Online social network individual depression detection using a multitask heterogenous modality fusion approach ☆

Yiding Wang, Zhenyi Wang, Chenghao Li, Yilin Zhang, Haizhou Wang *

School of Cyber Science and Engineering, Sichuan University, Chengdu 610065, PR China

ABSTRACT

In recent years, the number of people who endanger their lives has been increasing rapidly due to the mental burden of depression. The online social network (OSN) provides researchers with another perspective for detecting individuals suffering from depression. However, existing machine learning-based depression detection studies still leave relatively low classification performance, suggesting that there is significant improvement potential in their feature engineering. In this paper, we manually build and publish a large dataset on Sina Weibo (a leading OSN with the largest number of active users in the Chinese community), namely the Weibo User Depression Detection Dataset (WU3D). It includes more than 20,000 normal users and more than 10,000 depressed users, both of which are labeled and rechecked following the DSM-5 official medical and psychological depression document by professionals. Then, we conclude and propose ten statistical features by analyzing the user's text, social behavior, and posted pictures. In the meantime, text-based word features are extracted using the popular pretrained model XLNet. Moreover, we fuse these features from heterogeneous modalities and implement a multitask learning scheme to train our proposed deep neural network classification model, i.e. FusionNet (FN). The experimental results show that FN has excellent to recognize depressed users on the OSN, achieving the highest F1-Score of 0.9772 on the test set. Compared to existing studies, the proposed method has better classification performance and robustness for unbalanced training samples, as well as reasonable training and inference time. Our work provides a method to fuse multimodal information to detect individual-level depression and has reference significance for similar studies on other OSNs.

© 2022 Elsevier Inc. All rights reserved.

---

1. Introduction

1.1. Background

An Online Social Network (OSN) can be defined as a Web service for establishing a virtual connection between people with similar interests, backgrounds, and activities [1]. With the rapid development of the OSN such as Twitter and Facebook, people are more frequently using it to express opinions and emotions [2]. It provides researchers with a novel and effective way to detect the mood, communication, activity, and social behavior pattern of individuals [3–5]. In the past decade, researchers in various fields have conducted quantitative analyses of different illnesses and mental disorders based on OSN platforms [6–12]. Sina Weibo (hereinafter referred to as "Weibo") is one the most popular OSNs in the Chinese community [13]. A statistic shows the number of Sina Weibo users has reached more than 418 million in the year 2021[1].

Major depressive disorder, referred to as depression, is a common mental disease. According to a survey of the World Health Organization (WHO)[2], more than 300 million people worldwide suffer from depression. Depression can result in great psychological pain, even self-hurting and suicidal tendencies. Moreover, evidence from a health action plan of WHO[3] shows that people suffering from depression are much more likely to end their life prematurely than the general population. Despite the current availability of psychotherapy, medical therapy, and other modalities for the treatment of depression, 76%-85% of patients in low- and middle-income countries remain untreated [14]. The emergence of this phenomenon is not only the lack of medical resources but also the inability to make an accurate assessment in the early stage of depression, which leads to a large number of people with depression difficult to get diagnosis and treatment timely [15].

Pictures, text, videos, and other information posted on the OSN are feature sources that can reflect feelings of worthlessness, guilt, helplessness, and self-hatred, which can help researchers to specifically analyze and characterize depressed individuals [3,15–19]. However, there is an insurmountable problem in online depression detection using traditional analyzing methods: they mostly focus on analyzing the characteristics of users with depression rather than constructing predictive models. Therefore, it is difficult to give timely prediction results of new depressed users.

With the rapid development of artificial intelligence technologies, machine learning-based automatic detection approaches have made great contributions to the detection of depression [20–26]. An automated depression detection model based on machine learning usually needs to analyze various information such as tweets, pictures, videos, and social activity data of users. Then, it gives the classification results of the predicted objects, most of which are presented as a binary result of normal or depressive. If an individual is predicted for a potential depressive tendency, further resources and assistance can be provided, including later medical and psychological diagnoses. Such heuristic learning approaches are quite effective for helping in the early detection of depression since they are capable of handling a large number of instant interactive user data.

1.2. Challenges

However, current approaches to online depression detection still face many unresolved challenges.

Firstly, many current studies are not user-oriented modeling [11,21,27–31]. Those works usually aim to analyze and model the language style of the user. Through sentiment analysis and feature engineering of the tweet text, a classification model is developed to detect whether a specific tweet has a depressive tendency. These works analyzed fine-grained features and achieved pretty good results. However, such results cannot be directly applied to user-level depression detection, or they may lead to an incorrect prediction.

Second, in several existing studies[3,30,32,33]**, the size of the dataset used for modeling is insufficient**, with only a few hundred to a few thousand data samples being used for a machine learning model. Because of the difficulty of accurately obtaining and labeling depressed samples, researchers usually choose to construct small datasets or directly cite and use datasets from other works. As a consequence, the trained model fails to reach good generalization performance, thus hard to accurately predict depressed users on the OSN.

Moreover, not enough studies of user depression detection have been proposed on Weibo compare to Twitter and Facebook. To the best of our knowledge, there is no published large Weibo user depression detection dataset available currently.

Finally, many of the existing proposed models still do not reach a high level of classification performance, i.e. an F1-Score of 90% and above. Thus, these models need to be further improved to achieve better performance.

1.3. Contributions

Given the above problems and challenges, we hereby summarize the contributions of our work as below:

- **We build and publish a large-scale labeled dataset - Weibo User Depression Detection Dataset (WU3D)**[4]. WU3D includes more than 10,000 depressed users and more than 20,000 normal users, each of which contains enriched information fields, including tweets, the posting time, posted pictures, the user gender, etc. This dataset is labeled and further reviewed by professionals. The Web crawler for collecting the data has also been published along with the dataset for subsequent researchers to use.
- **We summarize ten features of depressed users, four of which are the first to be proposed.** Different from some existing work that directly uses the information fields as features, we made statistical analyzes of all the proposed features. These features show significant distribution differences between depressed and normal users in our experiments.
- **We use multitask training and multimodal inputs to construct a Deep Neural Network (DNN) classification model, i.e. FusionNet to predict online depressed users.** The multitask learning strategy can process heterogeneous modality information[5] simultaneously. Experimental results show that it achieves both the highest classification performance and the best robustness to unbalanced training samples.

The subsequent sections of this paper are organized as follows. In Section 2, related work and achievements in the field of depression detection on OSNs are introduced and analyzed. The proposed framework is elaborated in Section 3. Furthermore, Section 4 gives the significant evaluation of statistical features and the performance comparison experiments of several classification models (including our proposed FusionNet). At the last of the paper, Section 5 summarizes our work and discusses directions for future work.

We also note that a shorter conference version of this paper has been published on the proceedings of the 39th IEEE International Performance Computing and Communications Conference (IPCCC 2020) [48]. In this journal version, we further publish the script for dataset collection, enhance and modify the deep learning model, address more features, and detailed elaborate on their definitions in the feature engineering process. Moreover, additional experiments are designed and conducted to further illustrate the effectiveness of our proposed framework.

2. Related Work

The current methods for online depression detection mainly include two directions. **(i)** Manually extracting features and building Traditional Machine Learning (TML) models for classification. **(ii)** Using Deep Learning (DL) approaches to automatically extract features and construct DNN models as classifiers.

Among them, some of the research that uses DL also introduces TML methods to further improve their model performance. The research of each approach will be introduced below respectively.

2.1. Detection Approaches based on Traditional Machine Learning

Mining depression users based on TML mostly use features, i.e. numeric vectors that have been manually analyzed and extracted from users to represent the predicted object (a user, a tweet, a posted picture, etc.).

Choudhury et al. [3] presented a pioneering work in this field of research. They explored potential user behavior to perform a user-oriented depression detection. By measuring behavioral attributes on Twitter users relating to social engagement, emotion, language, and linguistic styles, they discovered useful signals for characterizing depression. Although their trained classifiers did not achieve high classification performance, as a pioneering work in this field, they provided a detailed feature engineering analysis process and a clear modeling approach.

Wang et al. [31] undertook further research using data from Twitter and Weibo. Compared with the work of [3] that made a more comprehensive feature analysis, this study implemented a sentiment analysis approach and proposed man-made rules by utilizing vocabulary to measure depressive tendencies of tweets. Their work indicated that text-based features play a crucial role in online depression detection.

Deshpande et al. [28] proposed a representation learning method based on natural language processing (NLP) to model the text information on Twitter. Different from the previously mentioned work [3,31], they used the Bag of Words (BOW) algorithm to represent the tweet text as a sparse vector, allowing the classifier to automatically learn latent features. Their trained Naive Bayes (NB) classifier reached an F1-Score of 0.8329, while the Support Vector Machine (SVM) classifier only reached an F1-Score of 0.7973.

After, Shen et al. [15] proposed an advanced detecting approach that can be used to detect depressed users timely. They constructed a well-labeled depression detection dataset on Twitter, which had been widely used by subsequent researchers. In the meantime, they extracted six depression-related feature groups covering the text, social behavior, and posted pictures. Their proposed multimodal depressive dictionary learning (MDL) approach can effectively learn the latent and sparse representation of user features. Experiments showed their proposed MDL model achieved an F1 of 0.85, indicating that the dictionary learning strategy and the ensemble of multimodal is quite effective.

---

4 **https://github.com/yiyepianzhounc/Weibo-User-Depression-Detection-Dataset**
5 The modalities include the text, images, and the user social behavior information

In recent years, more TML-based work has begun to emerge [29,32,33]. In particular, Mustafa et al. [32] implemented Frequency-Inverse Document Frequency (TF-IDF) algorithm to weigh the words in tweets. Their trained classifier based on a one-dimensional convolution neural network (CNN-1D) achieved an F1-Score of 0.89. Their work is the first to introduce a neural network model for detecting depressed users on the OSN.

2.2. Deep Learning-based Detecting Approaches

Modeling approaches based on DL are mainly for jointly considering user social behaviors and multimedia information such as the text, pictures, videos, etc. Among them, the modeling of the text information is the main research direction. Researchers have adopted NLP approaches to embed text into a high dimensional continuous vector to automatically mine word features. Some work has also fused manually extracted features into DNN classifiers as part of the input, or integrated traditional classifiers with DNN classifiers to improve performance. These multimodal and ensemble approaches have proven to be an effective way to accomplish various tasks on social network analysis including depression detection [34].

Several DNN classifiers that have achieved significant performance in the NLP classification task were selected and evaluated by Orabi et al. [35]. They used a pretrained Word2Vec [36] model to embed the text of tweets. Their experimental results showed that the CNN-1D with a max-pooling structure reported the highest performance. Compared to other recurrent structures including the recurrent neural network (RNN), the Long Short-Term Memory (LSTM) neural network [37], CNN-based models performed better in the task of depression detection.

Then, Sadeque et al. [38] proposed a latency-weighted F1 metric and applied it in a novel sequential classifier based on the Gated Recurrent Units (GRU). They treated all the text of tweets as documents and input them to the classifier asynchronously, named the "post-by-post" strategy. It allows the model to decide the depressive tendency of a user after each tweet is scanned. Thus, it somehow avoids the time consumption of scanning too many tweets under a certain and obvious depressed user (e.g, a user with 200 tweets recording its anti-depressant experience). This approach can scan and detect depressive tendencies of tweets more efficiently.

Later, based on the prior work [15], Shen et al. [16] discovered that the current research on a specific OSN may be unsuitable and not universal for depression detection on other platforms. Thus, they proposed a cross-domain DNN model with Feature Adaptive Transformation & Combination (DNN-FATC) strategy that can consider features of several aspects comprehensively and transfer the relevant information across heterogeneous domains.

