3DMKDR: 3D Multiscale Kernels CNN Model for Depression Recognition Based on EEG

Yun Su[✉], Zhixuan Zhang, Qi Cai, Bingtao Zhang, Xiaohong Li

Abstract: Depression has become a major health threat around the world, especially for older people, so the effective detection method for depression is a great public health challenge. Electroencephalogram (EEG) can be used as a biomarker to effectively explore depression recognition. Motivated by the studies that multiple smaller scale kernels could increase nonlinear expression compared to a larger kernel, this article proposes a model named the three-dimensional multiscale kernels convolutional neural network model for the depression disorder recognition (3DMKDR), which is a three-dimensional convolutional neural network model with multiscale convolutional kernels for depression recognition based on EEG signals. A three-dimensional structure of the EEG is built by extending one-dimensional feature sequences into a two-dimensional electrode matrix to excavate the related spatiotemporal information among electrodes and the collected electrode matrix. By the major depressive disorder (MDD) and the multi-modal open dataset for mental-disorder analysis (MODMA) datasets, the experiment shows that the accuracies of depression recognition are up to 99.86% and 98.01% in the subject-dependent experiment, and 95.80% and 82.27% in the subject-independent experiment, which are higher than alternative competitive methods. The experimental results demonstrate that the proposed 3DMKDR is potentially useful for depression recognition in older persons in the future.

Keywords: major depression disorder (MDD); electroencephalogram (EEG); three-dimensional convolutional neural network (3D-CNN); spatiotemporal features

1 Introduction

Depression is a common mental disease characterized by persistent low mood and cognitive impairment, which severely affects people's quality of life [1].

The world's population is aging at a fast pace [2]. Episodes of depression disorders might be expected to be more prevalent in older age due to the increased risk of adverse life events. Compared with younger adults, older people more often have substantial depressive symptomatology without meeting the diagnostic criteria for a depressive disorder. This condition affects nearly 1 in 10 older adults [3].

Therefore, we should pay more attention to the diagnosis of elderly people with depressive symptomatology. With the high incidence and low recognition rate of depression, exploring simple, objective, and accurate evaluation methods for depression detection for older people is a major public health challenge [4]. Electroencephalogram (EEG) has the ability to objectively reflect the user's inner true emotional experience [5]. Therefore, it is urgent and necessary to explore an effective EEG-based method

Yun Su, Zhixuan Zhang, Qi Cai and Xiaohong Li are with School of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China.

Bingtao Zhang is with School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China.

[✉] Corresponding author. Email: suyun@nwnu.edu.cn

for depression recognition.

Currently, deep learning (DL) methods have been used in depression and emotion detection based on EEG and can achieve superior performance compared to traditional machine learning methods [6]. For example, Sandheep et al.[7] proposed that an extensive data learning approach method and the convolutional neural network (CNN) model can distinguish depression patients better in resting state EEG data. Shao et al. [8] proposed a CNN model for classifying the EEG data of depressed and normal subjects. Deng et al. [9] proposed a model composed of five parallel convolutional filters that learned the space-frequency features of EEG, and its accuracy was up to 94.37% in the MODMA dataset [10]. The recognition of mental states requires consideration of not only the temporal dependence between data points but also the spatial relevance between different electrodes of EEG, which the three-dimensional convolutional neural network (3D-CNN) has the ability to automatically extract spatiotemporal features [11].

The existing recognition models have achieved high accuracy, while some researchers believe that multiple smaller-scale kernels have the ability to increase nonlinear expression more than a larger kernel [12]. Defining multiscale convolutional kernels based on temporal and spatial features is still an interesting topic in depression recognition research by the 3D-CNN model. Therefore, we proposed the 3D multiscale kernels convolutional neural network model for the depression disorder recognition (3DMKDR) in this paper. Specifically, we designed the spatial position of EEG electrodes according to the 10-20 electrode system [13], which retains the spatial information of all EEG electrodes. Different kinds of smaller-scale convolutional kernels and double linear convolutional structures are joined to our model to obtain efficient recognition ability. In summary, the main contributions of this paper are as follows.

1) The raw EEG signals are transformed into a three-dimensional (3D) dataset according to the topology of EEG electrodes. This method can completely capture the prior knowledge among electrodes, where the spatial and temporal information of electrodes are mined.

2) The 3D-CNN model with multiscale convolutional kernels is constructed to achieve more efficient depression recognition performance. Multiscale convolutional kernels have powerful potential for performing recognition based on EEG.

3) This study conducts extensive experiments to verify the effective recognition ability of 3DMKDR based on EEG from the MDD [14] and MODMA depression datasets. The results are higher than those of alternative competitive methods.

The paper is structured as follows. In Sections 2 and 3, the main methodology adopted in the work is discussed and followed by the experiment. The discussion with results derived from the experiment is presented in Section 4. Section 5 provides the main considerations and some future work.

2 Methods

In this study, we presented a relocated electrode method and 3DMKDR model. The procedure of the method is shown in Fig. 1. First, the original EEG signals from the EEG datasets are preprocessed and converted from the 2D to the 3D format by the located electrodes. Then, the 3DMKDR model is constructed and evaluated based on the recognition of major depressive disorders (MDDs) relative to healthy controls (HCs).

2.1 Preprocessing of EEG

It is necessary to preprocess the raw EEG signals to improve the recognition accuracy [15]. Some researchers relocated the EEG electrodes to the 2D topology based on the topological relation according to the 10-20 electrode system to preserve the spatial information of EEG electrodes, which can be conducive to improving the

recognition ability of the model [16].

