



Psychological Education Health Assessment Problems Based on Improved Constructive Neural Network

Yang Li^{1*}, Jia ze Li^{2*}, Qi Fan³, Xin Li¹ and Zhihong Wang¹

¹ School of Administration, Nanjing Forest Police College, Nanjing, China, ² School of Foreign Studies, Nanjing University, Nanjing, China, ³ Institute of Mental Health, Nanjing Xiaozhuang University, Nanjing, China

In order to better assess the mental health status, combining online text data and considering the problems of lexicon sparsity and small lexicon size in feature statistics of word frequency of the traditional linguistic inquiry and word count (LIWC) dictionary, and combining the advantages of constructive neural network (CNN) convolutional neural network in contextual semantic extraction, a CNN-based mental health assessment method is proposed and evaluated with the measurement indicators in CLPsych2017. The results showed that the results obtained from the mental health assessment by CNN were superior in all indicators, in which F1 = 0.51 and ACC = 0.69. Meanwhile, ACC evaluated by FastText, CNN, and CNN + Word2Vec were 0.66, 0.67, 0.67, and F1 were 0.37, 0.47, and 0.49, respectively, which indicates the use of CNN in mental health assessment has feasibility.

Keywords: LIWC dictionary, CNN, mental health assessment, assessment, psychological

OPEN ACCESS

Edited by:

Baiyuan Ding,
National University of Defense
Technology, China

Reviewed by:

Jun Geng,
Shandong University of Finance
and Economics, China
Wei Wang,
Shanxi Normal University, China

*Correspondence:

Yang Li
112121@nfdc.edu.cn.comand
Jia ze Li
mf1909034@smail.nju.edu.cn

Specialty section:

This article was submitted to
Health Psychology,
a section of the journal
Frontiers in Psychology

Received: 13 May 2022

Accepted: 20 June 2022

Published: 02 August 2022

Citation:

Li Y, Li Jz, Fan Q, Li X and Wang Z
(2022) Psychological Education
Health Assessment Problems Based
on Improved Constructive Neural
Network. *Front. Psychol.* 13:943146.
doi: 10.3389/fpsyg.2022.943146

INTRODUCTION

Driven by the development of the mobile Internet, people have developed a variety of social networking platforms, the more influential ones include WeChat, Twitter, QQ, online forums, Facebook, Renren, Weibo, etc. They have become essential tools for people's working lives. In terms of functional design, most social networking platforms support functions such as online communication, posting reviews, expressing emotions, and recording life. In particular, more and more users tend to express personal emotional thoughts or communicate their psychological state on social networking platforms, through which they can seek help when necessary. Therefore, social networking platforms have also become an important tool for studying mental health problems. By studying the information on social networking platforms, we can identify whether users have mental health problems such as self-harm and depression, and we can further trace the underlying causes of mental health problems. As deep learning algorithms have matured, the field of natural language processing has begun to apply deep learning algorithms more widely and has made positive progress. For example, Nanni Loris et al. used convolutional neural networks to classify sentiments of sentences and determine the positive and negative polarity of sentences by extracting n-gram features through multiple channels. The advantage of constructive neural network (CNN) is that it can capture the features of local sequence information, but it also has the obvious disadvantage that its convolutional kernel size is fixed, and thus it cannot model longer sequence information (Liu et al., 2021; Nanni et al., 2021; Wang et al., 2021). In this regard, Zhou used the Bi-LSTM structure to replace the convolution and pooling process of Text CNN to express the contextual information of text and extracted the valuable information in text by two-dimensional convolution and pooling in two dimensions, and

this method effectively improved the accuracy of text classification for various application scenarios such as news classification, question classification, and sentiment classification (Zhou et al., 2016; Shaikh et al., 2021). Liu used a mathematical model based on deep learning algorithms for the personality prediction of social network users (Joulin et al., 2016; Cui and Zang, 2021; Kanekar et al., 2021; Ren et al., 2021). Yates analyzed the process of detecting mental health problems of online forum users and concluded that feature construction is the most tedious task in the detection process, for which convolutional neural networks were invoked for feature construction, which greatly simplified the workflow (Yates et al., 1709; He et al., 2021). This study believes that the automatic assessment of mental health can be strengthened by combining the analysis of relevant data on social platforms. The automatic assessment of mental health mainly includes the detection and analysis of depression and suicide intention. In addition, the analysis of personality closely related to mental health problems and the analysis of reasons for emotional changes are helpful to the automatic assessment of mental health. In view of the need for psychological education health assessment, a set of mental health assessment models based on the linguistic inquiry and word count (LIWC) dictionary and CNN is constructed, and the feasibility of this model is demonstrated.

BASIC METHODS

Basic Principles of Constructive Neural Network

The Structure of Constructive Neural Network

In the CNN structure, the first layer is the input stage, the second and third layers are the first feature extraction stage, the fourth and fifth layers are the second feature extraction stage, and the sixth and seventh layers are the classification stage, in which the classifier is a two-layer fully connected neural network. The CNN architecture is shown in **Figure 1** below.

In terms of construction, CNN contains seven layers, namely, the input layer, convolutional layer, downsampling layer, fully connected layer, and output layer. Generally, there are one to three feature extraction stages in the CNN implementation process, after which the extracted feature data is delivered to the classifier. The 7-layer network structure shown in **Figure 1** above contains 2 feature extraction stages corresponding to a 4-layer network, while the classification stage occupies a 2-layer network, and this architecture design can meet the classification requirements in most application scenarios.