Recently, more studies based on DL have been widely proposed. Gui et al. [17] further discussed the change of classification accuracy of the model under the different proportions of depressed users and pointed out that the highest accuracy can be achieved when the proportion of normal and depressed user samples is close to balance. Moreover, they implemented a reinforcement learning (RL) approach to further improve the performance of the model. Lin et al. [39] used a popular pretrained model, i.e. BERT [40], to embed word vectors. Its hidden layer output was extracted to fuse both text and image features to further accomplish the downstream classification task.

3. Framework and Methodology

To detect the depressed users on the OSN more effectively, we propose a novel framework, as shown in Fig. 1. This framework mainly consists of three parts:

i. *User data collection and labeling.* This module contains two independent Web crawlers (*UserID-Crawler* and *UserInfo-Crawler*), which are used to collect user samples on Weibo.[6] Then, it is responsible for filtering and labeling the collected data to construct the Weibo User Depression Detection Dataset (WU3D).

ii. *Feature extracting.* This module is in charge of extracting the user's text information including nicknames, profiles, tweet text, and concatenating them into a long text sequence. Then, the sequence is entered to the pretrained model XLNet [41] to obtain the embedded word features. In the meantime, this module also extracts statistical features from user's post text, social behavior, and posted pictures. Finally, these features are jointly used in the classification model.

iii. *Model Training and predicting.* This module implements a depression detection model based on DNN, namely Fusion-Net, which receives features input from the *Feature Extracting* module. The proposed FusionNet can be trained in a multitask learning mode, in which word vectors and statistical features can be used jointly to optimize the classifier in each training step.

The following parts of this section will elaborate on the theoretical construction and implementation methods of these modules, respectively.

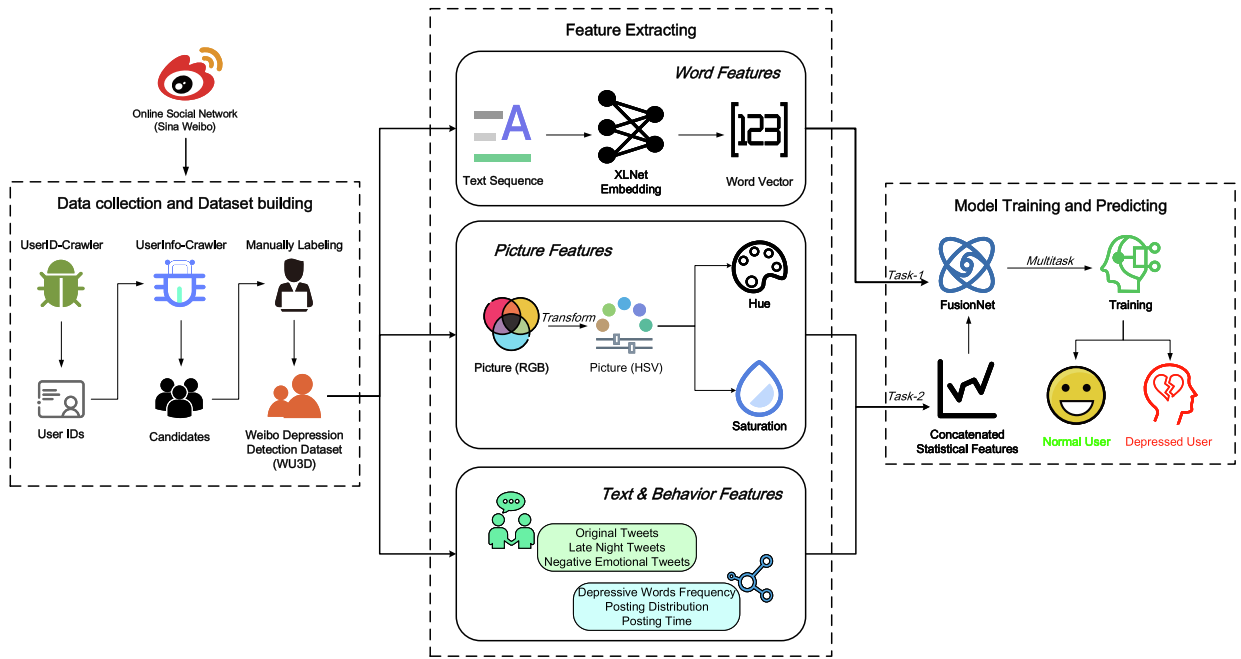[6] Both of the crawlers are published together with the WU3D dataset.

Fig. 1. The Framework of the Proposed Method.

3.1. User Data Collection and Labeling

3.1.1. Data collection

A user ID can be used to uniquely identify a user. With a user ID, the crawler can access the user's home page and collect information from it. First of all, *UserID-Crawler* is constructed to collect user IDs of depressed candidates. The API provided by Weibo official[7] is used to obtain as accurate information as possible. Our strategies for collecting user IDs of depressed candidates include:

(i) Collecting data from the Weibo Super Topic of "'抑郁症'" ("Depression" in English). The Super Topic is a social group on Sina Weibo that gathers users with common interests. It has been proved that individuals who share the same background are more likely to trust each other, thus will gather to form aggregations [42]. According to our investigation and analysis, there are a large number of active depressed users posting under the topic of "Depression". Collecting data in this way can greatly improve the efficiency of gathering depressed user samples. Therefore, *UserID-Crawler* collects depressed candidates under this topic and forms a list of their user IDs.

(ii) Collecting data through the function of "'微博搜索'" ("Weibo Search" in English) provided by Weibo official[8]. We use high-frequency words including "'抑郁症'" ("Depression" in English), "'自杀'" ("Suicide" in English), "'痛苦'" ("Pain" in English) and the late-night period (from 0:00 a.m. to 6:00 a.m.) as two main search conditions to crawl user IDs for collecting more depressed candidates.

Through the above two crawling strategies, we have collected sufficient user IDs of depressed candidates. Then, with the user ID list, *UserInfo-Crawler* is implemented to collect detailed user information from its homepage. The specific information fields collected by *UserInfo-Crawler* are shown in Fig. 2.

We divide the information for each user sample into two domains: the user domain and the tweet domain. The user domain contains the user's gender, birthday, profile (a short text of sentence of the user's self-description), the number of followers, the number of followings, and the list of tweets. For each tweet in the tweet domain, it contains the tweet text, the posting time, posted pictures, the number of likes, the number of forwards, the number of comments, and an identifier that identifies whether the tweet is original or not.

For normal candidates, we use *UserID-Crawler* to collect them under four Super Topics including "'日常'" ("Daily" in English), "'正能量'" ("Positive Energy" in English), "'榜姐每日话题'" ("Daily Topic" in English), "'互动'" ("Interaction"
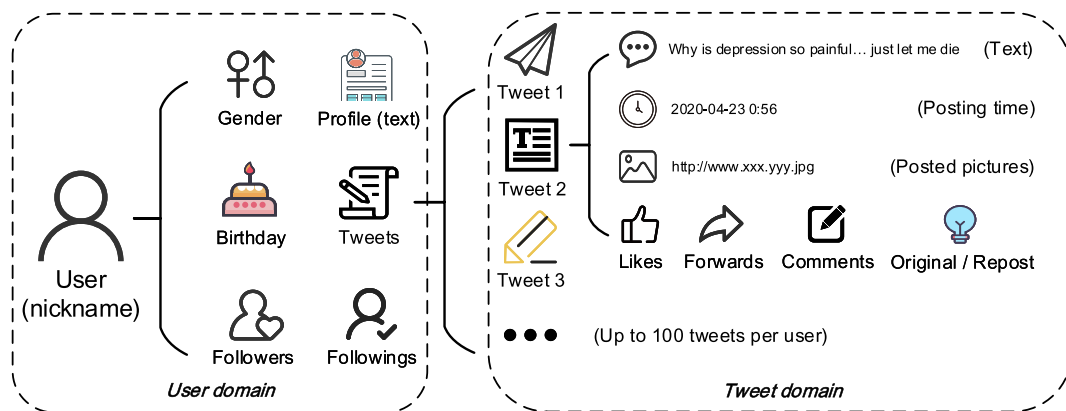

Fig. 2. Information Fields (per user) Collected by the UserInfo-Crawler.

in English) to form a list of normal candidate IDs. Then, the more detailed user information is collected through *UserInfo-Crawler* to form the same data fields and structure as the depressed candidates. Based on the previous steps, we have collected 125,479 depressed candidates and 65,913 normal candidates.

3.1.2. Data filtering and labeling

Automated scripts are implemented to filter out those non-personal accounts by identifying the user's *"account type"* field, including marketing accounts, official accounts, and social bots.

The automated filtered normal candidates are labeled as normal users directly without further manual labeling. For depressed users, we invite professional data labelers to complete the data labeling process of depressed candidates. To ensure that the results are highly reliable, the labeled data has been reviewed twice by psychologists and psychiatrists. Also, our labeling principles are made jointly by them and us authors, using the DSM-5 Major Depressive Disorder Fact Sheets [9] as the main reference. DSM-5 is a taxonomy and diagnostic tool published by the American Psychiatric Association, in which behavioral manifestations related to depression are the main criteria for the diagnosis of depression, and an important reference for analyzing the depression of online users. The principles of the data labeling can be described as follows:

i. Depressed candidates with a self-reported history of depression, confirmed diagnosis, currently taking antidepressants, and recording antidepressant experiences in multiple tweets will be labeled as depressed users.
ii. If a candidate's tweets have repeatedly appeared the content of describing psychological suffering, mental anguish, and strong suicide intention, the user will be identified as depressed. [10]
iii. If the posted pictures of a candidate repeatedly involve or show bloodshed and self-harming content and the tweet text includes keywords such as "'抑郁'" ("Depression" in English) and "'自残'" ("Self-harming" in English), the candidate will be labeled as a depressed one.
iv. Candidates who partially meet the above conditions but have too many unrelated contents such as forwarding lottery prizes, receiving red envelopes, advertising information, will be directly discarded.

Therefore, the target dataset, i.e. WU3D, is constructed. It contains both labeled normal and depressed users. Specifically, according to the labeling principles, there are two parts of users considered as "depressed" in our dataset: (1) users that have already been medically diagnosed with depression and (2) users that have depression tendencies without a confirmed diagnose. These two groups of users are both important in online social network depression detection. The specific information of the Candidates (CAND) and WU3D are given in Table 1. We counted the normal sample, the depressed sample, and the total for each of the two types. In particular, we give a detailed number of users and their posted tweets, and posted pictures.

All of the candidates were collected from March 2020 to May 2020. A total of over 200,000 user samples were collected, including 125,479 depressed candidates and 65,913 normal candidates. After strict data filtering and labeling, the number of depressed users in WU3D reached 10,325, with the retention rate of 8.23%; the number of normal users reached 20,338, with the retention rate of 29.34%. The total user data retention rate was 15.50%.

---

[9] https://psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM-5-Depression-Bereavement-Exclusion.pdf
[10] Since depression is a long-term mental illness, we focus on users who post this type of tweet for a relatively long period. In the analysis and statistics of our data labeling process, this "long-period" mostly refers to one month or longer, and the frequency of "repeatedly" mostly refers to five times or longer.

Table 1
Dataset Statistics.

Dataset	Category	User	Tweet	Picture
CAND	Depressed	125,479	5,478,806	2,354,701
	Normal	65,913	4,927,904	3,631,537
	Total	191,392	10,406,710	5,986,238
WU3D	Depressed	10,325	408,797	160,481
	Normal	22,245	1,783,113	1,087,556
	Total	32,570	2,191,910	1,248,037

Table 2
Manually Extracted User Features.