To improve the accuracy of depression detection, the topology of electrodes is further improved based on our previous work [17]. Located
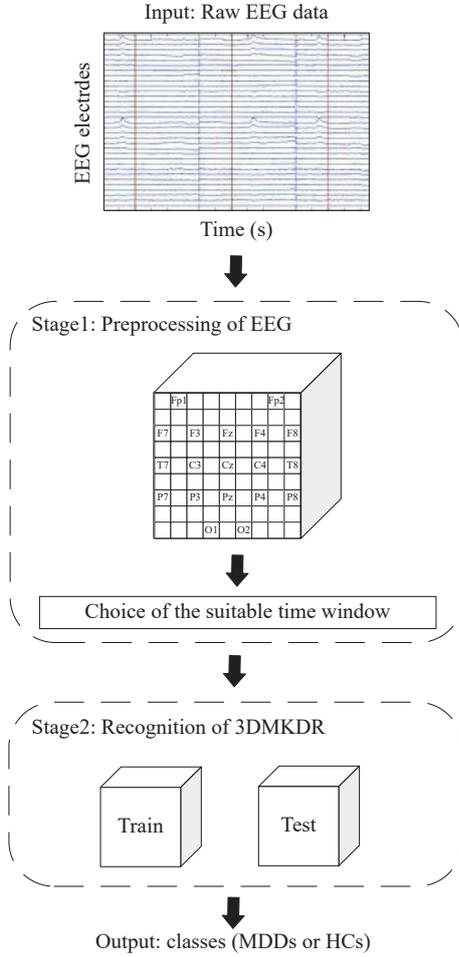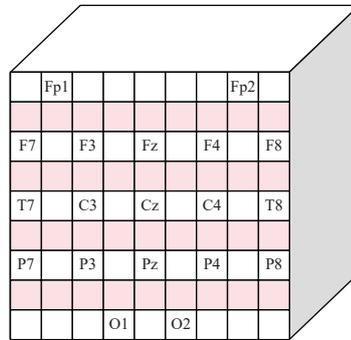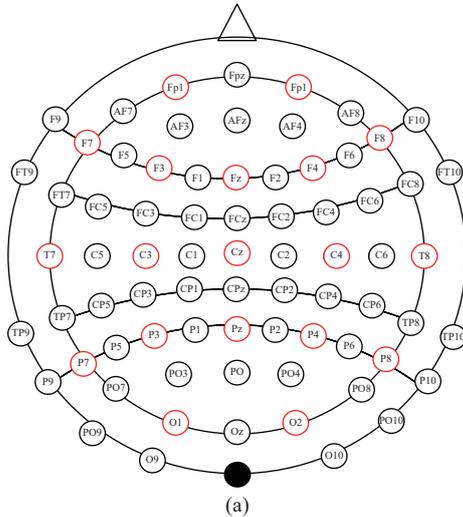
Input: Raw EEG data

Fig. 1　The procedure of the method

in the 10-20 electrode system shown in Fig. 2(a), a blank row is added between the two rows of electrodes during design processing based on a new dataset to ensure the topology sparseness among electrodes. This method can maintain reasonable spacing among all electrodes and ensure the exact electrode position. According to the farthest distance between the two electrodes, the size of the two-dimensional matrix is set to 9×9.

Then, the selected EEG signals are mapped into a 9×9 matrix, as shown in Fig. 2(b). Each one-dimensional EEG signal data vector is inserted into a two-dimensional topological matrix and the time sampling data vector of EEG acts as the third dimension of the three-dimensional matrix. The blank position is represented as a topological position of the unselected physiological signals, which is set to zero in the 9×9 matrix, and the matrix is normalized.

2.2 Recognition of 3DMKDR

Based on the generated 3D EEG dataset, the 3D-CNN model with multiscale convolutional kernels for depression recognition based on EEG signals (3DMKDR) is constructed to extract multiple feature matrixes that fuse temporal and spatial features.

In our model, $\boldsymbol{k}_{i_k \times j_k \times l_k}^{mn}$ is assumed to be the n-th ($n = 1, 2, ..., N$) convolutional kernel between the m-th convolutional layer and the

F: the frontal lobe　　　　C: the central region
T: the temporal lobe　　　Odd numbers: the left hemisphere
O: the occipital lobe　　　Even numbers: the right hemisphere
P: the parietal lobe

(a)　　　　　　　　　　　　　　(b)

Fig. 2　Brain electrode distribution diagram: (a) brain electrode distribution diagram [13]; (b) corresponding matrix

$(m+1)$-th convolutional layer. $\boldsymbol{b}^{mn}_{i_b \times j_b \times l_b}$ is the excursion matrix, and $\boldsymbol{C}^{mn}_{i_n \times j_n \times l_n}$ is a feature matrix in the m-th convolutional layer. $\boldsymbol{C}^{(m+1)n}_{i'_n \times j'_n \times l'_n}$ is the feature matrix in the $(m+1)$-th convolutional layer that is calculated based on $\boldsymbol{C}^{mn}_{i_n \times j_n \times l_n}$, which can be described as

$$\boldsymbol{C}^{(m+1)n}_{i'_n \times j'_n \times l'_n} = \\ \mathrm{ReLU}\left(\sum_{n=1}^{N} \boldsymbol{C}^{mn}_{i_n \times j_n \times l_n} * \boldsymbol{k}^{mn}_{i_k \times j_k \times l_k} + \boldsymbol{b}^{mn}_{i_b \times j_b \times l_b} \right) \quad (1)$$

where $*$ represents the convolutional processing. i, j and k vary in size in each 3D matrix and mean three distinct dimensions in the same matrix. $\boldsymbol{C}^{mn}_{i_n \times j_n \times l_n}$, $\boldsymbol{C}^{(m+1)n}_{i'_n \times j'_n \times l'_n}$, $\boldsymbol{k}^{mn}_{i_k \times j_k \times l_k}$ and $\boldsymbol{b}^{mn}_{i_b \times j_b \times l_b}$ are different 3D matrixes, and the lengths in three dimensions are represented by $i_n \times j_n \times l_n$, $i'_n \times j'_n \times l'_n$, $i_k \times j_k \times l_k$ and $i_b \times j_b \times l_b$, respectively. The nonlinear activation function $\mathrm{ReLU}(\cdot)$ (rectified linear unit) is described as

$$\mathrm{ReLU}(x) = \max(0, x) \quad (2)$$

3 Experiments

3.1 Datasets

To evaluate the effectiveness of our method, we conducted experiments on the MDD and MODMA datasets to analyze depressive states in different subjects.