The first layer is the input layer, which only accepts images of standard size N^*N .

The second and fourth layers are convolutional layers, which perform convolutional operations on the previous layer to obtain multiple feature maps. The upper layer performs convolutional operations with multiple trainable convolutional kernels of size $k*k$, plus a bias, and acquires a feature map of size $(N - k + 1)*(N - k + 1)$ and the same number as the convolutional kernels after an excitation function. The equation for the above process is as follows (Caldelli

et al., 2021; Khaydarova et al., 2021; Kumar and Hati, 2021; Moccia et al., 2021; Shi et al., 2021; Szajna et al., 2021):

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right) \quad (1)$$

In which, k denotes the convolution kernel, l denotes the number of layers, b denotes the bias, and M_j denotes the j th feature map.

The third and fifth layers are down-sampling layers, which sub-sample each feature map of the previous convolutional layer. This process reduces the resolution of the feature maps, effectively shortening the data processing time while ensuring that valuable information is not lost. The maximum or weighted summation operation is performed on the neighborhoods of size $m*m$ within each feature map, and a bias is added to obtain the same number of feature maps of size $(N - k + 1)/m$ after an excitation function as at the time of input. The equation for the above process is as follows:

$$x_j^l = f[\beta_j^l \text{down}(x_j^{l-1}) + b_j^l] \quad (2)$$

In which, down represents the downsampling function, and β, b represent the feature maps of each output, respectively.

The sixth layer is the classifier, which connects the feature maps extracted from the previous layer to build feature vectors for learning.

The seventh layer is the output layer, which outputs the corresponding tags.

Constructive Neural Network Training Process

Input

Build training sample set $X = \{x_i\}$, label $T = \{t\}_{i=1}$.

Initialization

Set the number of hidden layers, the learning rate, the number of neurons in the convolutional and downsampling layers, initialize the connection weights and biases between the networks, and choose a smaller random number.

Note: The training samples are denoted as $\langle x, t \rangle$, where x denotes the network input value vector, and t denotes the target output value.

The Training Process

Propagate the input forward along with the network. One sample is pulled from within the dataset and imported into the CNN, l denotes the current layer, and the output of the current layer is formulated as:

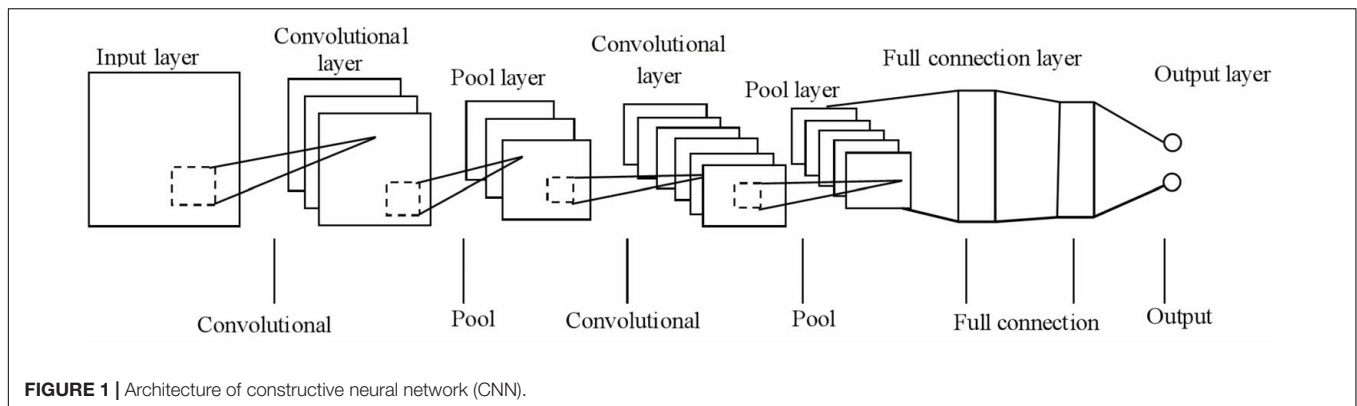
$$x^l = f(u^l), \text{ with } u^l = w^l x^{l-1} + b^l \quad (3)$$

In which, f denotes inputting activation function, b denotes the bias, and w denotes the weight matrix.

Each layer processes and passes the data until the final layer outputs the data as follows:

$$\text{output} = f_n(\dots(f_2(f_1(x_i w_1) w_2)) \dots w_n) \quad (4)$$

Propagate the error backward along with the network.



Sum the output units in the network and evaluate the error by referencing the squared error cost function, and the equation is as follows:

$$E_d = \frac{1}{2} \sum_k^n (t_k - o_k)^2 \quad (5)$$

In which, n denotes the set of each output unit, E_d denotes the error of the training sample d , o_k denotes the output value of unit k in the training sample d , and t_k denotes the target value of unit k for the training sample d .

For each output cell k , the error term δ_k for the current layer l is:

$$\delta_k \leftarrow f'(u^l)(t_k o_k) \quad (6)$$

Solve the error term δ_k for each hidden layer cell in the network as follows:

$$\delta_h \leftarrow f'(u^l) \sum w_{hk} \delta_k \quad (7)$$

The arithmetic of δ_k in the above equation is $\delta^l = w^{l+1} \delta^{l+1} \bullet f'(u^l)$.