Group	Feature name	Symbol	Source
Text: Ψ	Proportion of negative emotional tweets	ψ_{NP}	First proposed in our work
	Frequency of depression-related words	ψ_{FDW}	[3,31,33,16,15,10,12], modified in our work
Social behavior: Φ	Proportion of original tweets	ϕ_{POP}	[31,16,10]
	Proportion of late-night posting	ϕ_{PLNP}	[31,33,15,16,10], modified in our work
	Posting frequency (per week)	ϕ_{PF}	First proposed in our work
	Standard deviation of posting time	ϕ_{SDPT}	First proposed in our work
Picture: Γ	Frequency of picture posting	γ_{FPP}	First proposed in our work
	Proportion of cold color-styled pictures	γ_{PCP}	[16]
	Standard deviation of hue	γ_{SDH}	[15,16,10], modified in our work
	Standard deviation of saturation	γ_{SDS}	[15,16,10], modified in our work

Table 3
Variable and Function Symbol Definitions.

Symbol	Description
P	The posted tweet set of a user, including original and repost tweets.
t_p	The posting time of a tweet.
l_e	The emotional label of a tweet.
n_d	The number of depression-related words in a tweet.
T	A set of all user text information, including the nickname, the profile, and the tweet text.
π	The posted pictures set of a user.
$\mu = (h_\mu, s_\mu, v_\mu)$	The dominant color of a picture, a ternary contains hue, saturation and brightness of HSV color space. One picture has one dominant color.
$\overline{X}$	The mean sample value of an attribute.
S_ξ	The concatenated user long text sequence.
Δ	The max length of the long text sequence S_ξ.
C	The function that calculates the number of elements in a set.
L	The loss function of a neural network.
Θ	The parameter set of a neural network.
y	The true label of a user data sample.
$\hat{f}$	The objective function of a neural network. It inputs a user's feature vector and outputs the predicted label.
J	The joint optimization function of a neural network.

3.2. Statistical Feature Engineering

Several previous studies have defined features that are quite effective for detecting depressed users, such as the proportion of late-night tweets, the proportion of original tweets, and the mean value of hue and saturation. Based on their work, we firstly perform feature engineering of user features in three aspects: user's post text, social behavior, and posted pictures. We then summarize ten user-level features used for depression detection including four brand-new proposed and two modified [11]. These features are extracted using statistical approaches, including the scale, the mean value, the standard deviation, etc.

Descriptions of these features are shown in Table 2. The features are divided into three groups, including text-based features, social behavior-based features, and picture-based features. Here, we give specific descriptions and formulas to calculate each feature. In Table 3, symbol definitions that appear in this section and subsequent sections are given.

---

[11] We have conducted pre-experiments on our dataset. It proves that our modified features are more effective than the original ones.

3.2.1. Text-based features

Proportion of negative emotional tweets. In previous works for Twitter[3,38], by considering the number of tweets with negative emotions, they have achieved good results in distinguishing depressed users from normal ones. As users may have considerable differences in the total number of tweets, we instead adopt the "proportion" to normalize the feature and to minimize the fluctuation rather than directly using the "number". Although depressive tendencies do not fully equate to the expression of negative emotions when the proportion of tweets with negative emotions reaches a certain level, it can reflect that the user's mental state is depressed and painful, thus can reveal a tendency of depression. We use the Text Sentiment Analysis API of the Baidu Smart Cloud Platform[12] to label all the original tweets. The API returns three emotional labels: 0 for negative, 1 for neutral, and 2 for positive. We retain the negative emotions of label 0 and summarize labels 1 and 2 as a category of non-negative emotions. For all the original tweets under each user, we give the definition of ψ_{NP} in Eq. (1), in which $C_{(P_o)}$ is the total number of original tweets, $C_{(l_e)}$ is the total number of original tweets with negative emotions:

$$\psi_{NP} = \frac{1}{C_{(P_o)}} \times C_{(l_e)}, \ \psi_{NP} \in [0, 1] \tag{1}$$

Frequency of depression-related words. Researchers have focused on the lexical and semantic analysis of the tweet text and quantified these features by self-constructing or quoting depression-related semantic lists [3,31,33,15,16,10,12]. The results of the existing studies indicate that features based on high-frequency depression keywords can significantly improve the classification performance. We use "frequency" to describe how frequently depression-related words appeared in a user's tweets, reflecting its potential depressive tendencies. In our previous investigation and analysis on Weibo, we summarized a list of high-frequency words for depression. Here, it is used to calculate the frequency of depressive words in users' original tweets. The number of occurrences of depression-related words of each tweet n_d is counted by matching the keyword list. Then, ψ_{FDW} is calculated by:

$$\psi_{FDW} = \frac{1}{C_{(P_o)}} \times \sum_{i=1}^{C_{(P_o)}} n_{d_i}, \ \psi_{FDW} \in [0, \infty) \tag{2}$$

3.2.2. Social behavior-based features

Proportion of original tweets. Several related works have proved that depressed users are more likely to post a large number of original tweets to express their negative psychological state, with relatively few repost tweets [31,16,10]. Therefore, we use the proportion of original tweets to distinguish between depressed users and normal users. Here we use $C_{(P)}$ to calculate the total number of tweets, including original tweets and repost tweets. Then, ϕ_{POP} is defined by:

$$\phi_{POP} = \frac{1}{C_{(P)}} \times C_{(P_o)}, \ \phi_{POP} \in [0, 1] \tag{3}$$

Proportion of late-night posting. Late night is a time when depressive symptoms more frequently attack, thus depressed users tend to be more likely to post tweets in this period [31,33,15,16,10]. Moreover, the late-night period is the time when normal users sleep and rest. They rarely use social tools during this time and therefore send very few tweets. We use the proposed feature "Tweet Time" in Ref. [16] and make minor modifications. The time range of 0:00–6:00 is adopted as the late-night period. Moreover, all the tweets of a user are used to calculate, including original and repost ones. Then, ϕ_{PLNP} is given by Eq. (4), in which $C_{(t_p)}$ is used to calculate the total number of tweets posted in the late-night period from 0:00 a.m. to 6:00 a.m.

$$\phi_{PLNP} = \frac{1}{C_{(P_o)}} \times C_{(t_p)}, \phi_{PLNP} \in [0, 1] \tag{4}$$

Posting frequency (per week). The previous study for Twitter[31] has found that there is also a difference in posting frequency between normal and depressed users. Depressed users tend to post large numbers of tweets when they are suffering from depression and heavily rely on social media to express their painful feelings. Moreover, "Week" is a moderate time size and has stronger periodicity than "Month". We take the earliest t_{pE} and latest release time t_{pL} as the interval, count the total number of posted tweets $C_{(P_i)}$ during the time, and then divide it by the total number of weeks $C_{(W)}$ to get the weekly frequency value. Thus, ϕ_{PF} can be represented by Eq. (5):

$$\phi_{PF} = \frac{|t_{pL} - t_{pE}|}{C_{(W)}} \times C_{(P_i)} \tag{5}$$

Standard deviation of posting time. The posting time of depressed users tends to be concentrated in the late-night, while the relative distribution of post time of normal users is more discrete within a day [31,15,16,10]. Hence, we use the standard deviation to describe this phenomenon, to reflect the aggregation trend of users' posting time. The smaller the value

---

[12] http://ai.baidu.com/tech/nlp/sentiment_classify

of this feature, the more likely that the user would post at a specific period. Here, we consider all the original and repost tweets. The mean value of posting time $\overline{X}_{SDPT}$ is calculated by:

$$\overline{X}_{SDPT} = \frac{1}{C_{(P)}} \times \sum_{i=1}^{C_{(P)}} t_{P_i} \tag{6}$$

Then, ϕ_{SDPT} can be defined as:

$$\phi_{SDPT} = \sqrt{\frac{1}{C_{(P)}} \times \sum_{i=1}^{C_{(P)}} \left(t_{P_i} - \overline{X}_{SDPT}\right)^2} \tag{7}$$

3.2.3. Picture-based features

Frequency of picture posting. In existing works for Twitter and Weibo [16,10], "Tweet with pictures" is categorized into "Tweet type" to measure how often users post pictures in their tweets and has achieved good performance. Based on our prior research on Weibo, we also found that depressed users were more likely to use a lot of text to express their feelings and mental states, thus having fewer posted pictures than normal users. Therefore, we propose this feature to reflect users' habit of posting pictures. $C_{(\pi)}$ represents the total number of posted pictures. Then, we calculate γ_{FPP} by:

$$\gamma_{FPP} = \frac{1}{C_{(P_o)}} \times C_{(\pi)} \tag{8}$$

Proportion of cold color-styled pictures. Studies for Twitter [16,10] and Weibo [15,16,10] have shown that compared to normal users, depressed users tend to post pictures with a relatively colder color. Therefore, we extracted three hue and saturation-related features to distinguish depressed users from normal users.

However, the warmth and coolness of a picture is a relative concept, and the human eye will give different conclusions when contrasting different colors. Lin et al. [10] proposed a range definition of cold colors by analyzing the hue rings, which is used as our definition of the cold color range as $h_\mu \in (30, 110]$.

For the three color-related features, we compute them using values from the Hue, Saturation, Value (HSV) color model. Similar to Red-Yellow-Blow (RGB) color space, HSV is a color space that represents the intuitive properties of colors, which are composed of hue, saturation, and lightness. Among the three attributes, "hue" refers to the category of colored light. Different wavelengths of light give different colors and hues. It is measured by the angle value, with a range of 0–360 degrees. From red to counterclockwise, the red hue is defined as 0 degrees, the green is 120 degrees, and the blue is 240 degrees.

Saturation indicates the degree of color close to the color of the spectrum, and usually takes a value of 0 to 1. The larger the value, the more saturated the color. After converting the RGB value of each pixel to the HSV color space, we calculate the dominant color ternary $\mu = (h_\mu, s_\mu, v_\mu)$. The algorithm for extracting the dominant color is given in Algorithm 1.

Algorithm 1: Dominant Color Extraction

Input: τ, All the pixels of a picture, represented in the HSV color space.
Output: The dominant color pixel of the picture

1 **Initialize:** $threshold \leftarrow 30$
  /* the striking pixel threshold   */
2 **Initialize:** Array SP_Arr
  /* to storage the striking pixels   */
3 $\overline{\tau} \leftarrow$ the average of τ
4 **for** *every pixel **in** τ* **do**
5     $h_\tau = pixel[0]$;   // $pixel = (h_\tau, s_\tau, v_\tau)$
6     $\overline{h_\mu} = \overline{\tau}[0]$;   // $\overline{\tau} = (\overline{h_\mu}, \overline{s_\mu}, \overline{v_\mu})$
7   **if** $| h_\tau - \overline{h_\mu} |> threshold$ **then**
8       $SP\_Arr \leftarrow SP\_Arr + pixel$
9 **end**
10 $\mu \leftarrow$ the average of SP
11 **return** μ

The dominant color is the most attractive and the dominant color in a picture. Thus, we introduce the striking pixel (SP) to represent these colors. The SP plays an important role in the intuitive perception of the entire picture, usually measured by the absolute difference between a specific pixel and the average hue of the entire picture. The algorithm first inputs a picture with all the pixels represented by the HSV color space. It initializes a manually assigned threshold and an array SP_Arr to store striking pixels. Then, the algorithm calculates the average color $\overline{\Omega} = \left(\overline{h_\mu}, \overline{s_\mu}, \overline{v_\mu}\right)$ of the picture and iterates through each pixel, comparing the absolute value of the difference between its hue value and $\overline{h_\mu}$. If the difference is greater than the threshold, the currently iterated pixel is defined as an SP. Finally, by calculating the average value of the SP array, the dominant color ternary μ is calculated. Several rounds of tests have been run to choose the best value of the threshold (here set to 30).

