WEP Warwick Evans Publishing

Transactions on Social Science, Education and Humanities Research
ISSN: 2960-1770, eISSN: 2960-2262 | Volume 11
CEPC 2024

Prediction of Depression Index Based on LSTM and CNN

Yi Li *, Zida Cai, Jingyi Wang

Minzu University of China, Beijing, China, 100081

* Corresponding Author Email: muclliyyi@163.com

Abstract. In recent years, with the increase of social pressure and the acceleration of life pace, the incidence of depression has shown a rising trend, which makes the prevention, intervention and treatment of depression research is particularly important. At the same time, depression index is an important index used to measure and evaluate the degree of individual depression. In this paper, the MADRS scores corresponding to Montgomery Depression Scale were studied as depression index. In this paper, a depression index prediction model based on long short-term memory neural network (LSTM) and convolutional neural network (CNN) was established by collecting the data of several depressed patients in a hospital. The results show that the model has good predictive ability and stability, which can make a real-time judgment of depression in clinic and improve the efficiency of medical department.

Keywords: LSTM; CNN; Prediction of depression index.

1. Introduction

Depression is a serious mood disorder that can seriously affect a person's ability to lead a normal life. In severe cases, it can even lead to suicidal thoughts [1]. In the face of the high incidence of depression, early identification and intervention become the key. The Depression Index is an indicator used to measure and assess an individual's degree of depression, and was introduced to help medical professionals, researchers, or individuals themselves better understand and assess the severity of depressive symptoms. In addition, Montgomery Scale MADRS can better distinguish patients with depression [2], so this paper takes MADRS score as depression index to study.

propsed sytems before this one

In previous research on depression judgment, machine learning algorithms have been widely used: Sandheep P, Vineeth S, Poulose M et al proposed a deep local global attention convolutional neural network (DLGA-CNN) [3], which can recognize depression, but its efficiency is not high enough. Bashiri F, Mokhtarpour A et al extracted features by analyzing EEG signals and used support vector machine (SVM) to classify depression and normal [4]. Even though Li Y, Fang Y, Ren X et al improved the method of Bashiri F, Mokhtarpour A et al by using fuzzy labels [5], although the efficiency was significantly improved, different depressive states could not be classified. Zhang Y, Fu Z et al proposed for the first time an ERP based Bi-LSTM study on EEG recognition of depression [6]. Compared with other models, the recognition and classification accuracy is high, but it cannot achieve real-time prediction of depression state. Therefore, Hu X, Shu J, Jin Z and others improved it by using skit-Gram model, but they could only detect users' depression tendency [7]. Although Muzammel M, Salam H et al used the CNN model [8], they could only judge whether they were suffering from depression. Therefore, Thekkekara J.P., Yongchareon S, Liesaputra V et al. combined the CNN model with the Bi LSTM model and optimized it by using the attention mechanism [9], making the model superior to the existing deep learning model in depression detection. However, it is impossible to personalize the judgment of depression.

MAIN PROBLEM

Depression is a global mental health problem, and the automatic detection system for depression is of great help to the clinical diagnosis and early intervention of depression [10]. Most traditional diagnostic methods do not consider the basic characteristics of the patient, such as age and gender, and only study the clinical status of the patient is one-sided. Therefore, this paper combines the basic characteristics of the patient with the characteristics of clinical activities, uses deep learning algorithms to conduct research, and proposes a depression index prediction model based on LSTM and CNN models. It is expected that the study in this paper can help to solve the clinical judgment of depressive state.

2. Model building of LSTM and CNN

2.1. The structure of LSTM

Long short-term memory neural network (LSTM) is a special type of recurrent neural network (RNN) that is specifically designed to solve long-term dependence problems that are difficult to solve with standard recurrent neural networks. The following is the structure of a classic long short-term memory neural network as shown in Figure 1:

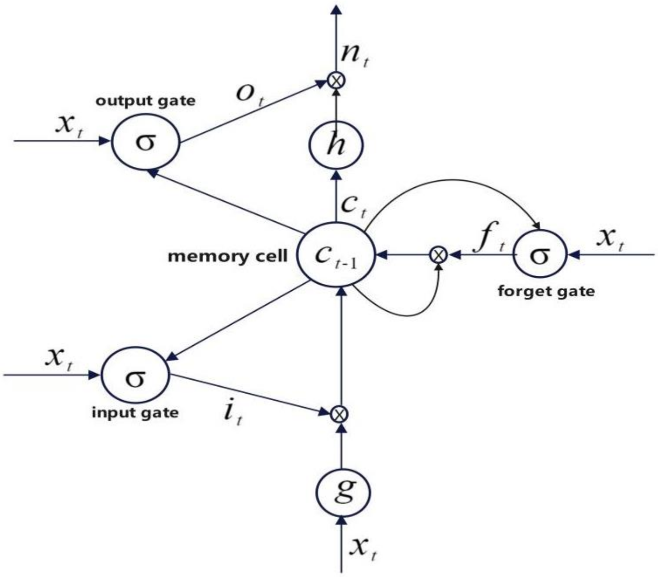

Figure 1. LSTM structure diagram

When training a general neural network model, it is usually used

$$s = f\left(W^T X + b\right) \tag{1}$$

Where W is the weight, X is the input, and b is the constant. The LSTM contains three gates, each of which requires a calculation similar to the one above, in addition to calculating the current state. The following is the detailed calculation formula:

$$i_t = \sigma W_{ix} x_t + W_{ih} h_{t-1} + W_{ic} c_{t-1} + b_i \tag{2}$$

$$D_t = \sigma W_{0x} x_t + W_{oh} h_{t-1} + W_{oc} c_{t-1} + b_o \tag{3}$$

$$f_t = \sigma W_{fx} x_t + W_{fh} h_{t-1} + W_{fc} c_{t-1} + b_f \tag{4}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tanh W_{cx} x_t + W_{ch} h_{t-1} + b_c \tag{5}$$

$$n_t = o_t \times \tanh c_t \tag{6}$$

The input expression of the model is $x = \left\{x_0, x_1, x_2, \cdots, x_t\right\}$, which x_t corresponds to the input parameters of the long short-term memory neural network at time t; which n_t Corresponding to the output parameters of the short-term and long-term memory neural network at time t; i_t. f_t, o_t, and c_t correspond to the input gate, forget gate, output gate, and memory cell at time t, respectively; W. B refers to the connection weights and bias values of neurons between the input layer and memory cells, as well as between memory cells and output layers.

$$\sigma(x)=\frac{1}{1+e^{-x}} \tag{7}$$

2.2. The structure of CNN

Convolutional neural networks (CNNS) are deep learning algorithms specifically designed to process data structures with grid-like structures, such as images or video frames. The design of such networks is inspired by biological processes, particularly the way the visual cortex is organized, where certain cells are highly sensitive to specific areas of input to the visual scene. The structures of convolutional neural networks are varied. Figure.2 shows a classical convolutional neural network, which is composed of convolutional layer, pooling layer and fully connected layer.

In this paper, based on LSTM, a feature extraction model is built according to the time series of patients' activity values. It consists of 2 layers of LSTM layer, attention layer, flat layer and 1 layer of fully connected layer. The specific structure is shown in Figure 2:

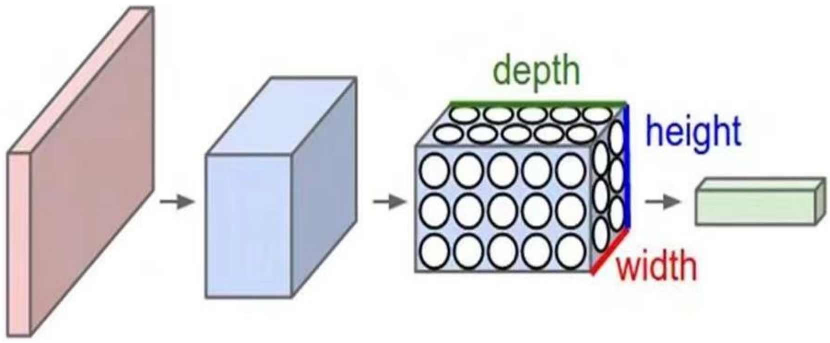

Figure 2. Classic CNN structure

2.3. The construction of LSTM

At the beginning of this paper, a feature extraction model based on LSTM was constructed for the time series of patients' activity values. The training set data is first selected, while each individual time series data is standardized to ensure that the data is on the same scale, helping the model converge faster and better. The time series is then segmented into fixed length sequences to fit the input requirements of the model. The LSTM network is then chosen because LSTM is particularly suited for working with time series data and is able to capture long-term dependencies in the data. In terms of LSTM layers, the number of cells ranging from 50 to 200 was tested, and different LSTM layer depths of 1 to 4 layers were tried. After many experiments and model performance comparison, it is finally determined that the configuration of 100 cells per layer and two-layer LSTM can capture the key information in the time series and avoid the overfitting problem caused by too many parameters. The attention mechanism is also added later to emphasize the key moments in the sequence and improve the sensitivity of the model to important events in the time data. The final model consists of 2 LSTM layers, attention layer, flat layer and 1 fully connected layer. The architecture of the model is shown in Figure 3:

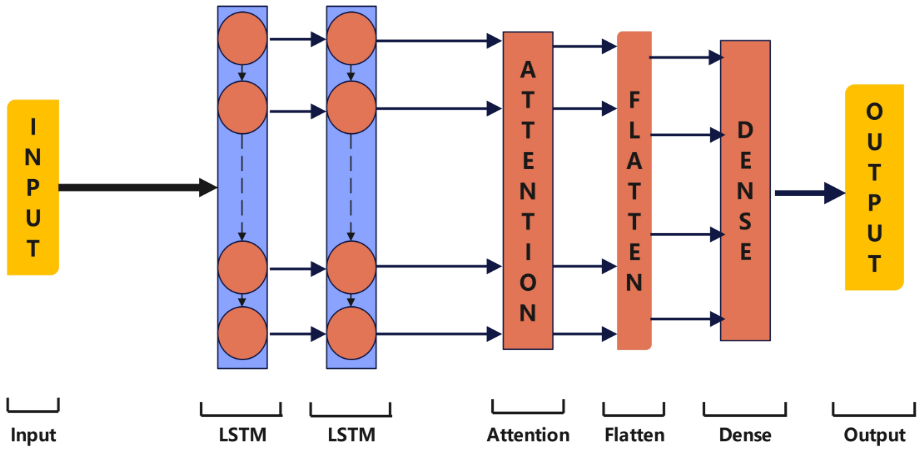

Figure 3. LSTM structure of this paper

2.4. The construction of CNN

In the process of building the CNN model, the initial model output showed some instability, which was reflected in the prediction results of all NaN. This may be due to problems with the ReLU activation function causing the gradient to disappear or explode in some cases, so I changed the activation function to ELU (exponential linear unit). Subsequently, the depth and width of the convolution layer were also adjusted. After a series of experiments, it was found that 64 initial convolutional filters, subsequently increased to 128, achieved an optimal balance on this dataset. Each layer of the network is followed by a maximum pooling layer to reduce dimensions and prevent overfitting. The flat layer is followed by two dense layers, first a dense layer of 100 neurons and then 50 neurons, both using the ELU activation function. The final output layer contains only one neuron, and since the goal of this article is regression analysis, linear activation functions are used here. The convolutional time network structure adopted by the training module in this paper is shown in Figure 4, which is composed of 2 convolutional layers, 2 pooling layers, 1 flattening layer and 2 fully connected layers.