Update each network weight w_{μ} as follows:

$$w_{\mu} \leftarrow w_{\mu} + \Delta w_{\mu} \quad (8)$$

$$\Delta w_{\mu} = -\eta \frac{\partial E_d}{\partial w_{ji}} = \eta \delta_j x_{ji} \quad (9)$$

Linguistic Inquiry and Word Count Dictionary

LIWC is one of nature language processing (NLP) technologies that can analyze various texts through computer programs. It can make a quantitative analysis of the text content and calculate the different categories of words in the imported text file, especially for psychological words, which can calculate the percentage of use of such words in the entire text. After more than a decade of development, revision, and expansion, LIWC has become increasingly stable, and it mainly includes two parts, namely, the program body and the dictionary. Among them, the core is the dictionary, and the words in the text can be compared with the dictionary one by one, and then the word frequency

results of various words are output. The current LIWC has a total of 80-word categories and about 4,500 words.

For existing LIWC-based text analysis methods, although this method is simple and effective, it does not consider the order between words and the contextual information of words, which will lead to large errors in text analysis. For example, when counting emotional words, the LIWC dictionary can only judge the category of emotional words but cannot accurately judge the specific emotional polarity of them. In addition, negative words and degree words in the context of emotion words cannot be accurately analyzed.

AUTOMATIC ASSESSMENT MODEL OF MENTAL HEALTH BASED ON CONSTRUCTIVE NEURAL NETWORK

In order to make full use of psychological feature information in LIWC, a set of convolutional neural networks based on LIWC (LIWC-CNN) is constructed (Tanana et al., 2021). First of all, the LIWC dictionary is used to determine the distinct degree of a certain word class on the mental health of posting users. Then the convolutional neural network is guided to extract the key psychological features in posts, which is used to train the automatic assessment model of mental health. In addition, LIWC can accurately represent the mental health status, and CNN has explained the context information of the words in the post (Tay, 2021), so the combination of the two can accurately assess the mental health status reflected in the post.

Weight Calculation of Word Classes

In 80 word classes in the LIWC dictionary, there are 16-word classes belonging to the descriptive word classes without actual semantics, such as the number of words per sentence, the total number of words, and so on, which are not the focus of this manuscript. The paper only studies the 64 semantically valid word classes in LIWC, including social process words, cognitive process words, negative words, personal pronouns words, and others. Considering that words in different parts of speech play different roles in identifying users' mental health status, this paper determines the weight of each word class through standard

deviation analysis, and the implementation process is as follows (Pandi et al., 2018):

Step 1: determine the word frequency of each word class. The frequency of different word classes in the posts of green, red, crisis, amber, and others is analyzed by using the language feature screening procedure.

The following linguistic features are screened by means of SD analysis:

Solve for the frequency of the word class. The frequency of occurrence of each word class in the LIWC dictionary in a sample of different mental health categories was solved respectively and expressed in mathematical language as follows: for sample $\{p_i = \langle t_1^i, t_2^i, \dots, t_k^i, \dots, t_{n_i}^i \rangle, i = 1, \dots, |D|\}$ in the training set, $|D|$ denotes the number of the samples, n_i denotes the text length of the sample p_i , t_k^i denotes the k th word in the sample p_i , the frequency of occurrence $TF(l, c)$ of word class l in sample class $c \in \{crisis, red, amber, green\}$ is as follows:

$$TF(l, c) = \frac{\sum_{i=1}^{|D|} \sum_{j=1}^{n_i} 1_l(t_j^i) \cdot 1_c(p^i)}{\sum_{i=1}^{|D|} n_i \cdot 1_c(p^i)} \quad (10)$$

$$A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Solve for the SD of the word class. The results of solving for the standard deviation of the frequency of each word class in the sample across mental health categories highlight the differences between the word classes, and larger standard deviations imply that the word class is effective in distinguishing the mental health category to which the sample belongs. Considering the significant differences in the frequency of occurrence of different word classes in the dataset, the peak frequency of occurrence of the word class in each mental health category was used for normalization when solving for the standard deviation. The equation is as follows:

$$\delta_l^2 = \sum_c [TF(l, c) - \mu]^2 \quad (11)$$

$$TF'(l, c) = \frac{TF(l, c)}{\max[TF(l, c)]} \quad (12)$$

$$\mu = \frac{\sum_c TF'(l, c)}{\sum_c 1} \quad (13)$$

In which, μ denotes the average of the normalized frequency of word class l over the four types of samples, and $TF'(l, c)$ denotes the normalization of $TF(l, c)$. δ_l^2 represents the standard deviation of the word l belonging to the speech class t .

Word class screening. The top- k LIWC word classes with the highest δ_1 are selected as linguistic features. According to the experimental results, the best classification results can be harvested when $k = 28$ is set, so the 28 most discriminative word classes from the 64 LIWC word classes are selected as linguistic features. The feature values corresponding to word class l in the

linguistic features of the training sample or test sample p_i are $LGF(l, p_i)$ as follows:

$$LGF(l, p_i) = \frac{1}{n_i} \cdot \sum_{j=1}^{n_i} 1_l(t_j^i) \quad (14)$$

Step 2: Solve the SD of the word frequency of each word class. The word frequency of different parts of speech in four posts is obtained in step one, then the SD of each word class word frequency is calculated. At the same time, the maximum value is used to perform normalization processing to evaluate the differences in LIWC word class in different mental health categories. That is to say, with the increase of the standard deviation, the differentiation degree of this word class for the mental health category is greater. The top 10 LIWC word classes with SD δ_l are listed in **Table 1**.