We count the total number of posted pictures with $h_\mu \in (30, 110]$ and $s_\mu < 0.7$, denote it as $C_{(\pi_{cold})}$. Then, γ_{PCP} is calculated by:

$$\gamma_{PCP} = \frac{1}{C_{(\pi)}} \times C_{(\pi_{cold})}, \gamma_{FPP} \in [0, 1] \tag{9}$$

Standard deviation of hue and **Standard deviation of saturation**. These two features are used to reflect the fluctuation of the user's picture color. The previous works used the mean values of hue and saturation as the picture features and achieved good results on Twitter [16,10] and Weibo [15,16,10]. In our research, we found that the hue of depressed users' pictures is more concentrated in colder ranges and the saturation value is relatively low. On the contrary, the hue and saturation distribution of normal users is more dispersed and average. We take the hue value $h\mu$ and the saturation value s_μ to calculate their mean values $\overline{X}_{SDH}$ and $\overline{X}_{SDS}$ by:

$$\overline{X}_{SDH} = \frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} h_{\mu_i} \tag{10}$$

$$\overline{X}_{SDS} = \frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} s_{\mu_i} \tag{11}$$

Then, γ_{SDH} and γ_{SDS} can be defined using the following equations:

$$\gamma_{SDH} = \sqrt{\frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} \left(h_{\mu_i} - \overline{X}_{SDH}\right)^2} \tag{12}$$

$$\gamma_{SDS} = \sqrt{\frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} \left(s_{\mu_i} - \overline{X}_{SDH}\right)^2} \tag{13}$$

3.3. Word Feature Extracting

3.3.1. Text sequence construction

Algorithm 2 gives the approach to construct the user text sequence. Considering that the user nickname and profile can also reflect its current emotional state, they are also concatenated to the tweet text.

Algorithm 2: User Long Text Sequence Construction

Input: T, a collection of user nicknames, profiles, and all tweets' text.
Output: The concatenated long text sequence S_ξ

1 **Initialize:** An empty string S_ξ
2 **Initialize:** The max length of text sequence Δ
3 **for** *text in T* **do**
4   **if** *the length of S_ξ > Δ* **then**
5   **break**
6   **if** *text belongs to an original post* **then**
7       $S_\xi \leftarrow S_\xi + text$
8   **else if** *text belongs to a repost* **then**
9   **if** *text = "Repost"* **then**
         `/* ignored the default repost reason`   `*/`
10   **continue**
11   **else**
12          $S_\xi \leftarrow S_\xi + text$
13   **else**
      `// User nickname or profile`
14       $S_\xi \leftarrow S_\xi + text$
15 **end**
16 **return** S_ξ

The algorithm firstly inputs all the text information (defined as T) of the user and constructs the concatenated user long text sequence S_ξ by traversing T. After entering the loop, the algorithm first determines if the current length of the concatenated text sequence is greater than the maximum length Δ; the algorithm ends if the condition is satisfied. Then, it concatenates the user's original tweet text in chronological order from the latest to the earliest. Moreover, when a user reposts a tweet, Weibo will ask the user to fill in the reason for the retweet. In particular, if the user does not fill in the reason, the text "'转发微博'" ("Repost" in English) will be automatically added as default. This default repost reason is not retained in the text sequence S_ξ since it does not express any opinions and feelings.

3.3.2. Word Embedding

To effectively vectorize the text sequence constructed above and apply this feature to the classification algorithm, the characteristics of this long text sequence are further discussed.

First, the sequence is strongly contextually linked. This link exists not only within a single tweet but also among the contexts of multiple tweets. For example, a user may post multiple tweets at different times about depression diagnoses, depression onset, medication treatment, and inner distress. The integration of all the information is usually the key to determining whether a user is depressed or has a depressive tendency.

Secondly, considering that under real circumstances, not all the posted tweets would have depression-related content even for users that are medically diagnosed with depression. That said, capturing text semantics such as "the user states that he has been diagnosed with depression" and "the user expresses a strong inclination toward suicide" is critical for the detection of depression using the long text sequence S_ξ.

Considering these aspects, several state-of-the-art word embedding algorithms are discussed here. Transformer [43] is a model that replaces the recurrent neural network with the attention mechanism. It calculates the weights of each unit in a long sequence to effectively capture the important semantic information. Moreover, BERT [40] is a two-way encoder that is proposed recently. However, due to the limitations of the "Position Embedding" structure in BERT (including its derivative models ALBERT and ROBERTa), the maximum sequence length *Delta* for single processing is restricted to 512 units. Furthermore, the existing truncation or batch processing of long text sequences used in Ref. [39] will significantly increase the time complexity of processing, which is considered as an unsuitable solution for performing a timely depression detection task. Therefore, the ideal word embedding model must have the ability to process long text efficiently and accurately.

A novel language model, namely XLNet [41], is then proposed by Yang et al. Since it combines the features of language models such as auto-regression and auto-encoding, XLNet has resolved the problem that BERT ignores the relationship between the Masked locations and can process longer text sequences. In this paper, XLNet-Chinese-base[13] is used as the upstream word embedding model, then a DNN classifier, FusionNet, is implemented to handle the downstream tasks.
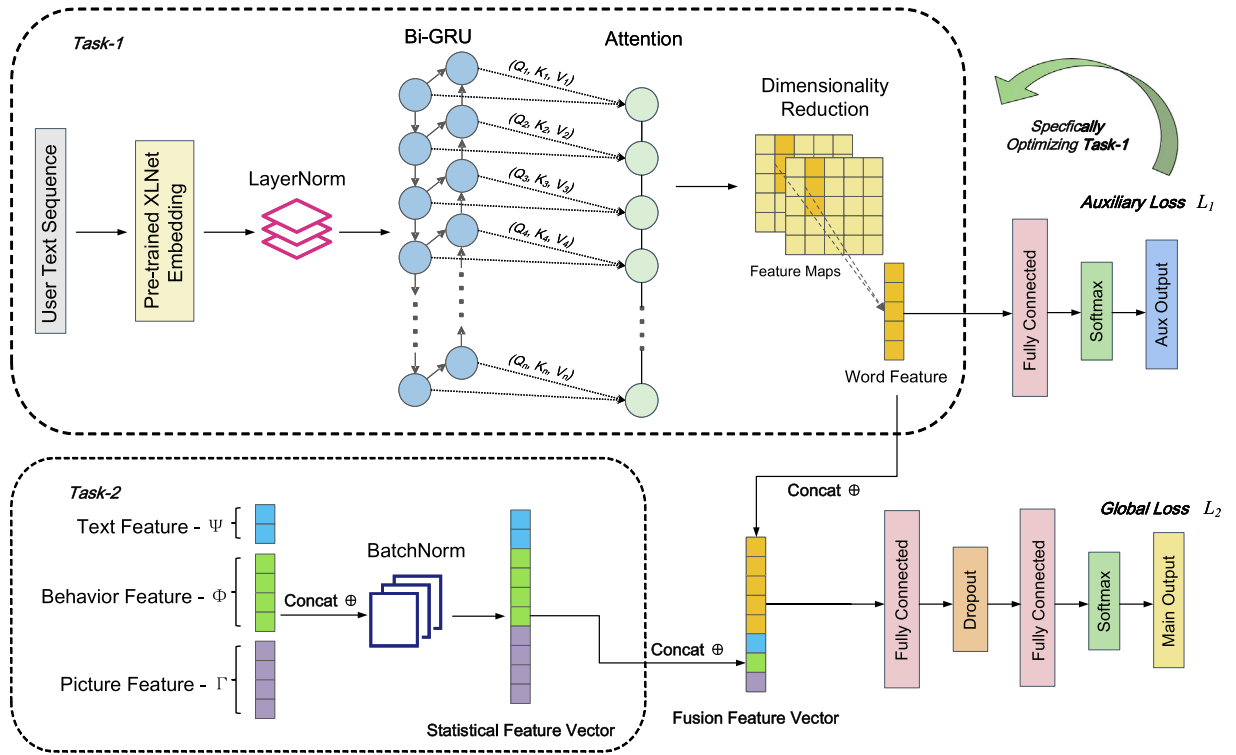
---

[13] https://github.com/ymcui/Chinese-XLNet

Fig. 3. The Structure of our Proposed FusionNet (FN).

3.4. Multitask Classification Model: FusionNet

Multitask learning is an integrated learning strategy that synchronizes model training in a way that multiple tasks share collective network structures and weights. Based on this strategy, we construct a DNN classifier with Bi-GRU with attention as its main structure to comprehensively use features of different modalities including the text, social behavior, picture, and the word. As shown in Fig. 3, the word vector classification task *(Task-1)* obtained from the upstream embedding model XLNet and the manually extracted statistical feature classification task *(Task-2)* are considered as two classification tasks for detecting depressed users. It is worth mentioning that we have set the XLNet to "no gradients", indicating the parameters in the pretrained XLNet are fixed and static. No optimization will be performed on XLNet in the training phase and it simply serves as a static "word feature extractor" or "plain text encoder". Loss functions L_1 and L_2 with different weights ω_1 and ω_2 are manually defined to simultaneously train and optimize the network.

Firstly, the user text sequence S_ξ is embedded by XLNet, and the output of the last hidden layer is connected to the layer normalization (LN) layer. Then, the LN layer is connected to the Bi-GRU layer with attention[14] to capturing the key information and reducing the dimensionality of the word vector.

For *Task-1*, this one-dimensional word feature connects the Full Connected (FC) layer, the Dropout layer, and the Softmax layer to directly output classification results. We set a auxiliary loss function L_1 for network optimization in' *Task-1*, to help accelerate its network convergence. For *Task-2*, the word feature is concatenated to the manually extracted statistical feature input. The statistical feature groups $[\Psi, \Phi, \Gamma]$ are regularized by the Batch Normalization (BN) layer [44].

Moreover, the fused feature vector accesses multiple FC layers with activation functions, activating the hidden layers' outputs to further improve the fitting capability of the network. Finally, the network is connected to the Softmax layer, and the classification result is given by the main output. The main loss function L_2 is used to optimize the whole FusionNet network.

We define the weight parameter set of *Task-1* network as Θ_{aux}, while the objective function is defined as $\hat{f}_1$. The global weight parameter set of the whole network is represented by Θ_g, while the objective function is defined as $\hat{f}_2$, so that $\Theta_{aux} \in \Theta_g$. Adopting the multitask learning strategy, the joint optimization function J can be described as:

---

[14] Bi-GRU & Attention has been proven to be an effective structure in dealing with NLP tasks, both in theory and in best practice. We select and use the best components here, then use experiments to prove their effectiveness and better performance.

$$\begin{cases} J_1 = \sum_{i=1}^{C_{(U)}} L_1\left(y_i, \hat{f}_1\left(S_{\xi_i}\right), \Theta_{aux}\right), \\ J_2 = \sum_{i=1}^{C_{(U)}} L_2\left(y_i, \hat{f}_2(\Psi_i, \Phi_i, \Gamma_i), \Theta_g\right). \end{cases} \tag{14}$$

$$J_{\left(\Theta_{aux}, \Theta_g\right)} = \omega_1 \times J_1 + \omega_2 \times J_2 \tag{15}$$

In Eq. (14), y_i represents the true label (normal or depressed) of a specific user sample i. The $\hat{f}_1$ and the $\hat{f}_2$ both output the predicted label of user sample i. In Eq. (15), ω_1 and ω_2 are the manually assigned weights of loss function L_1 and L_2.

Manually extracted features and several uncertain parameters will be evaluated in the following section.