The MDD dataset records resting state and task state EEG data from 34 major depression patients and 30 healthy control people. The task state data are used in our experiment. The experiment paradigm of the task state was that the subjects were asked to watch three kinds of patterns on the computer screen and press the space bar when the 5 cm circle in patterns appeared. The task state data were collected for 10 min, and the lengths of different trials were different in all subjects.

The MODMA dataset provided by Lanzhou University is the second open dataset to test our method in this paper. The EEG data in the MODMA consist of task state and resting state

data from 24 major depression patients and 29 healthy control people. The task state data are also chosen in our experiment. In the task state, all subjects were asked to identify the spatial position of the "dot" in one kind of picture and press the "1" or "4" button as soon as possible, when four graphs were randomly played on a computer screen. The whole task state experiment was completed in approximately 25 min, and the lengths of different trials were different in all subjects.

3.2 Preprocessing

In the experiments, the EEG data of the task state in the MDD and MODMA datasets are chosen, which includes more data of subjects and is helpful to obtain better recognition ability.

3.2.1 MDD Dataset

The task state data of the MDD dataset with 22 EEG electrodes from 64 subjects were recorded. Because 3 electrodes of 22 are ear-related electrodes and the data of three subjects (the 20th and the 21st healthy subjects, the 9th major depression subject) are invalid, this research selects 19 electrodes shown in the red circle of Fig. 2 (b) and 2 types of labels, i.e., major depressive disorders and healthy controls, from 61 subjects in the MDD dataset.

The processing step is as follows. First, artifacts are removed from the data by a bandpass frequency filter of 0.1–50.0 Hz and downsampled from 256 Hz to 128 Hz. The data lengths of trials in all subjects are different, so the dataset $(61 \times 19 \times 77\,440)$ is obtained by comparing and choosing the minimum data length $(77\,440)$ of all trials, where 61 is the number of subjects, 19 is the number of EEG electrodes and 77 440 is the number of sampled data points. Second, the dataset is changed to a new dataset with a size of $1\,209 \times 19 \times 128$, where 1 209 is the number of fragments by time window, 19 is the number of EEG electrodes, and 128 is the number of sampled data points. Finally, 73 749 3D matrixes $(9 \times 9 \times 128)$ are obtained based on the 10–20 elec-

trode system, where 73 749 is the product of 61 × 1 209 (subject number× fragment number), 9 × 9 is the devised 2D matrix, and 128 is the sampled data point.

3.2.2 MODMA Dataset

The task state EEG data in the MODMA record with 128 electrodes from 53 subjects. To ensure channel sparseness and alleviate the computation needed, 19 EEG electrodes that are the same as the MDD dataset and 2 labels (MDDs and HCs) are selected from this dataset for the analysis of the depressive disorder state.

The processing step is as follows. First, artifacts are removed from the data by a bandpass frequency filter of 0.1–50.0 Hz and downsampled from 250 Hz to 128 Hz. The dataset (53×19×99 836) is obtained by comparing and choosing the minimum data length (99 836) of each trial, where 53 is the number of subjects and 19 is the number of EEG electrodes. Second, 82 521 3D matrixes (9×9×128) are obtained for MODMA by the

same steps as the MDD dataset, where 82 521 is the product of 53×1 557 (subject number× fragment number), 1 557 is the number of fragments by the processing of time windows, 9×9 is the 2D matrix, and 128 is the sampled data points.

3.2.3 Time Window

To further improve the recognition accuracy of the model, a suitable time window is chosen by comparing two kinds of time windows. In the literature [18], people confirmed that the average classification accuracy of a 1 s time window based on EEG was superior to other types. However, researchers believe that the 2 s time window is the most common application [19].

