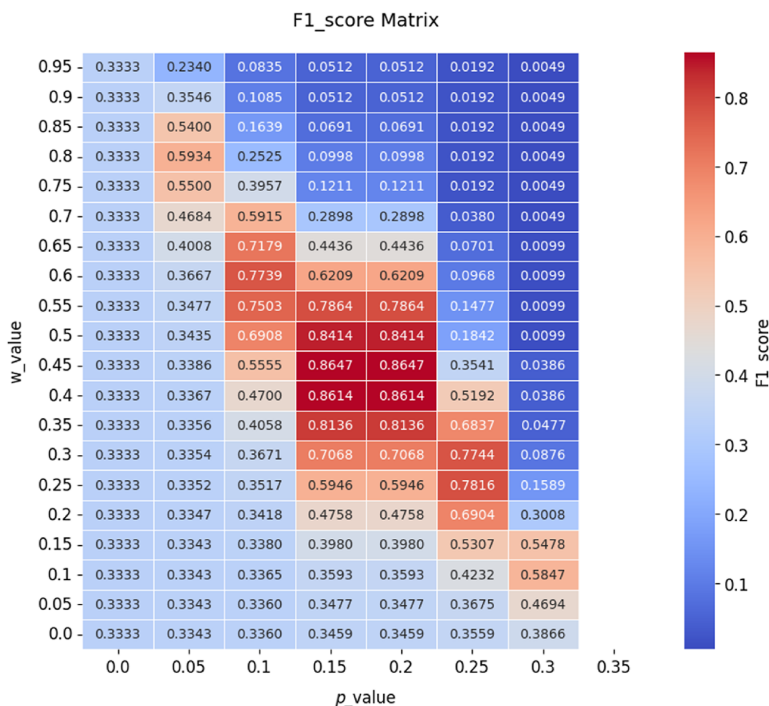


Figure 7. F1_scores under different combinations of hyperparameters. Source(s): Authors' own creation/work

Table 6. Scores at different learning rates

Learning rate	Accuracy/%	Precision/%	Recall/%	F1_Score/%
0.04	87.9	88.2	87.5	87.9
0.05	90.3	90.0	89.5	90.3
0.06	95.0	94.4	79.5	86.3
0.07	92.0	91.5	89.7	90.6
0.08	91.5	91.0	90.0	90.5
0.09	88.5	88.0	87.0	87.8

Note(s): *Italic fonts* indicate that a certain model, method or setting has achieved the highest score in a specific metric compared to others

Source(s): Authors' own creation/work

5.3.6 Comparison of different training epochs. In previous experiments, the number of training epochs is initially set to 100. Insufficient training epochs may prevent the model from adequately learning periodic and trend patterns in the data, particularly when long-term dependencies are critical for accurate prediction. Conversely, an excessive number of epochs increases the risk of overfitting, especially when the training data is limited, causing the model to learn noise or sample-specific artifacts, thereby reducing generalization. To identify the optimal training duration, we evaluate the model's performance across various epoch settings. The results are summarized in Table 8.

Table 7. Scores under different feature dimensions

Feature dimensions	Accuracy/%	Precision/%	Recall/%	F1_Score/%
16	82.5	79.8	84.2	81.9
32	87.3	84.5	<i>89.1</i>	86.7
64	91.8	93.5	88.2	<i>90.8</i>
128	<i>95.0</i>	<i>94.4</i>	79.5	86.3

Note(s): Italic fonts indicate that a certain model, method or setting has achieved the highest score in a specific metric compared to others

Source(s): Authors' own creation/work

Table 8. Scores under different training epochs

Training epochs	Accuracy/%	Precision/%	Recall/%	F1_Score/%
70	82.3	81.5	73.8	77.4
80	89.1	88.2	79.4	83.5
90	93.5	92.8	<i>84.6</i>	<i>88.3</i>
100	<i>95.0</i>	<i>94.4</i>	79.5	86.3
110	94.5	94.1	76.5	84.4

Note(s): Italic fonts indicate that a certain model, method or setting has achieved the highest score in a specific metric compared to others

Source(s): Authors' own creation/work

6. Conclusion

Depression, as a long-term mental disorder, requires prolonged observation and diagnosis. Clinically, a minimum observation period of two weeks is necessary for diagnosis (Shen *et al.*, 2017). According to Eichstaedt *et al.* (2018), a three-month window of social media activity is sufficient to effectively identify whether a user is experiencing depression. Some studies have developed modules for visualizing depression probability. For example, Cai *et al.* (2023) proposed a multivariate time series-based approach to track and visualize the temporal dynamics of users' depressive symptoms, providing both data and methodological support for future individual-level psychological research.