4. Experimental Evaluation and Analysis

4.1. Experiment Setup

4.1.1. Data cleaning

Since the original user data obtained by the crawler has irrelevant information, to minimize the experimental bias and improve the efficiency of model training, we have made the following assumptions and preprocess for the data:

i. Considering that our construction and analysis of user word features are fully text-oriented, all the non-text contents in the tweets are removed.

ii. The special expression of mixing English and Chinese: For example, the user may express "一直在被大黑狗折磨，我好痛苦啊……" ("The **black dog** has always been torturing me, I'm in so much pain…" in English) as "一直在被**black dog**折磨，我好痛苦啊……",", since some depressed users implicitly express "抑郁症" ("depression" in English) as "黑狗" ("black dog" in English). However, considering that the pre-trained corpus is constructed from multiple sources[15] including the informal text corpus from several Chinese online social platforms, and the formal text corpus from the Chinese Wikipedia, it is reasonably assumed that XLNet already has the robustness to this "mixed" multi-language plain text corpus. Also, this approach is also used in work [45] and has obtained excellent performance. Thus, we do not uniformly translate these mixed expressions.

4.1.2. Dataset slicing

In this part, WU3D is divided into four subsets: D_1, D_2, D_3, and D_4. All of the subsets are sampled using a fixed random seed without a crossover.

Among them, D_1 is used for DNN model training and the 10-fold cross-validation of TML classifiers. Furthermore, D_2 is used as a fixed dataset for validation in each round of neural network training. Finally, we test the performance of models on D_3 and give the evaluation metrics. As a supplementary dataset, D_4 contains only 325 depressed users and 12,245 normal users, which will be only used in Section 4.5 of unbalanced training samples. Statistics of the sliced datasets are given in Table 4 above.

4.1.3. Evaluation metrics

The experimental metrics used in this section mainly include supervised machine learning metrics. **True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)** are commonly used to describe the number of classes predicted by models in classification tasks. Among them, TP represents the number of depressed users correctly predicted, TN represents the number of normal users correctly predicted, FP represents the number of depressed users incorrectly predicted, and FN represents the number of normal users incorrectly predicted. Based on the above four metrics, we can further define the advanced metrics by:

$$Accuracy = \frac{|TP + TN|}{|TP + TN + FP + FN|} \tag{16}$$

$$Precision = \frac{|TP|}{|TP + FP|} \tag{17}$$

$$Recall = \frac{|TP + TN|}{|TP + FN|} \tag{18}$$

---

15 https://github.com/ymcui/Chinese-XLNet#%E5%9F%BA%E7%BA%BF%E7%B3%BB%E7%BB%9F%E6%95%88%E6%9E%9C

Table 4
Dataset Slicing Statistics.

Dataset	Depressed			Normal		
	User	Tweet	Picture	User	Tweet	Picture
WU3D	10325	408797	160481	22245	1783113	1087556
D_1	8000	319115	125096	8,000	630064	355353
D_2	1000	37315	14991	1000	79417	46060
D_3	1000	38941	15211	1000	80066	45066
D_4	325	13426	5183	12245	993566	641077

$$F_1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (19)$$

Moreover, Receiver Operating Characteristic (ROC) curve is a curve with False Positive Rate (FPR) as the horizontal axis and True Positive Rate (TPR) as the vertical axis, which can be used to visually reflect the classifier performance. Furthermore, in the experiment of the statistical features, we introduce the cumulative distribution function (CDF) curve as the evaluation metric.

4.1.4. Training settings

(i) Hardware and software environments. We complete all the experiments in this section on a workstation with Intel Xeon Silver 4212 CPU, NVIDIA RTX TITAN GPU with 24 GB GRAM, and 32 GB RAM. The programming-related settings are Python v3.7.5, Anaconda v4.8.3, TensorFlow v2.1.0, and Scikit-learn v0.23.1.

(ii) Baseline statistical feature classifiers. For the statistical feature classification task, we select several popular TML model from existing studies to demonstrate the effectiveness of the ten concluded features.

- **LR:** Logistic Regression is a commonly used linear model [27] and has good classification performance.
- **NB:** A Naive Bayesian classifier is a simple probabilistic classifier using Bayes' theorem as a basis. Its implementation is relatively simple and is used more often in related works [31,28,29,15,10,11].
- **SVM:** Support Vector Machine classifiers apply the principle of structural risk minimization to the field of classification and are the most used classifiers in previous studies [3,28,33,29,32,35,38,10,12]. In our work, we discussed different kernel modes including the linear kernel, polynomial kernel and radial basis kernel of the SVM, respectively.
- **RF:** Random Forest is an algorithm that integrates multiple trees through ensemble learning, which is also used widely in related works [33,32,10,11]. The basic unit of RF is the decision tree.
- **AB:** Adaptive Boosting is an ensemble learning algorithm that combines multiple simple classifiers [27].
- **GBDT:** Gradient Boosting Decision Tree is a classification model that uses an integrated additive model to continuously reduce the training residuals. GBDT is one of the algorithms with an excellent generalization ability in TML, however, to the best of our knowledge, there is no existing work using GBDT as a classification model.
- **BP:** The Back Propagation (BP) network is extracted from the main output part of our proposed FusionNet with the same parameter settings. To be specific, BP is composed of "FC + Dropout + FC + Softmax".

(iii) Baseline word vector classification networks. For word vector classifiers, we use several popular neural network structures as the main structures, appended by the FC layers and the Softmax layer to output the classification label.

- **CNN-1D:** One-dimensional convolutional neural networks are more widely used in natural language processing and have achieved good performance in the task of depression detection [32,35,30,10,27].
- **Bi-LSTM:** The bi-directional LSTM network splices two-way LSTMs together, which are more capable of handling time series data [35,38,39].
- **Bi-GRU:** Similarly, the bi-directional GRU splices the two-directional GRU network together and is similar to Bi-LSTM in its ability to handle time-series data [38,17].
- **TCN:** The temporal convolutional network is a new algorithm for processing time series that reduces the serial processing complexity of RNNs [46].
- **Attention:** The attention mechanism is proposed by Vaswani et al. [43], which can quickly filter out high-value information from large amounts of information. Attention is popular in many fields such as machine translation and speech recognition.
- **Bi-GRU with attention:** This DNN structure is extracted from our proposed FusionNet with the same parameter settings.
- **Bi-GRU with GAP:** Global Average Pooling (GAP) is used to replace the attention layer to reduce the dimensionality of the output of Bi-GRU, to compare the performance differences of these two similar structures.

For the baseline statistical feature classifiers and word vector classification networks, we have run a series of pre-experiments on each classifier and selected the structures and parameters with the best classification performance. Each classifier will be represented by its main structure's symbols (e.g., Bi-GRU-based classifiers are referred to as Bi-GRU for

Table 5
Neural Network Training Setup.

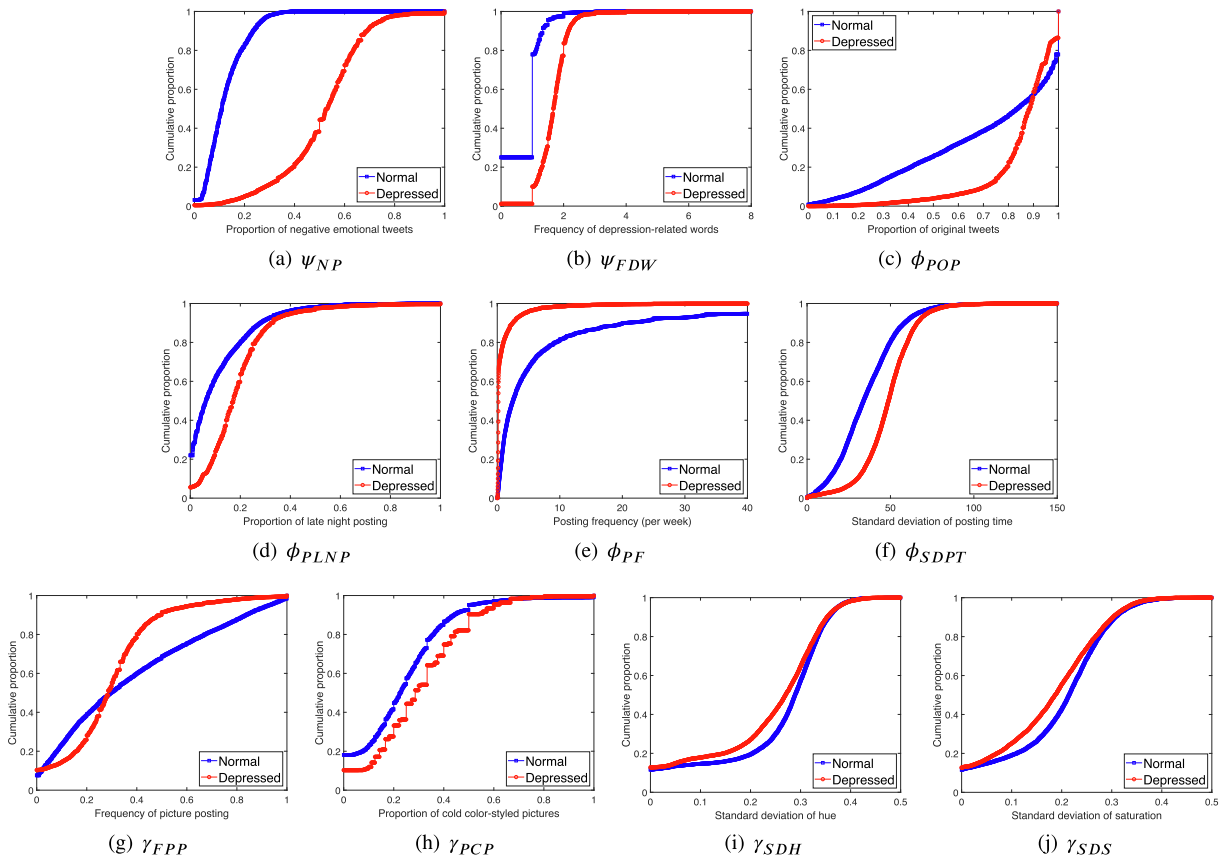| Item	Setup
Batch size	32
Epoch	80
Early Stopping	moni='val_acc', patience = 10
Check Point	moni='val_acc', mode='max'
Loss Function L_1, L_2	Categorical Crossentropy
Optimizer (for L_1)	NAdam (Init_lr = 3e-4)
Optimizer (for L_2)	NAdam (Init_lr = 1e-3)
FN $[\omega_1, \omega_2]$	[0.2, 1.0]

Fig. 4. The CDF Curves of Ten Statistical Features.

short). API default parameters are selected for both TML and DL classifiers that are not specifically described here. Moreover, separate experiments for the neural network structures of BP and Bi-GRU with attention are set to further demonstrate the superiority of FusionNet, which uses multitask learning to merge the two DNN structures.

The loss functions and callback settings for neural network training are given in Table 5.

4.2. Statistical Features Analysis

In this subsection, we evaluate the effectiveness of our newly proposed features and the modified features, as well as the other two unmodified features "Proportion of original tweets" and "Proportion of cold color-styled pictures" given in Table 4.2. Every individual feature will be evaluated in sub-sub Section 4.2.1, and feature groups will be evaluated in sub-sub Section 4.2.2. The features have already been categorized into three groups including text, pictures, and social behaviors.