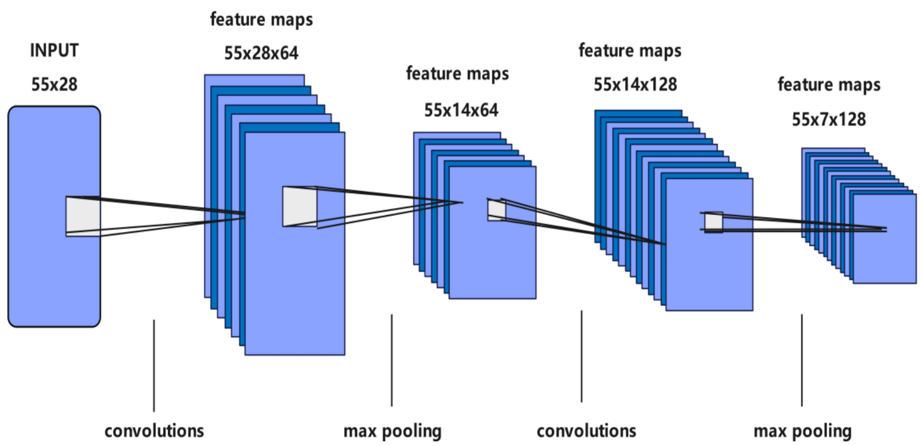

Figure 4. CNN structure of this paper

3. Results

3.1. LSTM output characteristic result analysis

When training, the Adam optimizer is used, the learning rate is set to 0.0005, the model is trained over 500 cycles and does not require target variables because the purpose of LSTM is feature extraction, not prediction. After the training, the model can extract a feature value from the time series data of each training set and test set individual, and integrate these feature values for the subsequent construction of CNN models. After the training, the model can extract an eigenvalue from the time series data of each patient and control group, output and organize the eigenvalues of all individuals, and the specific eigenvalues are shown in Table 1 and Table 2:

Table 1. Feature table of training set extraction

Number	Features
train_1	0.000135651
train_2	-0.00013632
train_3	0.000273293
train_4	5.07589E-05
train_5	-2.45273E-05
train_6	-8.96789E-05
train_7	0.000126158
train_8	1.21263E-05
train_9	0.000266552
train_10	-1.79233E-06
train_11	-3.81079E-05
train_12	-0.000336943
train_13	-0.000251947
train_14	-0.000164326
train_15	-0.000291284
train_16	6.27404E-05
train_17	4.83925E-05
train_18	3.13758E-05
train_19	0.000115936
train_20	0.000151799
train_21	0.000161206
train_22	1.24127E-05
train_23	0.000135489
train_24	1.43386E-05
train_25	-0.000126315
train_26	-0.000187648
train_27	-0.000475511
train_28	-0.000240396
train_29	4.95277E-06
train_30	-0.000977091
train_31	1.20946E-05
train_32	0.000185033

By constructing the LSTM time series model and using the attention mechanism to extract the characteristic values of the activity values of each patient over time, that is, extracting a separate characteristic value for each patient and control group member. It can be seen from Table 1 and Table 2 above that the range of eigenvalues in both the test set and the training set ranges from negative values to positive values. For example, train_2, train_5, train_6, train_10, train_11, train_12, train_13, train_14, train_15, train_25, train_26, train_27, train_28, train_30 in the training set are negative values, while train_1, train_3, train_4, train_7, and so on are positive values. In the test set, test_2, test_8, test_12, test_14 and so on are negative, and test_1, test_3, test_4, test_5, test_6, test_7, test_9, test_10, test_11, test_13 and so on are positive, which is enough to show that different patients in the same time the activity characteristics are different. It also showed differences in the activity data of patients in different test sets and training sets. As for the different sizes of the different characteristics, whether this has an effect on the change in the depression index will need to be known later. These eigenvalues represent the overall performance of the activity pattern or activity intensity in the time series, which reflects that our LSTM model has a very good effect on the feature extraction of time series.

Table 2. Feature table of testing set extraction

Number	Features
test_1	5.2418E-05
test_2	-5.22048E-05
test_3	4.89019E-05
test_4	0.000208022
test_5	0.000177852
test_6	9.21553E-05
test_7	0.000366256
test_8	-0.000191095
test_9	0.000121825
test_10	0.000274958
test_11	2.67166E-05
test_12	-0.000236095
test_13	0.000584203
test_14	-0.000210962
test_15	0.000361773
test_16	-5.46067E-05
test_17	-0.000281309
test_18	-0.000109449
test_19	-0.000134918
test_20	2.47231E-05
test_21	0.000167633
test_22	0.000621992
test_23	-0.000301256

3.2. Analysis of the results of the CNN predictive Depression Index

For model evaluation, this paper draws the CNN model training and test loss degree graph, as shown in Figure 5: training loss is represented by blue line, verification loss is represented by orange line.