To address this problem, we compare EEG classification results at two kinds of time window lengths in Fig. 3 (a) and Fig. 3 (b). It is obvious that the classification accuracy of the 1 s time window is higher than that of the 2 s time window with the same number of iterations. Compared with the 2 s time window, the data
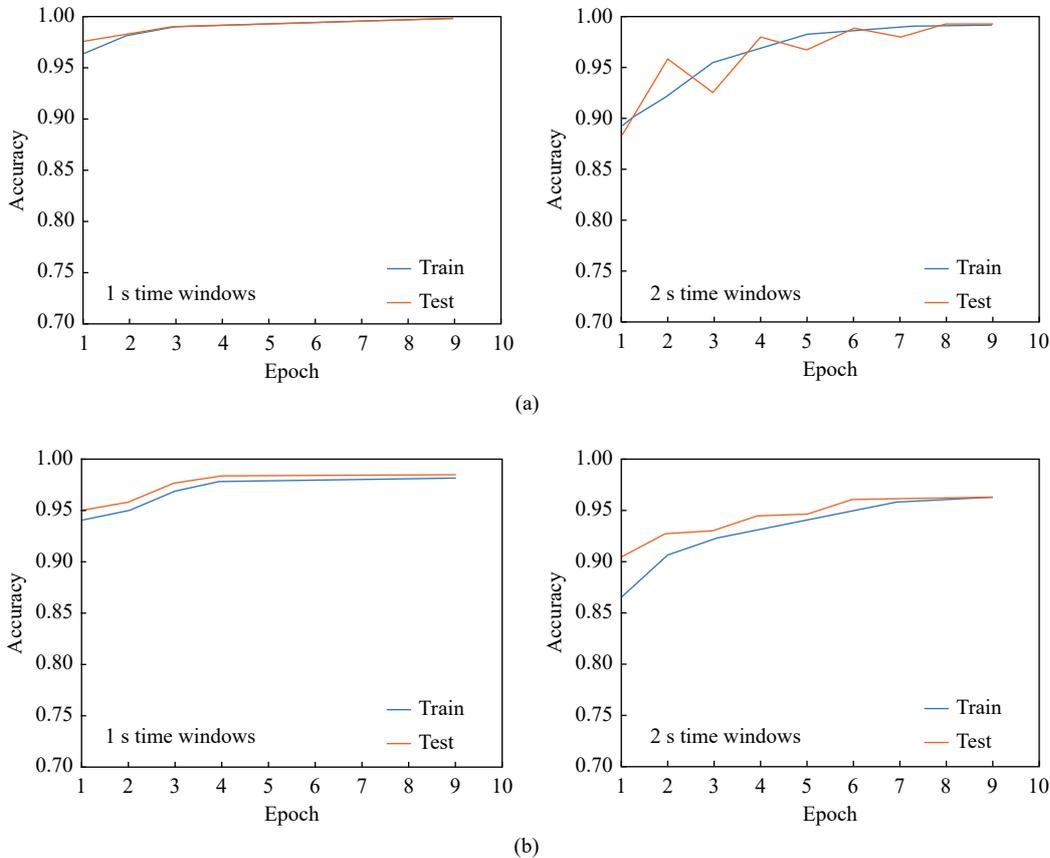

(a)

(b)

Fig. 3   Accuracy comparison of different time windows: (a) in the MDD dataset; (b) in the MODMA dataset

fragments by the 1 s time window can make the model obtain more detailed emotional characteristics and gain a suitable amount of data for the recognition of the model. Therefore, we select the 1 s time window length as suitable for the model, and a 0.5 s overlapping window is set to ensure high-quality time continuity of data and achieve high-efficiency recognition accuracy.

3.3   3DMKDR Model

The 3DMKDR model is constructed to cover the related information among the brain electrodes and helps to learn the advanced features of EEG signals from 3D data. Multiple convolutional kernels can increase the nonlinear expression and make the judgment function more efficient [20].

Moreover, the multiscale kernel can improve the robustness of the network to extract features, which advances the generalization ability and classification accuracy [21].

The detailed architecture of the 3DMKDR model is shown in Fig. 4 and Tab. 1. This model consists of three layers, in which layer 1 is implemented in parallel with layer 2. Layer 1 includes convolutional layer 1-1 and convolutional layer 1-2. The kernel sizes of the convolution layers in layer 1 and layer 3 are both $3\times3\times4$, where spatiotemporal features are extracted by the local spatial topology of 3×3 and temporal fragments of 4 sampled points. The kernel size of layer 2 is $3\times3\times5$.
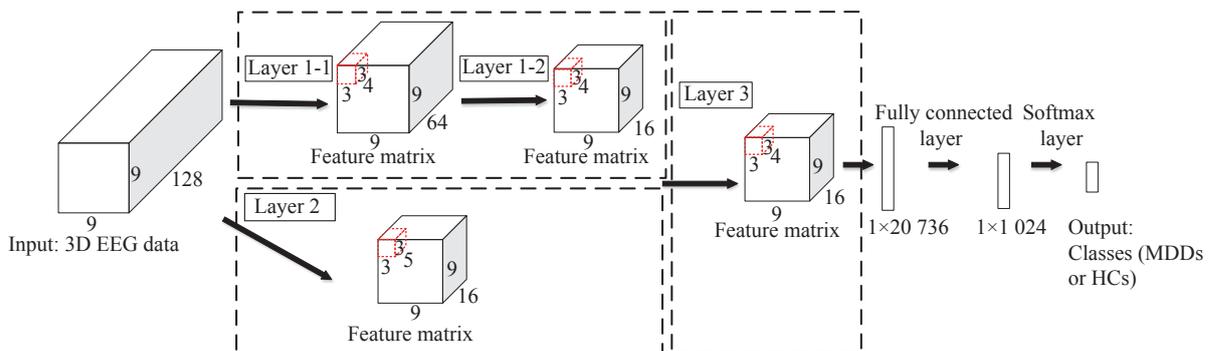

Fig. 4   The overall architecture of the model

Tab. 1   Detailed architectures of the model

No.	Layer type	Units	Kernel	Stride	Input
	Convolution	32	$3\times3\times4$	1,1,2	$9\times9\times128$
1	Convolution	32	$3\times3\times4$	1,1,2	$9\times9\times64$
	Pooling	32	$1\times1\times2$	1,1,2	$9\times9\times32$
2	Convolution	32	$3\times3\times5$	1,1,2	$9\times9\times128$
	Pooling	32	$1\times1\times2$	1,1,4	$9\times9\times32$
3	Convolution	64	$3\times3\times4$	1,1,2	$9\times9\times16$
	Pooling	64	$1\times1\times2$	1,1,2	$9\times9\times8$
Fully connected		\	\	\	20 736
Softmax		\	\	\	1 024

Layer 1 consists of layer 1-1 and layer 1-2, which can perform the coarse-grained calculation of spatiotemporal information based on the small convolutional kernel. Layer 2 accomplishes the same operation of coarse-grained calculation based on another type of small convolutional kernel. After the procession of layer 1 and layer 2,

layer 3 implements fine-grained screening of spatiotemporal information.

The activation function $\mathrm{ReLU}(\cdot)$ is used after each convolutional layer to perform nonlinear feature transformation and to prevent the loss of information from the input data. A 3D maximum pooling layer is set behind each convolutional layer and used to extract the features more efficiently here, which can reduce the quantity of data on the temporal dimension. The last pooling layer is followed by the fully connected layer and the Softmax layer is deployed as the output.