4.2.1. Cumulative distribution analysis

We evaluate the significance of the features by comparing the feature distribution curve of normal and depressed users. The CDF curve for each feature is plotted in Fig. 4. Since there are no quantitative parameters for the CDF curve to describe the results, we only evaluate the degree of coincidence between the two types of user curves.
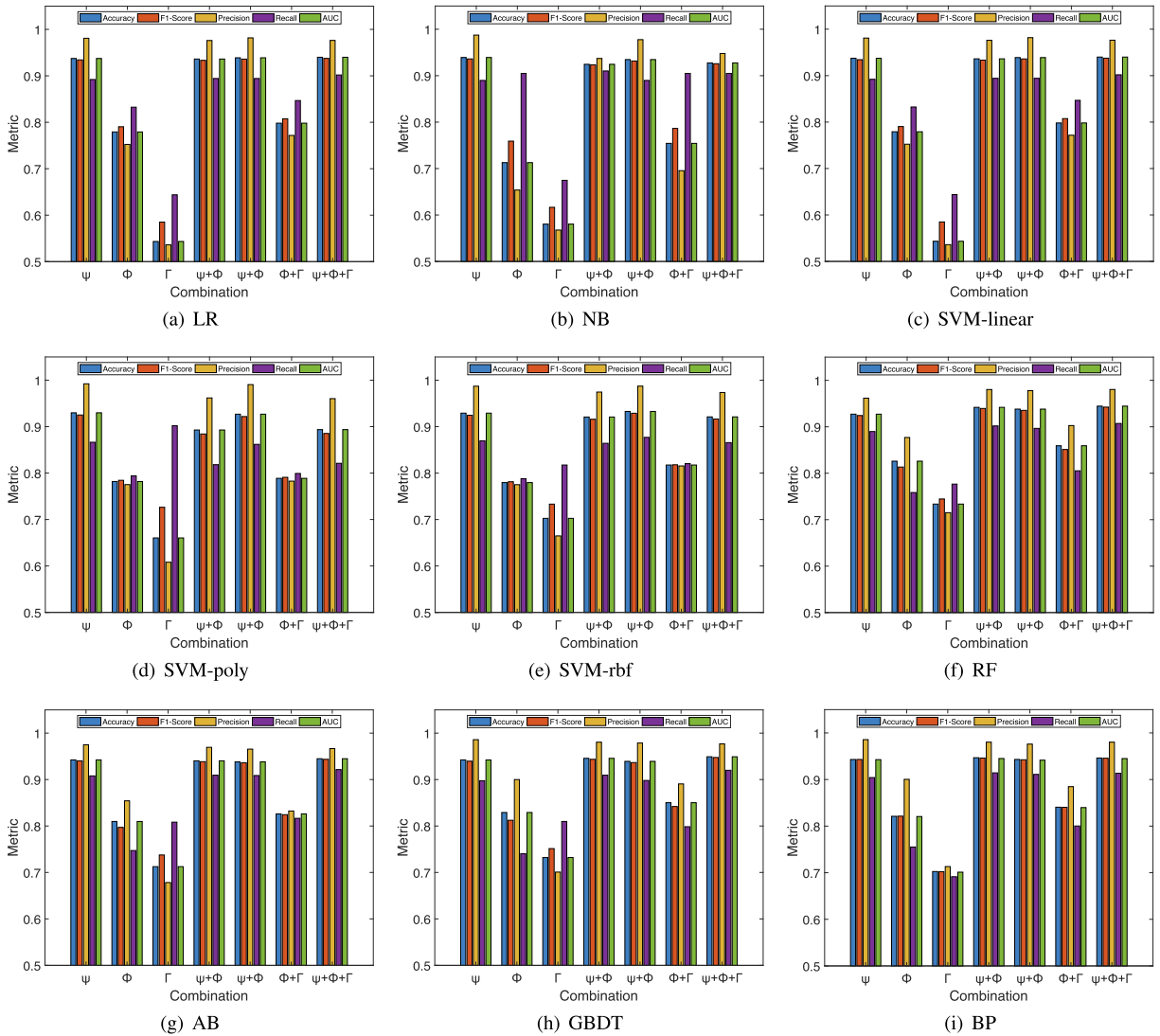
Fig. 5. The Contribution of Different Feature Group Combinations.

Among the features, ψ_{NP} has the highest distinction between the two types of users and the most significant difference in distribution. This phenomenon further demonstrates that text-based features play an important role in identifying depressed users on social networks. For our proposed features shown in Fig. 4(a), (e), (f), and (g), the curve of two types of users shows obvious separation and different trends, indicating there are significant differences in these features between the two types of users. For other features previously proposed in other work, the results are shown in Fig. 4. Among them, the CDF curves of Fig. 4(b), (c), (d), (h) indicate that these previously proposed features also show significant distinctions between the two types of users, which not only proves the effectiveness of these features but also indirectly certifies the feasibility of our proposed dataset WU3D. The CDF curves of Fig. 4(i) and (g) have a slightly higher degree of overlap, but they are also distinguishable features for the two types of users.

4.2.2. Feature Group Classification Contribution

In the previous Section 4.2, the contribution to the classification of each feature has been analyzed. In this part, We shift the concentration of analysis to feature "group", that is, considering the classification contribution of the text feature groups Ψ, the social behavior feature group Φ, and the picture feature group Γ. Baseline statistical feature classifiers are used to evaluate the contribution of different feature groups. We perform experiments on different combinations of features to determine the contribution to the classification task. The result is shown in Fig. 5.

The experimental result demonstrates that the classification performance of feature groups in each classifier is consistent with the results of the CDF curves, in which the text-based feature group Ψ contributed the most. The BP classifier has

(a) Δ - Time Consumption (Scaled)
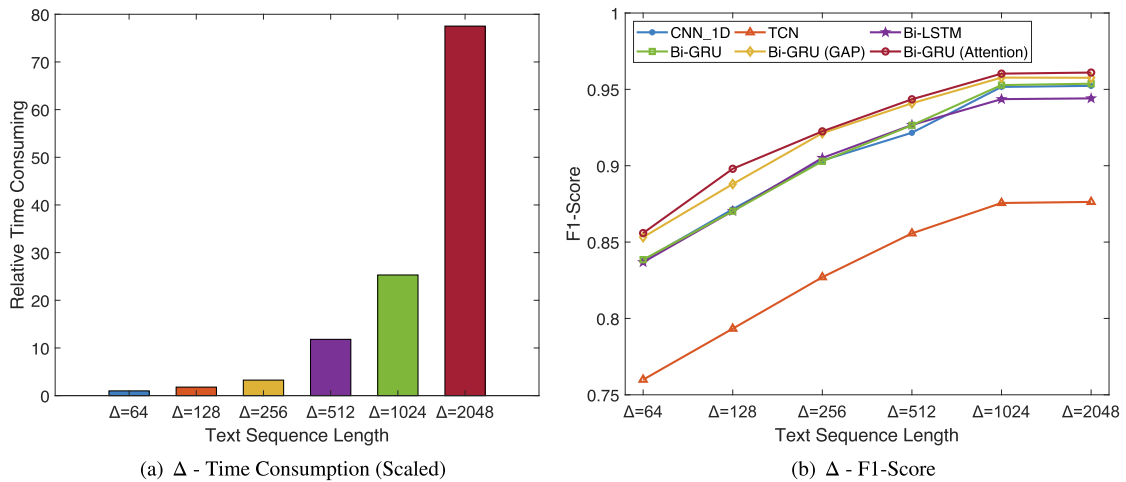
(b) Δ - F1-Score

Fig. 6. The Selection of Text Sequence Length.

already achieved a high F1-Score of 0.9431 when only text-based features are used. The contribution of the picture-based feature group Γ is relatively poor, only with the highest F1-Score of 0.7514 using the GBDT classifier. However, under the different combinations of features, the performance is improved at different levels for each group and each classifier. Especially, with the combination of all the feature groups ($\Psi + \Phi + \Gamma$), the GBDT classifier reports the highest F1-Score of 0.9465. GBDT and BP both achieved the highest performance metric for several rounds benefiting from the gradient descent related optimization method.

Therefore, it is concluded that all three types of feature groups can positively improve the performance of classification tasks, with the text-based features contributing the most.

4.3. Word Feature Analysis

According to Srivastava et al.[47], word features play a rather important role in mining the behaviors of online users. In our proposed method, the user text information consists of the user nickname, profile, and tweet text, all of which are concatenated to a long text sequence using Algorithm 2. However, due to the uncertainty of the total number of tweets and the number of words in a single tweet, it is necessary to explore the most effective embedding length of the long text sequence. In the meantime, since the pretrained Chinese XLNet has 12 layers of network structure, 768 hidden layers, and a total of 117 M parameters, it will also take considerable time to calculate the XLNet network outputs and to extract word vectors.

Therefore, in this section, we **(i)** select the optimal length of the text sequence by setting several different values of it, as well as **(ii)** examine the classification performance for the word feature (without statistical features). We record the time consumption of extracting word vectors using XLNet and the F1-Score of each classifier to explore the appropriate text sequence length, which should have reasonable embedding time consumption and a relatively high F1-Score.

4.3.1. Optimal Long Text Sequence Length Selection

According to our statistics, a user's tweet text may generally be longer than 32 Chinese characters. A too-short text length selection may result in anomaly premature truncation of the tweet text. Thus, our experiment starts with sequence length $\Delta = 64$, then we increase its value gradually. Six values of the text sequence length are selected for experiments, with Δ from 2^6 (64) to 2^{10} (1024). Then, we record the time consumption of embedding word vectors using XLNet and the classification F1-Score under different text sequence lengths.

The embedding time consumption of $\Delta = 64$ is selected as a base, normalizing the remaining groups to get the proportional values. Fig. 6(a) shows the relative time consumption of XLNet at each value of the sequence length Δ.

The result demonstrates that, as the text sequence length increases, the curve shows a nonlinear trend and the growth rate of the slope continues to increase. Specifically, the relative time of $\Delta = 512$ is 11.81, while the corresponding time of $\Delta = 1024$ is 25.31, and the corresponding time of $\Delta = 2048$ is 77.52. The time consumption increases by 114.31% and 206.28%, respectively. Therefore, when Δ increases by a binary exponential power, the time of embedding will increase at a significantly faster rate.

4.3.2. Word Feature Classification Contribution

The baseline word vector classification networks given in Section 4.1.4 are used to reflect the classification performance under different values of the long text sequence length Δ. Fig. 6(b) shows the F1-Score of each classifier under different Δ values. When Δ increases from 64 to 1024, the F1-Score of each classifier increases significantly. When Δ increases from

Table 6
Classification Performance of Target Classifiers.

Classifier	Accuracy	F1-Score	Precision	Recall
LR	0.9660	0.9655	0.9813	0.9502
NB	0.9555	0.9544	0.9779	0.9321
SVM-l	0.9623	0.9616	0.9796	0.9442
SVM-p	0.9562	0.9560	0.9604	0.9517
SVM-r	0.9600	0.9593	0.9773	0.9419
RF	0.9604	0.9597	0.9766	0.9434
AB	0.9649	0.9643	0.9820	0.9472
GBDT	0.9653	0.9647	0.9805	0.9494
FN	**0.9775**	**0.9772**	**0.9908**	**0.9639**