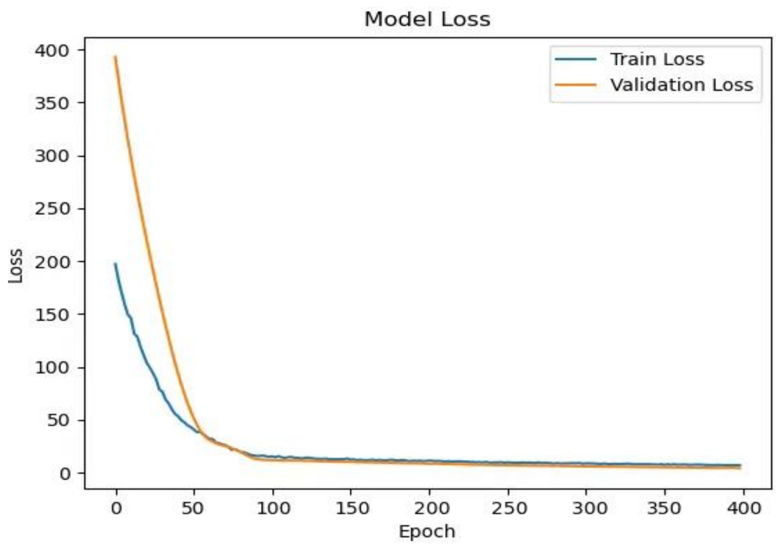

Figure 5. CNN model training and validation loss degree graph

As shown in Figure 5:

(1) Training loss: The loss starts at a high value and decreases rapidly over the first few epochs, after which it gradually flattens out and tends to a lower level. This shows that the model is learning the characteristics of the training data and improving its prediction ability.

(2) Validation loss: Its downward trend is similar to that of training loss, indicating that the model's performance on unseen data is also improving simultaneously. The initial decline was very rapid, and then it also leveled off.

(3) The loss tends to be stable: After the initial rapid decline, the two lines become very smooth and close to each other in about 100 cycles, which indicates that the model has reached a relatively stable loss value on the training set test set.

(4) No signs of overfitting: During the entire training process, training losses and validation losses were very close, and there was no continuous decline in training losses while validation losses increased.

(5) Convergence: The convergence of the two curves indicates that the model may have found a better solution for a given architecture and data set, and the loss value does not change much in the later stages of training.

Overall, the figure shows that model losses decline and stabilize with training, and show good consistency across the training and test sets.

Then the test set data is input into the model to evaluate the model's prediction effect. The specific prediction results are shown in Table 3:

Table 3. The test set predicted the MADRS score versus the actual MADRS score

number	Predicted_madrs2	Actual_madrs2
test_1	18.36671829	19
test_2	11.72689056	11
test_3	23.75287437	25
test_4	17.59941483	16
test_5	25.63259315	26
test_6	16.44964409	15
test_7	24.25936699	25
test_8	17.04142952	16
test_9	24.88441849	26
test_10	22.19405746	21
test_11	24.37545204	24
test_12	19.39029694	21
test_13	13.60834694	13
test_14	22.45866776	19
test_15	16.74028873	18
test_16	16.92998314	17
test_17	14.97418213	15
test_18	14.21745586	15
test_19	21.83620262	21
test_20	23.71932983	25
test_21	21.88807297	21
test_22	27.10066795	28
test_23	23.19208717	23

Table 3 shows the predictive power of the model for the severity of depression (MADRS score), with many of the predictions being quite close to the actual values, such as test_3 (prediction 23.75, actual 25), test_9 (prediction 24.88, actual 26), and test_23 (prediction 23.19, actual 23). This shows that the model can capture and reflect the depressed emotions of patients relatively accurately. In some cases, the predictions were very close or matched the actual results, such as test_13 (prediction 13.61, actual 13) and test_19 (prediction 21.84, actual 21), which further validated the validity of the model on a particular instance.

According to the results in Table 3, the discussion continues from the following aspects:

(1) Error analysis: Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are a few commonly used measures of the predictive accuracy of regression models. Among them:

1) The mean square error (MSE) is calculated as:

$$MSE = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{8}$$

MSE measures the average of the squares of the differences between the predicted and true values, which gives a higher weight to larger errors.