4   Results and Discussion

This model has been implemented in the PyTorch framework and deployed on the GeForce RTX 3060. The learning rate was set to

0.001 with the Adam Adadelta Optimizer, and the probability of the dropout operation was set to 0.6. Then, 10-fold cross-validation [22] and leave-one-subject-out cross-validation (LOSO) methods [23] were used to evaluate the performance of the 3DMKDR model.

To estimate the accuracy of depression recognition for the 3DMKDR model, the model was analyzed by the results of two experiments. We conducted both the subject-dependent and the subject-independent depression recognition experiments. Then, the average accuracy, F1 value, AUC value, UAR value and confusion matrix were obtained. Finally, there were three discussions: comparisons with other related studies, evaluation indicators and ablation experiments.

4.1 Results

4.1.1 Subject-Dependent Experiment

For the subject-dependent depression recognition experiment, EEG data from all subjects were trained into a model. This experiment was conducted under 10-fold cross-validation; that is, all subjects' data fragments mixed together 90% of the EEG data were used as training data to construct a model, and 10% of the data were used for testing the model. Therefore, the test and training datasets have the same data distribution.

The recognition results on the 4 bands, that is, theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–45 Hz), are shown in Tab. 2. According to the results, the recognition accuracy of depression of the theta band is lower than that of the other three bands. We speculate that the theta band is generated in the state of sleep and a less responsive depressive state. Furthermore, the excitement state of the alpha, beta, and gamma bands increased successively, so the recognition accuracy of depression gradually increased [24]. The iteration of the overall EEG as the efficiency standard of our model is also shown in Tab. 2.

Tab. 2   Comparison of the 3DMKDR model

Dataset	Band	Result	F1/AUC	Result	Iteration
MDD	Theta	95.83%	F1-HCs	99.89%	—
	Alpha	95.97%	F1-MDDs	99.84%	—
	Beta	96.46%	AUC-HCs	99.95%	—
	Gamma	97.83%	AUC-MDDs	99.98%	—
	EEG (all)	**99.86**%	—	—	620
MODMA	Theta	84.76%	F1-HCs	98.27%	—
	Alpha	85.33%	F1-MDDs	98.49%	—
	Beta	88.27%	AUC-HCs	98.76%	—
	Gamma	92.60%	AUC-MDDs	98.54%	—
	EEG (all)	**98.01**%	—	—	550

The F1 value, AUC value, UAR value and confusion matrix were calculated for two classes, MDDs and HCs, to analyze the performance of the 3DMKDR. The higher values of the two classes indicate that the model constructed in this paper has better performance. (0, 0) and (1, 1) are expressed as HCs and MDDs recognition rates, respectively, in which the confusion matrix of the MDD dataset in Fig. 5 (a) and the matrix of the MODMA dataset in Fig. 5 (b). The matrix verifies that our model is stronger than others in predicting depression. Unweighted average recall (UAR) is used as the evaluation criterion for the performance of this model, which is calculated by the following formula:

$$\mathrm{Recall}_i = \frac{\mathrm{TP}_i}{\mathrm{TP}_i + \mathrm{FN}_i} \qquad (3)$$

$$\mathrm{UAR} = \frac{\sum_{i=1}^{N} \mathrm{Recall}_i}{N} \qquad (4)$$

where TP_i and FN_i are the true positive and false negative of the i-th ($i=1,2$) class and predict the number of correct and incorrect for that class, respectively. Recall_i means the predicted correct proportion in all samples of this class. UAR represents the arithmetic average of the recall rates for all classes. In the MDD dataset, we can obtain that $\mathrm{UAR}_{\mathrm{MDD}}$ ($\mathrm{Recall}_1 = 99.87\%$, $\mathrm{Recall}_2 = 99.91\%$) is 99.89%. The $\mathrm{UAR}_{\mathrm{MODMA}}$ ($\mathrm{Recall}_1 = 97.87\%$, $\mathrm{Recall}_2 = 97.93\%$) is 97.90%. UAR can ignore imbalances of samples and realize the objective evaluation of the model. By analyzing

and comparing the recalls of two classes on the same dataset, higher recalls prove that our model has good overall performance.
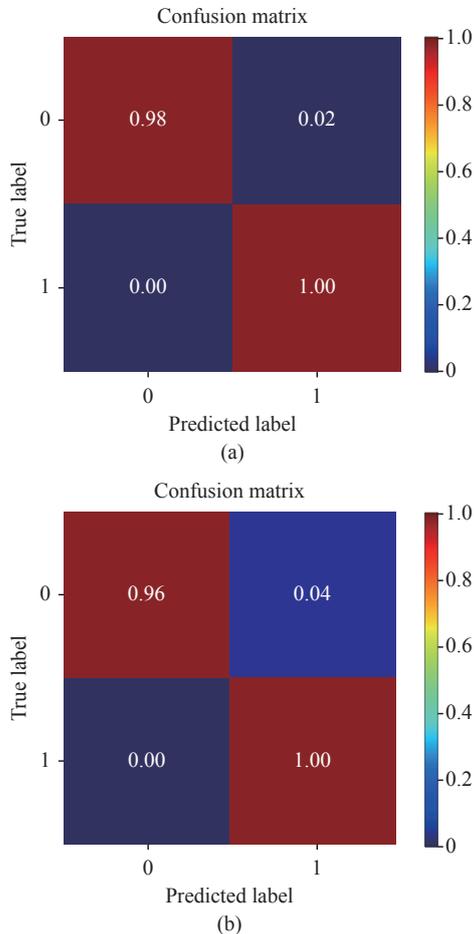

average accuracy is obtained. Tab. 3 and Tab. 4 show the recognition results of LOSO cross-validation in the MDD and MODMA datasets, respectively.