Step 3: Adjust the weight of words in the post. According to the attribution of words in the post in the LIWC category, the weight of words is adjusted to give a higher weight to those words that can clearly distinguish the post category. If the word in the post does not exist in LIWC, the weight of the word can be set to 1. The formula for the weight a_i of the word i is as follows:

$$a_i = \begin{cases} \frac{\exp(\delta_{l,i})}{\sum_l \exp(\delta_{l,i})} + 1 & \text{if } i \in l \\ 1, & \text{if } i \notin l \end{cases} \quad (15)$$

where, $\delta_{l,i}$ refers to the SD of the word i belonging to the word class l . If the word i belongs to more than one word class, the value with the maximum SD of the word class should be selected.

The Construction of the Mental Health Assessment Model Based on Constructive Neural Network

As a typical deep learning algorithm, CNN integrates convolution kernel operations, which can effectively increase the model depth. The traditional CNN consists of a pooling layer, a convolutional layer, and other parts. Here, the convolutional layer uses special convolution kernel operations to determine the local feature map of the text, and the pooling layer is used to locate effective features. To realize the automatic assessment of users' mental health, the advantages of Kim architecture and the LIWC dictionary are comprehensively applied to construct the LIWC-CNN model. The LIWC-CNN model architecture is shown in **Figure 2**.

Combined with the figure above, the LIWC-CNN model contains four parts, specifically:

Input layer: Decompose the post into multiple sequences with n words. If the number of words is less than n , it needs to use special markers to complete the post. If the number of words is greater than N , it needs to intercept the front n words as a substitute. Then, the Word2Vec model is adopted to train the word vector model, so as to represent the post in the form of a word vector matrix X , namely, $X = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i, \dots, \vec{x}_n]$, $\vec{x}_i \in R^{1 \times d}$. \vec{x}_i refers to the vector of the i th word, and d refers to the dimension of the word vector.

TABLE 1 | Top 10 speech classes with standard deviation of word frequency.

Ranking	Word class	Standard deviation	Ranking	Word class	Standard deviation
1	Relig_c	0.4269	6	Money_c	0.3304
2	Friends_c	0.3965	7	Assent_c	0.3237
3	Death_c	0.3867	8	Anger_c	0.2796
4	You_c	0.3658	9	Ingest_c	0.2823
5	We_c	0.3469	10	Sexual_c	0.2502

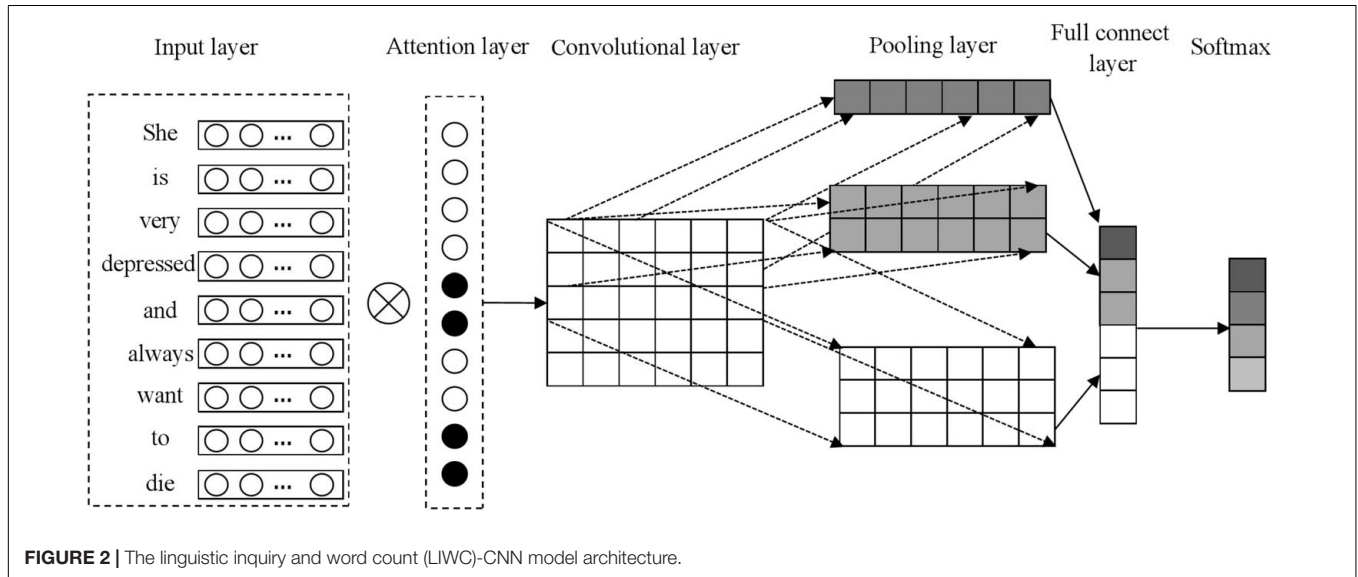


FIGURE 2 | The linguistic inquiry and word count (LIWC)-CNN model architecture.