In this study, a method for visualizing changes in users' depressive tendencies is proposed and tested on a general annotated dataset to verify the effectiveness of the proposed method. The visualization module proposed aims to present the temporal trend of a user's depression probability, which holds practical value for clinical intervention by physicians or health professionals. By examining the temporal trajectory of user posts, it becomes possible to identify periods of acute symptom deterioration, as well as potential emotional buffering or relapse risk phases. In addition, professional clinicians can make informed decisions regarding medication adjustments, cognitive behavioral therapy, or emergency psychological intervention based on these trends. Depression trend curves across different periods can also serve to quantify treatment effectiveness, support clinical inquiry and identify symptom peaks.

Compared with the existing research, this paper has an innovative breakthrough in proposing a visual analysis method. These research results provide new ideas and methods for using social media data for mental health monitoring and have important practical value for improving the early detection and accurate diagnosis of depression. Different from the previous researches that only focus on the prediction of a single time point, the visual analysis in this paper can provide better insight into the dynamic changes of users' emotions and

provide richer information for clinical diagnosis. Compared with depression detection methods based on image analysis or physiological signals, the advantage of using text data in this paper is that data acquisition is more convenient, which can be more widely used in large-scale population screening.

About the authors

Yukun Yang is undergraduate student majoring in information management and information systems at the School of Management, Beijing Institute of Technology. His research interests include management information systems, smart health and medical analytics, artificial intelligence and management applications, etc.

Wanxin Qiao is assistant professor in the Department of Management Engineering at School of Economics and Management, Beijing University of Technology, China. She received her Ph.D from Beijing Institute of Technology. Her research interests include user behavior on online community, healthcare & information systems, design science and social media. Dr Qiao has published in Journal of Management Information Systems, Information Processing and Management, Journal of Affective Disorders, among other journals.

Yuting Lin is master's student majoring in management science and engineering at the School of Management, Beijing Institute of Technology. She graduated from the School of Information Technology and Management, University of International Business and Economics with bachelor's degree in data science and big data technology. Her research interests include management information systems, smart health and medical analytics, artificial intelligence and management applications, etc.

Rui Liu is doctoral student majoring in management science and engineering at the School of Management, Beijing Institute of Technology, China. Her research interests include online health, online health communities and user behavior.