(a) Accuracy (b) F1-Score

(c) Precision (d) Recall
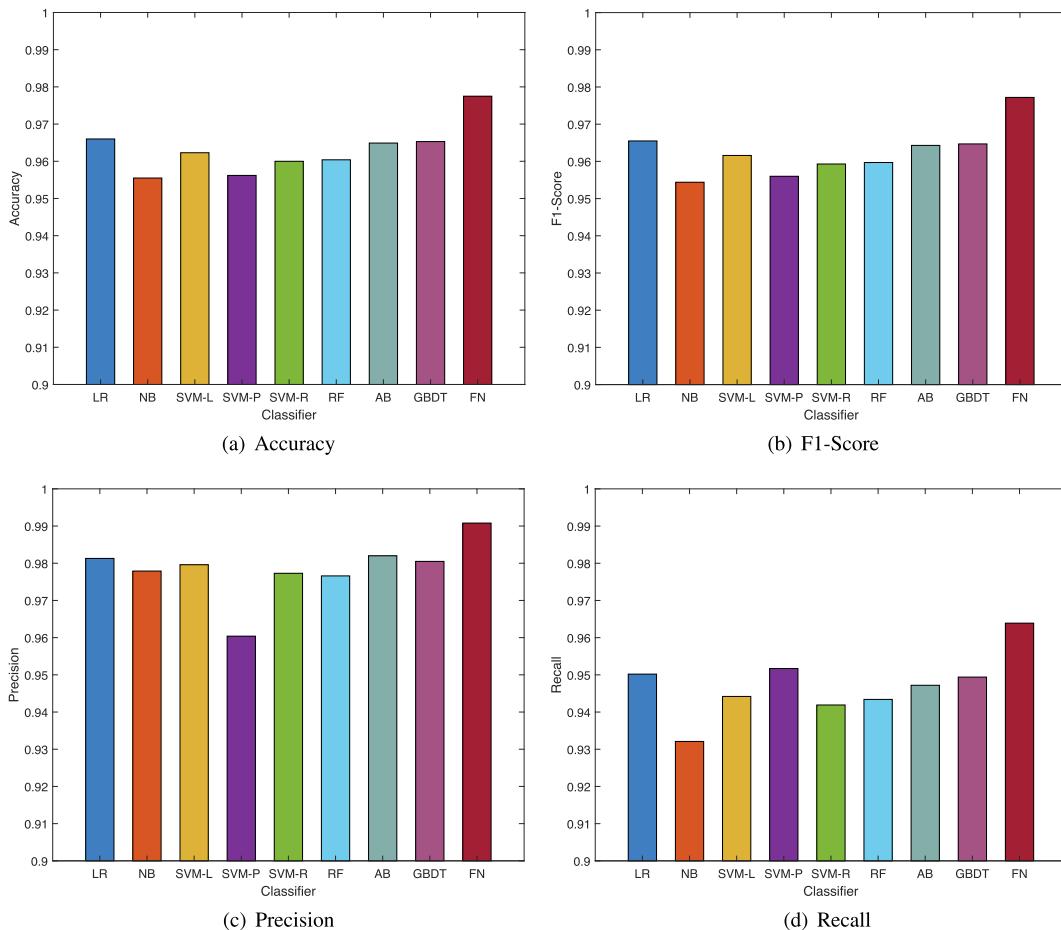
Fig. 7. Performance of the Target Classifiers.

1024 to 2048, the F1-Score almost no longer increases. Although the performance of a single attention structure is relatively poor, by adding it to other structures, the classification performance can be further improved. Thus, Bi-GRU with attention can better capture key information in long text sequences among the tested word vector classifiers. It achieves the highest classification performance and thus is used in our proposed FusionNet.

Combined with the two experiments, it is concluded that the length of text sequence at $\Delta = 1024$ has reached a performance bottleneck in this classification task, while the time computation of word embedding is relatively acceptable. Therefore, in subsequent experiments involving word vectors, we implement $\Delta = 1024$ as the default text sequence embedding length. Meanwhile, the highest F1-Score of more than 0.95 achieved by the Bi-GRU (Attention) attests to the classification significance of the word feature, indicating its importance in detecting depressed users on the OSN.
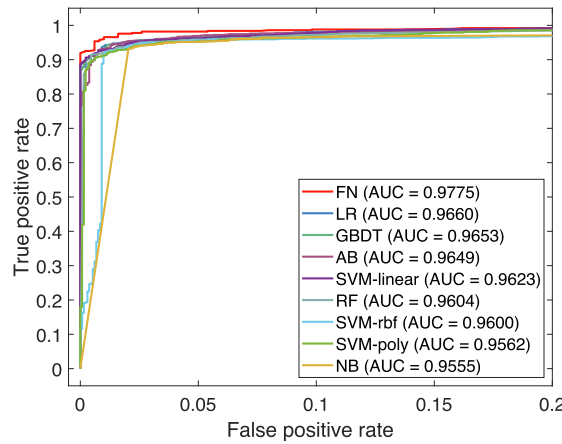
Fig. 8. ROC Curves of the Target Classifiers.

4.4. Performance of the Target Classifiers

To compare other classifiers with FusionNet, in this section, the output of the Bi-GRU with attention is extracted as the word feature, which is concatenated to the statistical feature. Then, this integrated feature vector is input into the baseline statistical feature classifiers to accomplish the classification task of depressed and normal users.

Thus, multimodel classifiers are constructed by both the word vector classification network and the baseline statistical feature classifier. Table 6 gives detailed metrics of all the target classifiers, while Fig. 7 visualizes these metrics.

In the performance experiment of the target classifiers, each classifier reaches an F1-Score above 0.95. Particularly, our proposed FusionNet achieves the highest F1-Score value of 0.9772. It also obtains the highest value under all the other metrics. Compared to the second-highest LR, FusionNet improves its F1-Score by 1.21%. Using the transfer learning scheme, the word feature is extracted as an input to the baseline statistical feature classifiers. However, transferring features through different classifiers may lead to a loss of information. Since multitask learning enables different tasks to share the same network structure and weights, it significantly reduces the loss of information caused by transfer learning, thus has better performance.

Furthermore, the ROC curves are given in Fig. 8. We obtain the classification probability values of the model output and plot these ROC curves by sampling the multipoint FPR and TPR. The result of the plot shows that all the curves are close to the upper left corner, which proves that all of the classifiers have excellent performance. The curve of FusionNet is closest to the upper left corner, achieving the best classification performance.

4.5. Robustness of Unbalanced Training Samples

In previous experiments, we used both depressed users and normal users with a proportion of 50% in dataset D_1, D_2, and D_3. However, in the real OSN environment, depressed users exist in a minority of the whole user community. Due to the difficulty of collecting depressed user data, it is hard to ensure that the training and optimization process of the classifier can fully guarantee a balanced data proportion.

Therefore, by changing the proportion of depressed user samples (denoted as ρ), we analyze the F1-Score fluctuations on the target classifiers to evaluate its robustness of training an unbalanced number of samples. Each classifier is treated as a group. For each group, we will test nine values of ρ from 0.1 to 0.9, with an interval of 0.1.

Here, we implement a new metric, namely the Intra-group F1-Score Variance (IFV), to calculate the variance of the F1-Score in each group. First, for each group, the mean value of the F1-Score is calculated and represented by $\overline{X}_{IF}$. The number of ρ values taken in each group is noted as λ. Therefore, the IFV metric is defined as:

$$IFV = \frac{1}{\lambda} \times \sum_{i=1}^{\lambda} \left(F1_i - \overline{X}_{IF} \right)^2 \tag{20}$$

Fig. 9(a) shows the F1-Score of each classifier under different ρ values. The experimental results show that when the proportion of data samples of depressed users and normal users are close to balance, the classifiers tend to achieve a higher F1-Score. It also shows that the proposed FusionNet achieves the highest F1-Score under every value of ρ. This testifies the effectiveness of the method of using multitask learning to integrate heterogeneous modalities for individual depression recognition and its excellent classification performance.

Table 7 and Fig. 9(b) shows the IFV metric of the target classifiers. Although LR once achieved a high F1-Score in the experiment of target classifiers, it has relatively low robustness for the unbalanced data due to the relatively poor decision
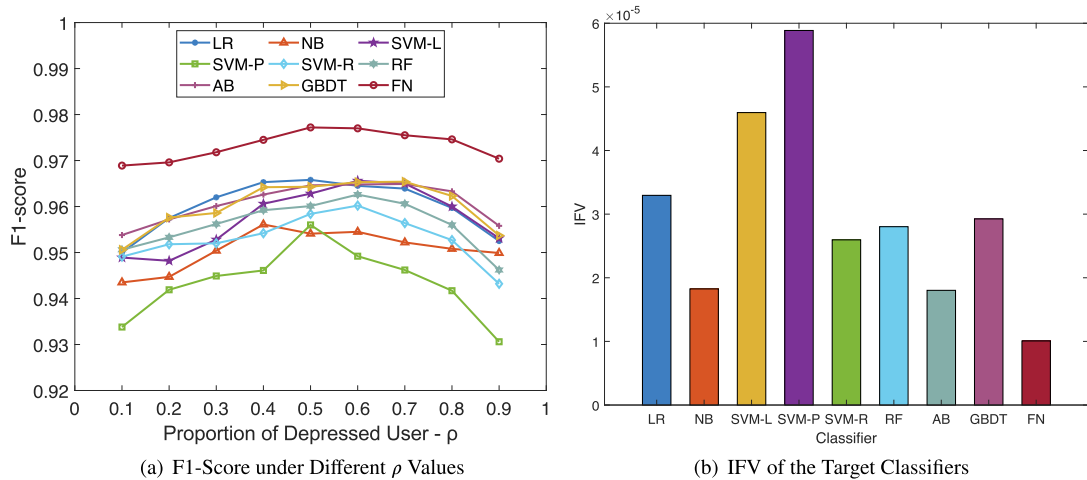
(a) F1-Score under Different ρ Values

(b) IFV of the Target Classifiers

Fig. 9. Unbalanced Training Samples.

Table 7
IFV of Target Classifiers.

Classifier	Intra-group F1-Score Variance
LR	3.30e-5
NB	1.82e-5
SVM-linear	4.60e-5
SVM-poly	5.89e-5
SVM-rbf	2.59e-5
RF	2.80e-5
AB	1.80e-5
GBDT	2.92e-5
FN (Proposed)	**1.01e-5**

Table 8
Time usage of model training and inferring.

Model Name	Training Time Usage (sec)	Inference Time Usage (sec)
CNN-1D	488	0.000607
Bi-LSTM	1061	0.001481
Bi-GRU	964	0.001250
TCN	1158	0.002495
Attention	4025	0.03868
FN (proposed)	948	0.001080

ability of the single classifier. With the kernel learning strategy, the two types of SVM classifiers have a better ability to fit the data, but the values of F1-Score also fluctuate obviously when the training samples are unbalanced. Ensemble classifiers including RF and GBDT obtain better robustness in the experiment. The IFV metric of our proposed FusionNet reaches the minimum value among the target classifiers, indicating that FusionNet has the best robustness, i.e. the most stable classification performance.

Benefiting from the strategic advantages of multitask learning, FusionNet successfully reduced the loss of feature information caused by transfer learning. Compared with the commonly used single-modal models in existing work, FusionNet has achieved a very distinguishable performance improvement and showed the best classification robustness when the training samples are unbalanced. In addition, it suggests that the adaptive learning rate of the optimizer Nadam also helps the proposed FusionNet find the global optimal solution more quickly.

4.6. Training and Inference Time Evaluation

In this part, neural network modules described in Section 4.1.4 are used to perform a training and inference time comparison evaluation. The time interval from when the training starts to when it reaches the total number of epochs specified in Table 5 or when an early stopper is triggered.

The results are given in Table 8, where the CNN-based classifier has both the minimum training time usage and inference time usage due to the parallel feature scanning characteristic of the convolutional kernel. All of the time usages of recurrent neural network-based classifiers are relatively slower but acceptable, except for the raw Attention (Transformers) model that takes too long to reach convergence.

As for FusionNet, although the time used for training and inference is not as fast as CNN-1D, it is still slightly faster than the other classifiers, which further indicates the benefit of multitasking training. The results provided here can only be treated as a rough baseline comparison since the time usage of model training and inference varies due to different network structures, training settings, and hardware (especially for CPU and GPU) performance.

5. Conclusion

In this work, we proposed a multitask learning-based approach to fuse heterogeneous modality information to predict depressed users on the OSN, enclosed with a published large-scale dataset, i.e. WU3D, aiming to facilitate the work of the depression detection and recognition task on the OSN, especially Sina Weibo.

Firstly, the size of the proposed dataset WU3D will be sufficient for subsequent researchers to complete further research. Also, the Web crawlers are published along with the dataset so that researchers may use them to collect more relevant user samples. Secondly, the experiments show that all of our proposed and modified statistical features can contribute positively to the classification task, in which we also justify that the text information is the most vital part of depression detection on the OSN.