Attention layer: In combination with the analysis results, the attention layer further improves the recognition performance of posts by utilizing the ability of the LIWC word class to distinguish samples. After determining the weight of words in the post, it is weighted with the word vector matrix X of the input layer to construct a new word vector matrix M . The formula is as follows:

$$\begin{aligned}
 M &= \vec{\alpha} \cdot \vec{X} \\
 \vec{a} &= \langle a_1, \dots, a_i, \dots, a_n \rangle \\
 \vec{X} &= \langle \vec{x}_1, \vec{x}_2, \dots, \vec{x}_n \rangle
 \end{aligned}
 \tag{16}$$

Convolutional layer: Utilize the convolutional kernel of each window to obtain the local semantics of the text sequence, so as to realize feature detection. In operation, the convolution kernel of sliding window h is selected to obtain the local semantics of sentences in posts, and the word vector matrix is constructed by splicing semantic information. Decompose the sentence with n words into $\langle x_{0:h-1}, x_{1:h-1}, \dots, x_{i:1+h-1}, \dots, x_{n-h+1:h} \rangle$, and apply the following Equation 16 to perform a convolution operation on each component x . At the same time, call the activation function $\delta(x)$ to realize the de-linearization process, so as to solve the feature vector \vec{g}_i . Finally, the feature matrix G that can present global semantics can be obtained by splicing the feature vector, as follows:

$$\vec{g}_f = \partial[f(w \cdot X_{j:j+h-1} + b)]
 \tag{17}$$

$$\delta(x) = \frac{1}{1 + e^{-x}}
 \tag{18}$$

$$G = \vec{g}_1 \oplus \vec{g}_2 \oplus \dots \oplus \vec{g}_j \oplus \dots \oplus \vec{g}_{n-h+1}
 \tag{19}$$

Among them, $f(\cdot)$ refers to the convolution operation function, $w \in R^{h \times k}$ refers to the weight parameter of the convolution kernel, g_i refers to the eigenvalue obtained by the convolution operation against the component $X_{i:i+h-1}$, $b \in R$ refers to the bias value, and \oplus refers to the vector splicing action.

Pooling layer: Use the appropriate pooling functions to find out effective feature values in global semantic information, so as to achieve the target task more efficiently. Max function is most commonly used, as follows:

$$\vec{G} = [\max(\vec{g}_1), \max(\vec{g}_2), \dots, \max(\vec{g}_i), \dots, \max(\vec{g}_{n-h+1})]
 \tag{20}$$

Full connect layer: Feature vector is obtained through the above steps, which is set as semantic feature information \vec{F} representing the original text, and then softmax function is called to solve the post-classification result y , as follows:

$$\vec{F} = \sum_{i=1}^{n-h+1} (w \cdot \vec{G}_i + \vec{b}_i)
 \tag{21}$$

$$y = \text{soft max}(V \cdot \vec{F} + b)
 \tag{22}$$

$$\text{soft max}(x) = \frac{e^x}{\sum_{k=1}^K e^x}
 \tag{23}$$

MODEL EVALUATION AND EXPERIMENTAL ANALYSIS

Experimental Data and Evaluation Indicators

The data set adopts social posts provided by the CLPsych2017 Workshop and some posts in the emotion column of Sina Weibo. After selecting the sample data, the data content is preprocessed, such as deleting the information content that cannot be decoded and recognized, data conversion, and evaluation indicators such as Urgent F1, Non-Green F1, All F1, Flagged F1, and others are used to evaluate the model performance.

Among them, for different types of posts, after analysis, it can be known: For green post, it does not show any signs of self-harm tendencies; for amber posts, it contains clues that are likely to show signs of mental health problems; for red post, it indicates that the user is suffering from a more serious problem and needs the help of the forum moderator as quickly as possible; for crisis post, it shows a strong tendency to self-harm. The statistics found that the negative emojis in the crisis and red samples are mostly negative, and the emojis in green sample are mostly positive. Therefore, the emojis that only appear in the crisis or red samples (collectively referred to as the urgent class) are uniformly marked as negative, the emojis that appear only in the green or amber samples (collectively referred to as non-urgent classes) are uniformly marked as positive, and if the emojis appear in both the urgent class and non-urgent class, they are uniformly marked as neutral.

The posts are initialized with the Word2Vec word vector, and the cross-entropy loss function is minimized by the back propagation mode. What's more, a set of optimal model parameters is obtained. To avoid the overfitting phenomenon, L2 regularization is performed for the weight coefficient W (Ahmad et al., 2017; McHaney et al., 2018; Razavi et al., 2021; Wiggins et al., 2021).

Data Preprocessing

The basic operations of data preprocessing are data conversion, emoji labeling, and data filtering. Where data conversion refers to link replacement and case conversion. The former represents the unified replacement of all links in the post with "URL," so as to cut off the connection with the link object. Data filtering represents the removal of non-digital, non-English, and non-punctuation symbols that are difficult to be decoded and recognized in the sample. Emoticon labeling refers to the emoticons in samples being found out and using them to analyze users' emotional states. According to statistics, there are a large number of emoticons reflecting negative expressions in the red and crisis samples, while the green samples are filled with emoticons reflecting positive expressions. Therefore, the emoticons in the urgent samples are uniformly labeled as negative and those in the non-urgent samples as positive. If a certain type of emoticon exists in both samples, they are uniformly labeled as neutral.

TABLE 2 | Model hyper-parameter setting.