2) The root mean square error (RMSE) is calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2} \tag{9}$$

RMSE measures the average deviation between the predicted results and the actual values. The lower its value, the better the performance of the prediction model. It is expressed in the same magnitude as the original data and therefore visualizes the magnitude of the error.

3) The mean absolute error (MAE) is calculated as:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| \tag{10}$$

Where n is the number of observations, y_i is the true value of the ith observation, and $\hat{y}_i$ is the predicted value of the ith observation. MAE is the average of the absolute value of the difference between all individual observations and the predicted value. Unlike MSE, MAE weights all individual differences equally on the mean, so it highlights outliers better.

Based on the results of Table 3 the MSE, PMSE and MAE are calculated as shown in Table 4:

Table 4. Error analysis table

MSE	RMSE	MAE
1.427211	1.194659	0.974004

The value of RMSE is about 1.19, indicating that the average deviation of the predicted depression index from the actual depression index is only 1.19, which has basically no effect on the clinical judgment of depression status and reflects the good predictive ability of the model.

(2) Consistency: the model's prediction results for individuals in different test sets are somewhat consistent, with no extreme misclassifications, indicating that the model is stable overall.

(3) Model bias: It is also important to observe whether the model systematically overestimates or underestimates MADRS scores. From the data provided, there is no significant model bias, with most predictions fluctuating closely around actual values.

According to these prediction results, the CNN model shows good predictive ability and stability, which is suitable for the evaluation of depression severity, that is, real-time prediction of depression index. Although there were some prediction errors, most predictions were very close to the actual results, indicating that the model captured the patient's state relatively accurately.

4. Conclusion

Depression has become a major problem in the world's public health field. In this paper, the LSTM model was first used to extract features from the time series of patients, and it was found that the LSTM model was particularly suitable for processing time series data, and the time features of the extracted patients could fully reflect the differences between different individuals, reflecting that the constructed LSTM model had a good effect on the feature extraction of time series. Secondly, the CNN model is used to predict the depression index, and the model is evaluated by the test set data. It is found that the CNN model shows good prediction ability and is very stable. The above results show that this model can be used to make a real-time judgment of depression in clinic and improve the efficiency of medical departments.

References

[1] Wang Y, Li Z. Depression Detection with Dynamic and Static Visual Features [J]. Journal of circuits, systems and computers, 2023 (18): 32.

[2] Knapskog A B, Barca M L, Engedal K. A comparison of the validity of the Cornell Scale and the MADRS in detecting depression among memory clinic patients [J]. Dement Geriatr Cogn Disord, 2011, 32 (4): 287-294.

[3] Sandheep P, Vineeth S, Poulose M, et al. Performance analysis of deep learning CNN in classification of depression EEG signals [C] // TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON). IEEE, 2019.

[4] Bashiri F, Mokhtarpour A. Depression classification and recognition by graph-based features of EEG signals [J]. International journal of medical engineering and informatics, 2022 (3): 14.

[5] Li Y, Fang Y, Ren X, et al. EEG-based depression recognition using feature selection method with fuzzy label [J]. Journal of King Saud University - Computer and Information Sciences, 2024, 36 (3).

[6] Zhang Y, Fu Z. The study of EEG Recognition of Depression on Bi-LSTM based on ERP P300 [J]. E3S Web of Conferences, 2020, 185 (18): 02007.

[7] Hu X, Shu J, Jin Z. Depression tendency detection model for Weibo users based on Bi-LSTM [J]. IEEE, 2021.

[8] Muzammel M, Salam H, Hoffmann Y, et al. AudVowelConsNet: A Phoneme-Level Based Deep CNN Architecture for Clinical Depression Diagnosis [J]. 2020.

[9] Thekkekara J P, Yongchareon S, Liesaputra V. An attention-based CNN-BiLSTM model for depression detection on social media text [J]. Expert Systems with Applications, 2024, 249.

[10] Lin L, Chen X, Shen Y, et al. Towards Automatic Depression Detection: A BiLSTM/1D CNN-Based Model [J]. Applied Sciences, 2020, 10 (23): 8701.