References

- American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed., American Psychiatric Association.
- Beck, A.T. (1961), "An inventory for measuring depression", *Archives of General Psychiatry*, Vol. 4 No. 6, p. 561, doi: [10.1001/archpsyc.1961.01710120031004](https://doi.org/10.1001/archpsyc.1961.01710120031004).
- Biradar, A. and Totad, S.G. (2019), "Detecting depression in social media posts using machine learning", *2nd International Conference on Recent Trends in Image Processing and Pattern Recognition, RTIP2R 2018*, Vol. 1037, pp. 716-725, Springer Verlag, Solapur, December 21, 2018 - December 22, 2018, doi: [10.1007/978-981-13-9187-3_64](https://doi.org/10.1007/978-981-13-9187-3_64).
- Cai, Y., Wang, H., Ye, H., Jin, Y. and Gao, W. (2023), "Depression detection on online social network with multivariate time series feature of user depressive symptoms", *Expert Systems with Applications*, Vol. 217, 119538, doi: [10.1016/j.eswa.2023.119538](https://doi.org/10.1016/j.eswa.2023.119538).
- Chen, Z., Wang, D., Lou, L., Zhang, S., Zhao, X., Jiang, S., Yu, J. and Xiao, J. (2025), "Text-guided multimodal depression detection via cross-modal feature reconstruction and decomposition", *Information Fusion*, Vol. 117, 102861, doi: [10.1016/j.inffus.2024.102861](https://doi.org/10.1016/j.inffus.2024.102861).
- Crawford, J.R. and Henry, J.D. (2003), "The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample", *The British Journal of Clinical Psychology*, Vol. 42 No. Pt 2, pp. 111-131, doi: [10.1348/014466503321903544](https://doi.org/10.1348/014466503321903544).
- De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E. (2013), "Predicting depression via social media", *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, Cambridge, MA, pp. 128-137.
- Dejong, T.M., Overholser, J.C. and Stockmeier, C.A. (2010), "Apples to oranges?: a direct comparison between suicide attempters and suicide completers", *Journal of Affective Disorders*, Vol. 124 Nos 1-2, pp. 90-97, doi: [10.1016/j.jad.2009.10.020](https://doi.org/10.1016/j.jad.2009.10.020).
- Deshpande, M. and Rao, V. (2017), "Depression detection using emotion artificial intelligence", *2017 International Conference on Intelligent Sustainable Systems (ICISS), presented at the 2017 International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, Palladam, pp. 858-862, doi: [10.1109/ISS1.2017.8389299](https://doi.org/10.1109/ISS1.2017.8389299).

- Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preotiuc-Pietro, D., Asch, D.A. and Schwartz, H.A. (2018), "Facebook language predicts depression in medical records", *Proceedings of the National Academy of Sciences*, Vol. 115 No. 44, pp. 11203-11208, doi: [10.1073/pnas.1802331115](https://doi.org/10.1073/pnas.1802331115).
- Ghosh, T., Banna, M.H.A., Nahian, M.J.A., Uddin, M.N., Kaiser, M.S. and Mahmud, M. (2023), "An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla", *Expert Systems with Applications*, Vol. 213, 119007, doi: [10.1016/j.eswa.2022.119007](https://doi.org/10.1016/j.eswa.2022.119007).
- Guntuku, S.C., Yaden, D.B., Kern, M.L., Ungar, L.H. and Eichstaedt, J.C. (2017), "Detecting depression and mental illness on social media: an integrative review", *Current Opinion in Behavioral Sciences*, Vol. 18, pp. 43-49, doi: [10.1016/j.cobeha.2017.07.005](https://doi.org/10.1016/j.cobeha.2017.07.005).
- Guo, Z., Ding, N., Zhai, M., Zhang, Z. and Li, Z. (2023), "Leveraging domain knowledge to improve depression detection on Chinese social media", *IEEE Transactions on Computational Social Systems*, Vol. 10 No. 4, pp. 1528-1536, doi: [10.1109/TCSS.2023.3267183](https://doi.org/10.1109/TCSS.2023.3267183).
- Hamilton, M. (1960), "A rating scale for depression", *Journal of Neurology, Neurosurgery and Psychiatry*, Vol. 23 No. 1, pp. 56-62, doi: [10.1136/jnnp.23.1.56](https://doi.org/10.1136/jnnp.23.1.56).
- Harris, K.M. and Goh, M.T.-T. (2017), "Is suicide assessment harmful to participants? Findings from a randomized controlled trial", *International Journal of Mental Health Nursing*, Vol. 26 No. 2, pp. 181-190, doi: [10.1111/inm.12223](https://doi.org/10.1111/inm.12223).
- Harris, K.M., Syu, J.-J., Lello, O.D., Chew, Y.L.E., Willcox, C.H. and Ho, R.H.M. (2015), "The ABC's of suicide risk assessment: applying a tripartite approach to individual evaluations", *PLoS One*, Vol. 10 No. 6, e0127442, doi: [10.1371/journal.pone.0127442](https://doi.org/10.1371/journal.pone.0127442).
- Islam, M.R., Kamal, A.R.M., Sultana, N., Islam, R., Moni, M.A. and Ulhaq, A. (2018), "Detecting depression using K-Nearest neighbors (KNN) classification technique", *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering, IC4ME2 2018*, Rajshahi.
- Jahnavi, Y., Elango, P., Raja, S.P., Parra Fuente, J. and Verdú, E. (2023), "A new algorithm for time series prediction using machine learning models", *Evolutionary Intelligence*, Vol. 16 No. 5, pp. 1449-1460, doi: [10.1007/s12065-022-00710-5](https://doi.org/10.1007/s12065-022-00710-5).
- Kroenke, K., Spitzer, R.L. and Williams, J.B.W. (2001), "The PHQ-9: validity of a brief depression severity measure", *Journal of General Internal Medicine*, Vol. 16 No. 9, pp. 606-613, doi: [10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x).
- Kumar, E.R. and Venkatram, N. (2024), "Predicting and analyzing suicidal risk behavior using rule-based approach in Twitter data", *Soft Computing*, Vol. 28 No. 23, pp. 13821-13829, doi: [10.1007/s00500-023-08095-y](https://doi.org/10.1007/s00500-023-08095-y).
- Kumar, M., Dredze, M., Coppersmith, G. and De Choudhury, M. (2015), "Detecting changes in suicide content manifested in social media following celebrity suicides", *Proceedings of the 26th ACM Conference on Hypertext and Social Media - HT '15, presented at the 26th ACM Conference*, ACM Press, Guzelyurt, pp. 85-94, doi: [10.1145/2700171.2791026](https://doi.org/10.1145/2700171.2791026).
- Liu, D., Feng, X.L., Ahmed, F., Shahid, M. and Guo, J. (2022), "Detecting and measuring depression on social media using a machine learning approach: systematic review", *JMIR Mental Health*, Vol. 9 No. 3, e27244, doi: [10.2196/27244](https://doi.org/10.2196/27244).
- Ma, X., Hancock, J. and Naaman, M. (2016), "Anonymity, intimacy and self-disclosure in social media", *34th Annual Conference on Human Factors in Computing Systems, CHI 2016*, May 7, 2016 - May 12, 2016, Association for Computing Machinery, San Jose, CA, pp. 3857-3869, doi: [10.1145/2858036.2858414](https://doi.org/10.1145/2858036.2858414).
- Mann, P., Paes, A. and Matsushima, E.H. (2021), "Screening for depressed individuals by using multimodal social media data", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35 No. 18, pp. 15722-15723, doi: [10.1609/aaai.v35i18.17858](https://doi.org/10.1609/aaai.v35i18.17858).
- Montgomery, S.A. and Åsberg, M. (1979), "A new depression scale designed to be sensitive to change", *British Journal of Psychiatry*, Vol. 134 No. 4, pp. 382-389, doi: [10.1192/bjp.134.4.382](https://doi.org/10.1192/bjp.134.4.382).