Hyper-parameter	Description	Value
D	Word vector dimension	128
N	Learning rate	0.001
H	Convolution kernel sliding window size	23,5,7
Batch size	Batch size	50
P	Dropout rate	0.1
M	Number of convolution kernels	48
\wedge	Regularization coefficient	0.1

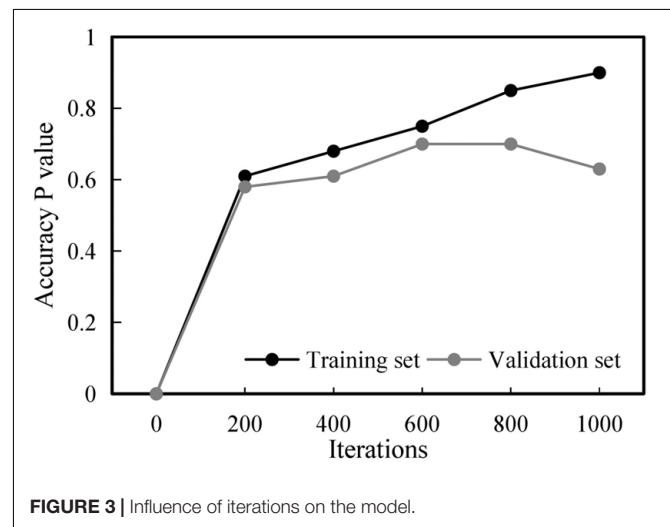


FIGURE 3 | Influence of iterations on the model.

It is important to note that the stop word removal operation is usually not performed during preprocessing. The reasons are as follows: first of all, the distribution of stop words is uneven in each sample, which can reflect the users' usage habits and indirectly help the classification process. Then, the stop words may change the emotional orientation displayed by the text content. For example, the apparent proportion of "not" in the crisis sample is significantly higher than that in the green sample.

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N [y_i \ln \hat{y}_i + (1 + y_i) \ln(1 - \hat{y}_i)] + \frac{1}{2} \lambda \|\omega\|_2^2 \quad (24)$$

Parameter Settings

After obtaining the sample data, the relevant verification experiments are carried out immediately. The hyper-parameter values of the CNN model are shown in Table 2.

Considering that the number of artificially labeled training samples is small, if the number of iterations is too small, the learning effect will be not ideal; if the number of iterations is too large, the model overfitting problem will occur. To verify the effect of the number of iterations on the classification effect, this section divides the original training set into training set (90%) and validation set (10%) for cross-validation. The results are shown in Figure 3.

In the training set, increasing the number of iterations can effectively improve the classification accuracy, up to more than

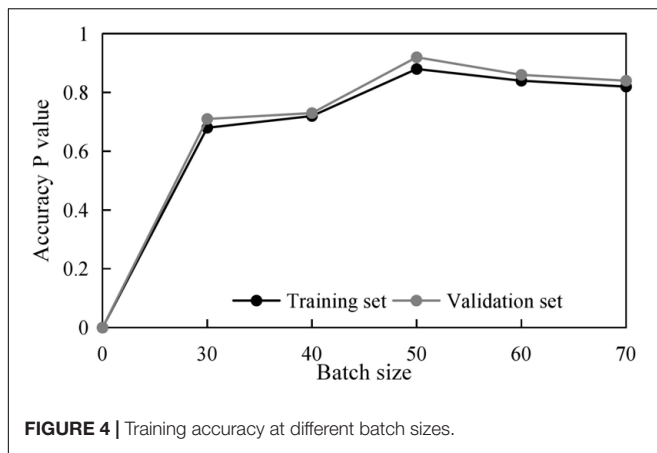


FIGURE 4 | Training accuracy at different batch sizes.

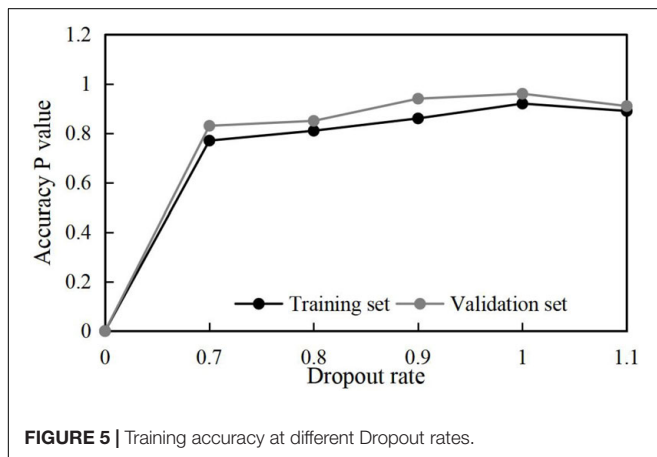


FIGURE 5 | Training accuracy at different Dropout rates.

95%. However, in the validation set, increasing the number of iterations, the classification accuracy first goes through a rapid rise stage and then begins to decline gradually. The accuracy peak of the validation set occurs at the 550th iteration. Therefore, the model at the 550th iteration is adopted as the final test model.

Meanwhile, to obtain some optimal parameters, the influences of different batch sizes and dropout rates on the training results are discussed, and the results shown in Figures 4, 5 are obtained.

Therefore, combining the above optimization training results, some parameters are set as shown in Table 2.

Experimental Results and Analysis

For the test of the application effect of introducing the LIWC dictionary, this section compares the LIWC-CNN model with the traditional CNN model and compares and analyzes the LIWC-CNN model with the model established by Yates et al., so as to test the validity of the LIWC-CNN model.

1. Traditional CNN model. The first is to preprocess the text and establish a Gaussian distribution word vector matrix with a mean $\mu = 0$ and SD $\delta = 1.0$. The second is to perform the operation process of each level. Finally, the classification prediction results are obtained.
2. FastTextf model. The original word vector text matrix is averaged to obtain the hidden layer, and the Softmax layer

TABLE 3 | Comparison of experimental results of various methods.