Tab. 3   Accuracy in the MDD dataset

Sub	ACC(%)	Sub	ACC(%)	Sub	ACC(%)
1	97.68	22	96.94	43	81.39
2	97.72	23	93.88	44	91.98
3	96.44	24	96.61	45	89.74
4	96.69	25	97.19	46	98.68
5	96.36	26	98.43	47	99.50
6	95.86	27	91.98	48	98.68
7	94.95	28	92.47	48	98.35
8	96.36	29	99.92	50	98.76
9	97.27	30	96.77	51	96.03
10	95.04	31	93.71	52	95.78
11	96.77	32	98.18	53	90.82
12	96.36	33	97.11	54	92.39
13	96.61	34	98.35	55	92.27
14	95.20	35	97.93	56	98.51
15	94.04	36	98.01	57	96.03
16	93.38	37	97.98	58	97.27
17	96.86	38	98.59	59	96.69
18	96.03	39	94.21	60	95.62
19	94.79	40	98.51	61	96.53
20	97.68	41	98.26	Mean	**95.80**
21	97.72	42	84.04		

Fig. 5   The confusion matrixes of datasets: (a) in the MDD dataset; (b) in the MODMA dataset

Tab. 4   Accuracy in the MODMA dataset

Sub	ACC(%)	Sub	ACC(%)	Sub	ACC(%)
1	50.39	19	50.39	37	99.87
2	99.87	20	100	38	99.87
3	50.39	21	50.39	39	99.87
4	99.87	22	99.87	40	99.87
5	99.87	23	82.08	41	99.87
6	99.87	24	99.87	42	50.39
7	50.39	25	50.39	43	99.87
8	50.39	26	99.87	44	50.39
9	99.87	27	99.87	45	50.39
10	99.87	28	99.87	46	50.39
11	50.39	29	99.87	47	99.87
12	89.58	30	99.87	48	99.87
13	86.25	31	99.87	48	99.87
14	50.39	32	99.87	50	99.87
15	99.87	33	99.87	51	99.87
16	72.92	34	99.87	52	99.87
17	99.87	35	99.87	53	37.92
18	99.87	36	99.87	Mean	**82.27**

4.1.2 Subject-Independent Experiment

For the subject-independent recognition experiment, the model can be trained using data acquired from a limited number of subjects and can be applied for a subject who has never experienced the system prior to the experiment. This experiment was conducted under the leave-one-subject-out cross-validation. For example, in the MDD dataset, a model is trained based on EEG data acquired from 60 participants and is tested on the data acquired from the remaining one subject. This is repeated until each subject is a test subject once. Due to some individual differences between each participant, the test and training sets have different data distributions. Then, the

The experimental results of comparing the two datasets can be summarized as follows. Each

subject in the two datasets has different recognition accuracies because there are differences in the psychological state of different individuals. In addition, during the data collection process of the MDD dataset, the subjects' attention was more concentrated. This leads to these subjects' recognition accuracy in the MDD dataset having relatively small fluctuations, and the average accuracy is higher than that with the MODMA dataset. Some lower recognition results appear in Tab. 4, such as 50.39%, which may be due to the subject distractions in the EEG acquisition experiment of the MODMA dataset. The model structure can obtain information among the EEG electrodes and promote better recognition ability.

4.2 Discussion

4.2.1 Comparison

The 3DMKDR model in this article compared to the previous literature report shows that our model has the ability to improve the accuracy of EEG-based depression recognition. In Tab. 5, the average accuracy of our model is up to 99.86% in the MDD datasets, which is higher than the result of previously researched models, in which the highest accuracy was 99.58%. The average accuracy for MDDs and HCs is up to 98.01% in the MODMA datasets, which is higher than the 94.00% achieved.

Tab. 5 Comparison with previous studies

Dataset	Research	Year	Method	Acc（%）
MDD Dataset	Kaur et al. [27]	2021	Empirical Mode Decomposition	98.51
	Ke et al. [28]	2021	Frequency Electrode+ CNN	99.08
	Loh et al. [25]	2022	CNN+Short-Time Fourier Transform	99.58
	Khosla et al. [29]	2022	Quadratic SVM	99.14
	Ours	2022	3DMKDR	**99.86**
MODMA Dataset	Li et al. [26]	2018	KNN	94.00
	Zhu et al. [30]	2019	CBEM	92.73
	Wang et al. [31]	2021	Self-organizing incremental NN	92.23
	Ours	2022	3DMKDR	**98.01**

It is known that the model based on DL proposed by [25,26] currently has the best performance in the MDD and MODMA datasets. However, the results of the 3DMKDR model are approximately 0.28% and 4.01% higher than those models. Compared with the model [27–29], the multikernel makes the 3DMKDR have higher speed efficiency and better recognition performance than other models in the MDD dataset. Our model adopts the method of repositioning the EEG electrodes to the 2D topology so that the model has higher speed efficiency and better identification performance than other models [30, 31] in the MODMA dataset. The parallel structure and the dot addition part of this model can improve the timeliness and efficient recognition ability of our model. From our experimental results shown in Tab. 5, it can be seen that our results are superior to those of previous studies reported in the literature.

4.2.2 Ablation Experiment

The generic dimension and volume type of the kernel are 2×2, 3×3, 5×5, and 7×7. Conv2×2 (the size of the convolutional kernel is 2×2) cannot find the central point of the convolution, which causes the features of the padding process to offset constantly [32]. A plurality of conv3×3 approximately equal to one conv5×5 or conv7×7 can further extract the features and reduce the number of parameters, and the recognition capability of the model can be controlled by the kernel volume [33]. Thus, this paper applies the conv3×3 multiscale convolutional kernels to increase the feature amount of the model calculation and improve the model recognition.