- Peng, Z., Hu, Q. and Dang, J. (2019), "Multi-kernel SVM based depression recognition using social media data", *International Journal of Machine Learning and Cybernetics*, Vol. 10 No. 1, pp. 43-57, doi: [10.1007/s13042-017-0697-1](https://doi.org/10.1007/s13042-017-0697-1).
- Roy, A., Nikolitch, K., McGinn, R., Jinah, S., Klement, W. and Kaminsky, Z.A. (2020), "A machine learning approach predicts future risk to suicidal ideation from social media data", *NPJ Digital Medicine*, Vol. 3 No. 1, 78, doi: [10.1038/s41746-020-0287-6](https://doi.org/10.1038/s41746-020-0287-6).
- Seabrook, E.M., Kern, M.L., Fulcher, B.D. and Rickard, N.S. (2018), "Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates", *Journal of Medical Internet Research*, Vol. 20 No. 5, e168, doi: [10.2196/jmir.9267](https://doi.org/10.2196/jmir.9267).
- Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T.-S. and Zhu, W. (2017), "Depression detection via harvesting social media: a multimodal dictionary learning solution", *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, presented at the Twenty-Sixth International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, Melbourne, pp. 3838-3844, doi: [10.24963/ijcai.2017/536](https://doi.org/10.24963/ijcai.2017/536).
- Statista (2025), "Internet and social media users in the world 2025 | Statista", available at: <https://www.statista.com/statistics/617136/digital-population-worldwide/> (accessed 18 September 2025).
- Tadesse, M.M., Lin, H., Xu, B. and Yang, L. (2019), "Detection of depression-related posts in Reddit social media forum", *IEEE Access*, Vol. 7, pp. 44883-44893, doi: [10.1109/access.2019.2909180](https://doi.org/10.1109/access.2019.2909180).
- Uddin, A.H., Bapery, D. and Mohammad Arif, A.S. (2019), "Depression analysis of Bangla social media data using gated recurrent neural network", *1st International Conference on Advances in Science, Engineering and Robotics Technology, ICASERT 2019*, May 3, 2019 - May 5, 2019, Institute of Electrical and Electronics Engineers, Dhaka, doi: [10.1109/ICASERT.2019.8934455](https://doi.org/10.1109/ICASERT.2019.8934455).
- Wang, Y., Wang, Z., Li, C., Zhang, Y. and Wang, H. (2020), "A multimodal feature fusion-based method for individual depression detection on Sina Weibo", *2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC)*, pp. 1-8, doi: [10.1109/IPCCC50635.2020.9391501](https://doi.org/10.1109/IPCCC50635.2020.9391501).
- Wei, P.-C., Peng, K., Roitberg, A., Yang, K., Zhang, J. and Stiefelhagen, R. (2023), "Multi-modal depression estimation based on sub-attentional fusion", *Workshops Held at the 17th European Conference on Computer Vision, ECCV 2022*, October 23, 2022 - October 27, 2022, Vol. 13806 LNCS, Springer Science and Business Media, Deutschland GmbH, Tel Aviv, pp. 623-639, doi: [10.1007/978-3-031-25075-0_42](https://doi.org/10.1007/978-3-031-25075-0_42).
- World Health Organization (2017), "Depression and other common mental disorders: global health estimates Internet", available at: <https://www.who.int/publications/i/item/depression-global-health-estimates> (accessed 10 October 2025).
- World Health Organization (2019), "Suicide", available at: <https://www.who.int/news-room/fact-sheets/detail/suicide> (accessed 10 October 2025).
- World Health Organization (2023a), "WHO launches new resources on prevention and decriminalization of suicide", available at: <https://www.who.int/news/item/12-09-2023-who-launches-new-resources-on-prevention-and-decriminalization-of-suicide> (accessed 18 September 2025).
- World Health Organization (2023b), "Depressive disorder (depression)", 19 February, available at: <https://www.who.int/news-room/fact-sheets/detail/depression> (accessed 18 September 2025).
- Yan, Z., Peng, F. and Zhang, D. (2025), "DECEN: a deep learning model enhanced by depressive emotions for depression detection from social media content", *Decision Support Systems*, Vol. 191, 114421, doi: [10.1016/j.dss.2025.114421](https://doi.org/10.1016/j.dss.2025.114421).
- Yoon, J., Kang, C., Kim, S. and Han, J. (2022), "D-vlog: multimodal vlog dataset for depression detection", *Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI 2022*, Vol. 36 No. 11, pp. 12226-12234, Virtual, Online, doi: [10.1609/aaai.v36i11.21483](https://doi.org/10.1609/aaai.v36i11.21483).

Yu, B., Zhang, X., Wang, C., Sun, M., Jin, L. and Liu, X. (2020), "Trends in depression among adults in the United States, NHANES 2005-2016", *Journal of Affective Disorders*, Vol. 263, pp. 609-620, doi: [10.1016/j.jad.2019.11.036](https://doi.org/10.1016/j.jad.2019.11.036).

Zhang, W., Xie, J., Zhang, Z. and Liu, X. (2024), "Depression detection using digital traces on social media: a knowledge-aware deep learning approach", *Journal of Management Information Systems*, Vol. 41 No. 2, pp. 546-580, doi: [10.1080/07421222.2024.2340822](https://doi.org/10.1080/07421222.2024.2340822).

Corresponding author

Rui Liu can be contacted at: 18846041760@163.com