Furthermore, we evaluate the performance of the pretrained model XLNet as the embedding model to solve downstream text classification tasks. It shows that when the appropriate embedding length is selected, XLNet has excellent performance and efficiency in handling long text sequences. Finally, we demonstrated the effectiveness and superiority of the method that fuses user information of heterogeneous modalities based on simultaneous multitasking for the recognition of depressed individuals on the OSN.

For future work, two directions will be further explored. **(i)** The size of the dataset will be further expanded. We will construct larger datasets for training and evaluating classifiers to achieve better generalization performance. **(ii)** The characteristics and behavior patterns of depressed users will be further analyzed. We will continue to work on this topic to propose more effective feature solutions for individual-level depression detection on the OSN.

CRediT authorship contribution statement

Yiding Wang: Methodology, Software, Formal analysis, Validation, Writing - original draft, Writing - review & editing. **Zhenyi Wang:** Methodology, Writing - original draft, Writing - review & editing, Visualization. **Chenghao Li:** Methodology, Software, Validation, Investigation. **Yilin Zhang:** Investigation, Validation, Writing - original draft. **Haizhou Wang:** Conceptualization, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

References

[1] Shailendra Rathore, Pradip Kumar Sharma, Vincenzo Loia, Young-Sik Jeong, and Jong Hyuk Park. Social network security: Issues, challenges, threats, and solutions. Information Sciences, 421:43–69, 2017.
[2] Mengfan Yao, Charalampos Chelmis, and Daphney-Stavroula Zois. Cyberbullying ends here: Towards robust detection of cyberbullying in social media. In The World Wide Web Conference, pages 3427–3433, 2019.
[3] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, pages 128–137, Cambridge, MA, USA, Jul 2013.

[4] Hong-Han Shuai, Chih-Ya Shen, De-Nian Yang, Yi-Feng Carol Lan, Wang-Chien Lee, S Yu Philip, and Ming-Syan Chen. A comprehensive study on social network mental disorders detection via online social media mining. IEEE Transactions on Knowledge and Data Engineering, 30(7), 1212–1225, 2017.

[5] Despoina Chatzakou, Ilias Leontiadis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, Athena Vakali, Nicolas Kourtellis, Detecting cyberbullying and cyberaggression in social media, ACM Transactions on the Web (TWEB) 13 (3) (2019) 1–51.

[6] Tao Wang, Markus Brede, Antonella Ianni, and Emmanouil Mentzakis. Detecting and characterizing eating-disorder communities on social media. In Proceedings of the 10th ACM International Conference on Web Search and Data Mining, pages 91–100, Cambridge, UK, Feb 2017.

[7] Christian Karmen, Robert C Hsiung, and Thomas Wetter. Screening internet forum participants for depression symptoms by assembling and enhancing multiple nlp methods. Computer Methods and Programs in Biomedicine, 120(1), 27–36, Jun 2015.

[8] Elvis Saravia, Chun-Hao Chang, Renaud Jollet De Lorenzo, and Yi-Shin Chen. Midas: Mental illness detection and analysis via social media. In Proceedings of the 8th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pages 1418–1421, San Francisco, CA, USA, Aug 2016.

[9] Lu. Qingpeng Zhang, Siyang Gao Zhong, Xiaoming Li, Optimizing hiv interventions for multiplex social networks via partition-based random search, IEEE Transactions on Cybernetics 48 (12) (Dec 2018) 3411–3419.

[10] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. User-level psychological stress detection from social media using deep neural network. In Proceedings of the 22nd ACM International Conference on Multimedia, pages 507–516, Orlando, FL, USA, Nov 2014.

[11] Sairam Balani and Munmun De Choudhury. Detecting and characterizing mental health related self-disclosure in social media. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pages 1373–1378, Seoul, Republic of Korea, Apr 2015.

[12] Qijin Cheng, Tim MH Li, Chi-Leung Kwok, Tingshao Zhu, and Paul SF Yip. Assessing suicide risk and emotional distress in chinese social media: A text mining and machine learning study. Journal of Medical Internet Research, 19(7):e243, Jul 2017.

[13] Qi Gao, Fabian Abel, Geert-Jan Houben, and Yong Yu. A comparative study of users' microblogging behavior on sina weibo and twitter. In Proceedings of the 20th ACM International Conference on User Modeling, Adaptation, and Personalization, pages 88–101, Berlin, Heidelberg, July 2012. Springer.

[14] Stefan Weinmann, Markus Koesters, Mental health service provision in low and middle-income countries: recent developments, Current Opinion in Psychiatry 29 (4) (2016) 270–275.

[15] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. Depression detection via harvesting social media: A multimodal dictionary learning solution. In Proceedings of the 26th ACM International Joint Conference on Artificial Intelligence, pages 3838–3844, Melbourne, Australia, Aug 2017.

[16] Tiancheng Shen, Jia Jia, Guangyao Shen, Fuli Feng, Xiangnan He, Huanbo Luan, Jie Tang, Thanassis Tiropanis, Tat Seng Chua, and Wendy Hall. Cross-domain depression detection via harvesting social media. In Proceedings of the 27th ACM International Joint Conference on Artificial Intelligence, pages 1611–1617, Stockholm, Sweden, Jul 2018.

[17] Tao Gui, Liang Zhu, Qi Zhang, Minlong Peng, Xu Zhou, Keyu Ding, and Zhigang Chen. Cooperative multimodal approach to depression detection in twitter. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence, pages 110–117, Honolulu, HI, USA, Jan 2019.

[18] Soumitra Ghosh, Asif Ekbal, Pushpak Bhattacharyya, What does your bio say? inferring twitter users' depression status from multimodal profile information using deep learning, IEEE Transactions on Computational Social Systems (2021).

[19] Fatma Najar, Nizar Bouguila, Exact fisher information of generalized dirichlet multinomial distribution for count data modeling, Information Sciences 586 (2022) 688–703.

[20] Yoshihiko Suhara, Yinzhan Xu, and Alex'Sandy' Pentland. Deepmood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In Proceedings of the 26th ACM International Conference on World Wide Web, pages 715–724, Perth, Australia, Apr 2017.

[21] Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. Depaudionet: An efficient deep model for audio based depression classification. In Proceedings of the 6th ACM International Workshop on Audio/Visual Emotion Challenge, pages 35–42, Amsterdam, The Netherlands, Oct 2016.

[22] Le Yang, Dongmei Jiang, Xiaohan Xia, Ercheng Pei, Meshia Cédric Oveneke, and Hichem Sahli. Multimodal measurement of depression using deep learning models. In Proceedings of the 7th Annual ACM Workshop on Audio/Visual Emotion Challenge, pages 53–59, Mountain View, CA, USA, Oct 2017.

[23] Hamdi Dibeklioğlu, Zakia Hammal, Jeffrey F Cohn, Dynamic multimodal measurement of depression severity using deep autoencoding, IEEE Journal of Biomedical and Health Informatics 22 (2) (Mar 2017) 525–536.

[24] Yu. Zhu, Yuanyuan Shang, Zhuhong Shao, Guodong Guo, Automated depression diagnosis based on deep networks to encode facial appearance and dynamics, IEEE Transactions on Affective Computing 9 (4) (Oct 2018) 578–584.

[25] Bart Desmet, Véronique Hoste, Online suicide prevention through optimised text classification, Information Sciences 439 (2018) 61–78.

[26] Hamad Zogan, Imran Razzak, Shoaib Jameel, Xu. Guandong, Depressionnet: learning multi-modalities with user post summarization for depression detection on social media, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 133–142.

[27] Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, Nazli Goharian, Smhd: a large-scale resource for exploring online language usage for multiple mental health conditions, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1485–1497.

[28] Mandar Deshpande and Vignesh Rao. Depression detection using emotion artificial intelligence. In Proceedings of the 19th IEEE International Conference on Intelligent Sustainable Systems, pages 858–862, Palladam, Tirupur, India, Dec 2017.

[29] Nafiz Al Asad, Md Appel Mahmud Pranto, Sadia Afreen, and Md Maynul Islam. Depression detection by analyzing social media posts of user. In Proceedings of 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems, pages 13–17, Dhaka, Bangladesh, Apr 2019.

[30] Marcel Trotzek, Sven Koitka, Christoph M Friedrich, Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences, IEEE Transactions on Knowledge and Data Engineering 32 (3) (Mar 2020) 588–601.

[31] Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, Zhana Bao, A depression detection model based on sentiment analysis in micro-blog social network, in: Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2013, pp. 201–213.

[32] Raza Ul Mustafa, Noman Ashraf, Fahad Shabbir Ahmed, Javed Ferzund, Basit Shahzad, and Alexander Gelbukh. A multiclass depression detection in social media based on sentiment analysis. In Proceedings of the 17th IEEE International Conference on Information Technology—New Generations, pages 659–662, Las Vegas, NV, USA, Apr 2020. Springer.

[33] Maxim Stankevich, Vadim Isakov, Dmitry Devyatkin, and Ivan Smirnov. Feature engineering for depression detection in social media. In Proceedings of the 7th IEEE International Conference on Pattern Recognition Applications and Methods, pages 426–431, Funchal, Madeira, Portugal, Jan 2018.

[34] Feiran Huang, Xiaoming Zhang, Jie Xu, Zhonghua Zhao, and Zhoujun Li. Multimodal learning of social image representation by exploiting social relations. IEEE Transactions on Cybernetics, pages 1–13, Mar 2019.

[35] Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, and Diana Inkpen. Deep learning for depression detection of twitter users. In Proceedings of the 5th Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic, pages 88–97, New Orleans, LA, USA, Jun 2018.

[36] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th ACM International Conference on Neural Information Processing Systems, pages 3111–3119, Lake Tahoe, NV, USA, Dec 2013.

[37] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Computation 9 (8) (Nov 1997) 1735–1780.

[38] Farig Sadeque, Xu. Dongfang, Steven Bethard, Measuring the latency of depression detection in social media, in: Proceedings of the 11th ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, Feb, 2018, pp. 495–503.

[39] Chenhao Lin, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. Sensemood: Depression detection on social media. In Proceedings of the 28th ACM International Conference on Multimedia Retrieval, pages 407–411, Dublin, Ireland, Jun 2020.

[40] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1, pages 4171–4186, 2019.

[41] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. In Advances in neural information processing systems, pages 5753–5763, Vancouver, Canada, Dec 2019.

[42] Pasquale De Meo, Emilio Ferrara, Domenico Rosaci, and Giuseppe ML Sarné. Trust and compactness in social network groups. IEEE Transactions on Cybernetics, 45(2), 205–216, Feb 2015.

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Proceedings of the 31st ACM International Conference on Neural Information Processing Systems, pages 5998–6008, Long Beach, CA, USA, Dec 2017.

[44] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd ACM International Conference on International Conference on Machine Learning, pages 448–456, Lille, France, Jul 2015.

[45] Han Zhang, Shaoqi Sun, Hu. Yongjin, Junxiu Liu, Yuanbo Guo, Sentiment classification for chinese text based on interactive multitask learning, IEEE Access 8 (2020) 129626–129635.

[46] Ashutosh Pandey and DeLiang Wang. Tcnn: Temporal convolutional neural network for real-time speech enhancement in the time domain. In Proceedings of the 44th IEEE International Conference on Acoustics, Speech and Signal Processing, pages 6875–6879, Brighton, UK, May 2019.

[47] Deepesh Kumar Srivastava and Basav Roychoudhury. Words are important: A textual content based identity resolution scheme across multiple online social networks. Knowledge-Based Systems, 195:105624, 2020.

[48] Yiding Wang, Zhenyi Wang, Chenghao Li, Yilin Zhang, Haizhou Wang. A multimodal feature fusion-based method for individual depression detection on sina weibo. In Proceedings of 39th IEEE International Performance Computing and Communications Conference, pages 1-8, Austin, Texas, USA, November 2020.