Methods	Non-green F1	Flagged		Urgent		All	
		F1	Acc	F1	Acc	F1	Acc
FastText	0.21	0.82	0.89	0.00	0.00	0.37	0.66
CNN	0.34	0.79	0.73	0.50	0.36	0.47	0.67
CNN + Word2Vec	0.38	0.83	0.81	0.55	0.69	0.49	0.67
LIWC-CNN	0.40	0.77	0.85	0.59	0.76	0.51	0.69

in the model outputs the final classification prediction results (Shumaly et al., 2021).

3. CNN model based on Word2vec. The trained Word2Vec word vector model is initialized. Then, under the cooperative operation of the convolution layer, pooling layer, fully connected layer, and Softmax layer, the classification prediction results are obtained (Liu et al., 2020).

Using the same source data, the test results of different models are listed in Table 3. By comparison, it is found that the effectiveness of the LIWC-CNN model is the best, and its Non-greenF1 value is equal to 0.39, which is significantly 0.33 higher than that of the traditional CNN model and 0.35 higher than the CNN model based on Word2Vec. When its UrgentF1 value is equal to 0.54, it is significantly 0.46 higher than traditional CNN.

Combined with Table 3 above, the analysis is as follows:

1. The classification effect of FastText model for posts with mental health problems is not ideal. The FastText model has the highest recognition accuracy of 0.89 for flugged samples, which is the highest score among all models, but it has only 0.21 recognition accuracy for non-green samples and is not outstanding on other indicators. In addition, flagged samples exist relative to green samples, but the two kinds of samples are very uneven. The FastText model mostly attributes posts to the relatively rich green type of training samples, which indirectly affects the FastText model for the recognition effect of urgent class samples.
2. The application performance of the traditional CNN model can be significantly improved by using pre-trained Word2Vec in the traditional CNN model. All the evaluated indexes are better than those of the traditional CNN model, confirming that Word2Vec word vector features can describe the semantics of short texts more accurately and profoundly. Then the users' psychological health status is explored. Generally speaking, Word2Vec word vector is obtained by training a language model with a full data set, which integrates richer semantic information and can realize fast convergence.
3. LIWC-CNN can efficiently and accurately identify posts that require urgent intervention. The Non-green F1 value as well as UrgentF1 value reach the highest level. At this time, the model has the best performance, but it is not as good as LIWC-CNN model.

To sum up, after introducing the LIWC dictionary into the CNN model, the new model can efficiently and accurately identify the posts with different mental health characteristics, especially

the recognition effect for the post needing continued intervention is good. It can be seen that the psychological knowledge has guiding significance in the process of deep learning feature extraction, and also shapes the main advantages of the LIWC–CNN model. In essence, the LIWC–CNN model is weighted to represent the word vector in the input layer, and then the convolution kernel is called to extract the deeper features. Under this mechanism, after updating model parameters in each round of back propagation, the weight of the input word vector is adjusted, so as to achieve the maximum application effect of LIWC, which makes the model better study the psychological characteristics reflected by the text.

Overall, the best evaluation result of each model is that the non-greenF1 value is less than 0.4, which is obviously lower than the evaluation model based on the multi-feature fusion method. The main reason is that the number of samples obtained in this paper is insufficient. By using deep learning methods to construct mental health evaluation models, the psychological characteristics contained in the text can be extracted by optimizing the parameters of network structure so as to effectively avoid the tedious process of traditional methods. However, the size of the data set puts forward a higher requirement. If the number of data sets is insufficient, it will affect the evaluation effect. Combined with the reality of the scene, the number of posts in urgent need of intervention is not many, which has become a major problem constraining the promotion and application of the LIWC–CNN model.

CONCLUSION

The idea of modeling in this article is to determine the weight of each word class in the LIWC lexicon in identifying different mental health problem samples based on its distribution differences in different categories of samples, which is used

REFERENCES

- Ahmad, Z., Lutfi, S. L., Kushan, A. L., Khairuddin's, M. H., Zolkeplay's, A. F., Rahmat, M. H., et al. (2017). Construction of the Malay Language Psychometric Properties Using LIWC from Facebook Statuses. *Adv. Sci. Lett.* 23, 7911–7914.
- Caldelli, R., Galteri, L., Amerini, I., and Bimbo, A. D. (2021). Optical Flow based CNN for detection of unlearned deepfake manipulations. *Pattern Recognit. Lett.* 146, 31–37.
- Cui, K., and Zang, B. (2021). Text Classification based on ANN. *Int. Core J. Eng.* 7, 47–50.
- He, M., Wang, X., Zou, C., Dai, B., and Jin, L. (2021). A Commodity Classification Framework Based on Machine Learning for Analysis of Trade Declaration. *Symmetry* 13, 964–964.
- Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2016). “Bag of tricks for efficient text classification,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, (Valencia: Association for Computational Linguistics), 427–431. doi: 10.48550/arXiv.1607.01759
- Kanekar, S. A., Sharma, A., Patkar, G. S., and Tilve, A. K. (2021). Building semantically annotated corpus for text classification of Indian defence news articles. *Int. J. Inf. Technol.* 13, 1539–1544.

to guide the CNN to accurately extract the valuable semantic information in the posts and facilitate a more efficient and accurate assessment of the user's mental health status. The experimental results confirm that the proposed method can effectively identify posts with mental health problems, which is attributed mainly to the fact that the CNN takes into account both the semantic information of the text and the semantic information of the words in the posts when modeling the content of the posts, which also creates the comparative advantage of the model in this manuscript.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

All authors participated in the preparation and presentation of the manuscript and approved the submitted version.