To confirm the effectiveness of multiscale

kernels, this paper takes the four types of convolutional kernels, i.e., conv3×3×k, conv3×3×k*, conv5×5×k and conv7×7×k. Different types of kernels carry out the same dataset, and each classification result is compared in Tab. 6. Conv3×3×k* represents the multiscale kernel of our model, conv3×3×k represents the same-scale kernel, and k represents the length of the data that was originally set in the third dimension. Based on the comparison of the results of the ablation experiment, this work finds that conv3×3×k is 4.18% and 1.75% higher than conv5×5×k and conv7×7×k, respectively, and conv3×3×k* is 6.76% and 4.33% higher than conv5×5×k and conv7×7×k, respectively, in the MDD dataset. Similarly, the results of conv3×3×k and conv3×3×k* are better than those of conv5×5×k and conv7×7×k in the MODMA dataset.

Tab. 6   Comparison of different kernels

Kernel size	MDD (%)	MODMA (%)
conv3×3×k^*	99.86	98.01
conv3×3×k	97.28	96.89
conv5×5×k	93.10	93.94
conv7×7×k	95.53	94.04

It is obvious that the smaller convolutional kernels pay more attention to the spatiotemporal information between the two electrodes and have the ability to extract effective spatiotemporal features. The larger convolutional kernels cover a larger area in electrode matrixes and extract spatial information between multiple electrodes and contain considerable noise in the features.

Conv3×3×k* is even more than conv3×3×k, with 2.58% and 1.12% more in the two datasets. The recognition results of the multiscale convolutional kernel are higher than those of the same-scale convolutional kernel. The reason is that the multiscale convolutional kernel can control the extraction of features based on different lengths of temporal sampling points. This method has the ability to diversely increase the network

capacity so that the decision function is more distinguished for different categories.

5   Conclusion

In this study, the multiscale kernel 3D-CNN model (3DMKDR) is proposed to effectively recognize MDDs relative to HCs based on EEG signals. By analyzing and comparing the accuracy, F1, AUC and UAR values of the subject-dependent and subject-independent experiments are conducted in the task state EEG data of the MDD and MODMA datasets, which are open depression datasets. The results show that our model can quickly and effectively detect whether the subject is in a major depressive disorder state. The method has been confirmed to have remarkable potential to recognize depression disorders in older people.

While we have achieved superior accuracy when compared to alternative methods as discussed in this paper using EEG data, we consider that there are further potential improvements. For older persons who are inconvenient to move, our future work will explore a better model that uses EEG data from fewer electrodes to efficiently recognize depression based on the resting state EEG data collected by portable EEG acquisition devices. It is anticipated that these methods can facilitate the performance of real-time depressive disorder recognition for seniors, which will improve quality of life for older people with depressive symptomatology.

References：

[1] M. Marcus, M. Taghi Yasamy, M. van Ommeren, D. Chisholm, and S. Saxena, "Depression: A global public health concern, " *WHO Department of Mental Health Substance Abuse*, vol. 1, pp. 6-8, 2012.

[2] World Health Organization, "WHO launches digital app to improve care for older people, " *Geneva,* 2019. https://www.who.int/zh/newsroom/detail/30-09-2019-who-launches-digital-app-to-improve-care-for-older-people.

[3] World Health Organization, "Integrated care for

older people: Guidelines on community-level interventions to manage declines in intrinsic capacity, ” *Geneva*, 2017. https://www.who.int/publications/i/item/9789241550109.

[4] T. Hu, X. Zhao, M. Wu, Z. Li, L. Luo, C. Yang, and F. Yang, “Prevalence of depression in older adults: A systematic review and meta-analysis, ” *Psychiatry Research*, vol. 311, pp. 114511, 2022.

[5] Y. Lin, Y. Long, H. Zhang, Z. Liu, and Z. Zhang. “Psychological stress assessment using multiple physiological signals based on XGBoost, ” *Transactions of Beijing Institute of Technolog,* vol. 42, no. 8, pp. 871-880, 2022.

[6] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, “Feature-level fusion approaches based on multimodal EEG data for depression recognition, ” *Information Fusion*, vol. 59, pp. 127-138, 2020.

[7] P. Sandheep, S. Vineeth, M. Poulose, and D. P. Subha, “Performance analysis of deep learning CNN in classification of depression EEG signals, ” in *IEEE Region 10 Conference (TENCON)*, Kochi, India, pp. 1339-1344, 2019.

[8] W. Shao, Z. You, L. Liang, X. Hu, C. Li, W. Wang, and B. Hu, “A multi-modal gait analysis-based detection system of the risk of depression, ” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 10, pp. 4859-4868, 2022.

[9] X. Deng, X. Fan, X. Lv, and K. Sun, “SparNet: A convolutional neural network for EEG space-frequency feature learning and depression discrimination , ” *Frontiers in Neuroinformatics*, vol. 16, pp. 914823, 2022.

[10] H. Cai, Z. Yuan, Y. Gao, S. Sun, N. Li, F. Tian, H. Xiao, J. Li, Z. Yang, X. Li, Q. Zhao, Z. Liu, Z. Yao, M. Yang, H. Peng, J. Zhu, X. Zhang, G. Gao, F. Zheng, R. Li, Z. Guo, R. Ma, J. Yang, L. Zhang, X. Hu, Y. Li, and B. Hu, “A multi-modal open dataset for mental-disorder analysis, ” *Scientific Data*, vol. 9, no. 1, pp. 1-10, 2022.

[11] E. Salama, R. El-Khoribi, M. Shoman, and S. M. Wahby, “EEG-based emotion recognition using 3D convolutional neural networks, ” *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, pp. 329-337, 2018.