FUNDING

This work was sponsored by Key Projects of Teaching Reform in Departmental Universities (The Construction of College Students' Mental Health Teaching Model from the Perspective of Positive Psychology) (ZD18005), Jiangsu Social Science Foundation (15JYB018) and Jiangsu Universities in Philosophy and Social Science Foundation (2015SJD257), and research on the integrated education mechanism of public security colleges and universities under the cooperation strategy of school bureaus (ZD21005).

- Khaydarova, R. R., Mourontsev, D. I., Lapaev, M. V., and Fishenko, V. D. (2021). ROCK-CNN: distributed Deep Learning Computations in a Resource-Constrained Cluster. *Int. J. Embedd. Real Time Commun. Syst.* 12, 14–31.
- Kumar, P., and Hati, A. S. (2021). Transfer learning-based deep CNN model for multiple faults detection in SCIM. *Neural Comput. Appl.* 33, 1–12.
- Liu, D., Wan, C., Liu, X., Jiang, T., Liao, G., Qiu, X., et al. (2020). Suicidal Ideation Cause Extraction From Social Texts. *IEEE Access* 8, 169333–169351.
- Liu, F., Ma, J., and Wang, Q. (2021). Atom-substituted tensor dictionary learning enhanced convolutional neural network for hyperspectral image classification. *Neurocomputing* 455, 215–228.
- McHaney, R., Tako, A., and Robinson, S. (2018). Using LIWC to choose simulation approaches: a feasibility study. *Decis. Support Syst.* 111, 1–12.
- Moccia, S., Fiorentino, M. C., and Frontoni, E. (2021). Mask-R[Formula: see text]CNN: a distance-field regression version of Mask-RCNN for fetal-head delineation in ultrasound images. *Int. J. Comput. Assist. Radiol. Surg.* 16, 1–8. doi: 10.1007/s11548-021-02430-0
- Nanni, L., Maguolo, G., Brahnas, S., and Paci, M. (2021). An Ensemble of Convolutional Neural Networks for Audio Classification. *Appl. Sci.* 11, 5796–5796.
- Pandi, C., Santosh, D. T., and Vardhan, B. V. (2018). Finding and classifying the deceptive spam reviews using liwc dictionary variables and decision tree classifier. *Int. J. Eng. Technol.* 10, 308–315.

- Razavi, A., Collins, S., Wilson, A., and Okereke, E. (2021). Evaluating implementation of International Health Regulations core capacities: using the Electronic States Parties Self-Assessment Annual Reporting Tool (e-SPAR) to monitor progress with Joint External Evaluation indicators. *Glob. Health* 17, 69–69. doi: 10.1186/s12992-021-00720-5
- Ren, J., Wu, W., Liu, G., and Chen, Z. (2021). Bidirectional Gated Temporal Convolution with Attention for text classification. *Neurocomputing* 455, 265–273.
- Shaikh, S., Daudpota, S. M., Imran, A. S., and Kastrati, Z. (2021). Towards Improved Classification Accuracy on Highly Imbalanced Text Dataset Using Deep Neural Language Models. *Appl. Sci.* 11, 869–869.
- Shi, X., Huang, G., Hao, X., Yang, Y., and Li, Z. (2021). A Synchronous Prediction Model Based on Multi-Channel CNN with Moving Window for Coal and Electricity Consumption in Cement Calcination Process. *Sensors* 21, 4284–4284. doi: 10.3390/s211134284
- Shumaly, S., Yazdinejad, M., and Guo, Y. (2021). Persian sentiment analysis of an online store independent of pre-processing using convolutional neural network with fastText embeddings. *PeerJ Comput. Sci.* 7:e422. doi: 10.7717/peerj-cs.422
- Szajna, A., Kostrzewski, M., Ciebiera, K., Stryjski, R., and Woźniak, W. (2021). Application of the Deep CNN-Based Method in Industrial System for Wire Marking Identification. *Energies* 14:3659.
- Tanana, M. J., Soma, C. S., Kuo, P. B., Bertagnolli, N. M., Dembe, A., Pace, B. T., et al. (2021). How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. *Behav. Res. Methods* 53, 2069–2082. doi: 10.3758/s13428-020-01531-z
- Tay, D. (2021). Modelability across time as a signature of identity construction on YouTube. *J. Pragm.* 182, 1–15.
- Wang, Y., Ko, T. S., and Hsu, C. C. (2021). Interpreting convolutional neural network for real-time volatile organic compounds detection and classification using optical emission spectroscopy of plasma. *Anal. Chim. Acta* 1179:338822. doi: 10.1016/j.aca.2021.338822
- Wiggins, N., Maes, K., Palmisano, G., Avila, L. R., Rodela, K., and Kieffer, E. (2021). A Community Participatory Approach to Identify Common Evaluation Indicators for Community Health Worker Practice. *Prog. Commun. Health Partnersh.* 15, 217–224. doi: 10.1353/cpr.2021.0023
- Yates, A., Cohan, A., and Goharian, N. (1709). Depression and self-harm risk assessment in online forums. *arXiv [Preprint]*. doi: 10.48550/arXiv.1709.01848
- Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., Xu, B., et al. (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. *arXiv [Preprint]*. doi: 10.48550/arXiv.1611.06639
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Li, Li, Fan, Li and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.