[12] X. Song, D. Yan, L. Zhao, and L. Yang, “LSDD-EEGNet: An efficient end-to-end framework for EEG-based depression detection, ” *Biomedical Signal Processing and Control*, vol. 75, pp. 103612, 2022.

[13] H. H. Jasper, “The ten twenty electrode system of the international federation, ” *Electroence-*

phalography and Clinical Neuroph siology*, vol. 10, no. 2, pp. 371-375, 1958.

[14] W. Mumtaz, L. Xia, M. A. Mohd Yasin, S. S. Azhar Ali, and A. S. Malik, “A wavelet-based technique to predict treatment outcome for major depressive disorder, ” *PloS One*, vol. 12, no. 2, pp. e0171409, 2017.

[15] J. Zhu, C. Jiang, J. Chen, X. Lin, R. Yun, X. Li, and B. Hu, “EEG based depression recognition using improved graph convolutional neural network, ” *Computers in Biology and Medicine*, vol. 148, pp. 105815, 2022.

[16] Q. Zhong, Y. Zhu, D. Cai, L. Xiao, and H. Zhang, “Electroencephalogram access for emotion recognition based on a deep hybrid network, ” *Frontiers in Human Neuroscience*, vol. 14, pp. 589001, 2020.

[17] Y. Su, Z. Zhang, X. Li, B. Zhang, and H. Ma, “The multiscale 3D convolutional network for emotion recognition based on electroencephalogram, ” *Frontiers in Neuroscience*, vol. 16, pp. 872311, 2022.

[18] Y. Zhao, J. Yang, J. Lin, D. Yu, and X. Cao, “A 3D convolutional neural network for emotion recognition based on EEG signals, ” in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, pp. 1-6, 2020.

[19] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen, and S. Su, “Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine, ” in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Italy, pp. 7250-7253, 2015.

[20] A. M. Roy, “ An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces, ” *Biomedical Signal Processing and Control*, vol. 74, pp. 103496, 2022.

[21] J. Li, S. Zhang, and T. Huang, “Multi-scale 3D convolution network for video based person re-identification, ” in *the Thirty-Third AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii, USA, pp. 8618-8625, 2019.

[22] Y. Shimizu, J. Yoshimoto, S. Toki, and Y. Takamura, Masahiro Yoshimura, and Shinpei Okamoto, “Nested 10-fold cross-validation, ” *PLoS One. Figure*, 2015. https://figshare.com/articles/figure/_Nested_10_fold_cross_validation_/1400630.

[23] K. Liao, M. V Mollison, T. Curran, and V. R. d.

Sa, "Single-trial EEG predicts memory retrieval using leave-one-subject-out classification, " in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, pp. 2613-2620, 2018.

[24] Y. Wei, Y. Xu, S. Wang, L. Jin, and T. Ding, "Workers' cerebrocortical activity in hot and humid condition: An electroencephalogram study, " *Journal of Beijing Institute of Technology*, vol. 31, no. 1, pp. 112-122, 2022.

[25] H. W. Loh, C. P. Ooi, E. Aydemir, T. Tuncer, S. Dogan, and U. R. Acharya, "Decision support system for major depression detection using spectrogram and convolution neural network with EEG signals, " *Expert Systems*, vol. 39, no. 3, pp. e12773, 2022.

[26] X. Li, J. Li, B. Hu, J. Zhu, X. Zhang, L. Wei, N. Zhong, M. Li, Z. Ding, J. Yang, and L. Zhang, "Attentional bias in MDD: ERP components analysis and classification using a dot-probe task, " *Computer Methods and Programs in Biomedicine*, vol. 164, pp. 169-179, 2018.

[27] C. Kaur, P. Singh, and S. Sahni, "EEG artifact removal system for depression using a hybrid denoising approach, " *Basic and Clinical Neuroscience (BCN)*, vol. 12, no. 4, pp. 465-476, 2021.

[28] H. Ke, C. Cai, F. Wang, F. Hu, J. Tang, and Y. Shi, "Interpretation of frequency channel-based CNN on depression identification, " *Frontiers in Computational Neuroscience*, vol. 15, pp. 773147, 2021.

[29] A. Khosla, P. Khandnor, and T. Chand, "Automated diagnosis of depression from EEG signals using traditional and deep learning approaches: A comparative analysis, " *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 108-142, 2022.

[30] J. Zhu, Z. Wang, S. Zeng, X. Li, B. Hu, X. Zhang, C. Xia, L. Zhang, and Z. Ding., "Toward depression recognition using EEG and eye tracking: An ensemble classification model CBEM, " in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA, USA, pp. 782-786. 2019.

[31] D. Wang, C. Lei, X. Zhang, H. Wu, S. Zheng, J. Chao, and H. Peng, "Identification of depression with a semi-supervised GCN based on EEG data, " in *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Houston, TX, USA, December 9-12. pp. 2338-2345, 2021.

[32] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks, " in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Zurich, Switzerland, pp. 818-833, 2014.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition, " arXiv preprint arXiv: 1409.1556, 2015.

Yun Su received her Ph. D. degree from Lanzhou University. She is an associate professor at Northwest Normal University. Her current research interests include machine learning in a range of domains including affective computing, bioinformatics, and knowledge base modeling of multimodal physiological signals.

Zhixuan Zhang is currently pursuing the Master 's degree with the School of Northwest Normal University. Her research interests include emotion recognition and validation of agents.

Qi Cai is currently pursuing the Master's degree with the School of Northwest Normal University. His research interests include emotion recognition.

Bingtao Zhang received the Ph. D. degree in Computer Application Technology from Lanzhou University in 2019. His research interests include data mining, biomedical engineering.

Xiaohong Li received her Master's degree from Lanzhou University of Technology. She is an associate professor in Northwest Normal University. Her current research interests include machine learning and intelligent information processing, with research areas focusing on community discovery, recommendation systems, and text processing