Enhancing EEG-Based Classification of Depression Patients using Spatial Information

Chao Jiang, Yingjie Li*, *Member, IEEE*, Yingying Tang, Cuntai Guan*, *Fellow*, *IEEE*

***Abstract*—Background: Depression has become a leading mental disorder worldwide. Evidence has shown that subjects with depression exhibit different spatial responses in neurophysiological signals from the healthy controls when they are exposed to positive and negative stimuli. Methods: We proposed an effective electroencephalogram-based detection method for depression classification using spatial information. A face-in-the-crowd task, including positive and negative emotional facial expressions, was presented to 30 participants, including 16 depression patients and 14 healthy controls. Differential entropy and the genetic algorithm were used for feature extraction and selection, and a support vector machine was used for classification. A task-related common spatial pattern (TCSP) was proposed to enhance the spatial differences before the feature extraction. Results and discussion: We achieved a leave-one-subject-out cross-validation classification result of 84% and 85.7% for positive and negative stimuli, respectively, using TCSP, which is statistically significantly higher than 81.7% and 83.2%, respectively, acquired without the TCSP (p<0.05). We also evaluated the classification performance using individual frequency bands and found that the contribution of the gamma band was predominant. In addition, we evaluated different classifiers, including k-nearest neighbor and logistic regression, which showed similar trends in the improvement of classification by employing TCSP. Conclusion: The results show that our proposed method, employing spatial information, significantly improves the accuracy of classifying depression patients.**

***Index Terms*— Depression, EEG classification, Task-related common spatial pattern**

This research was supported by National Natural Science Fund of China (No. 61571283), Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01) and ZJLab; Shanghai Science and Technology Committee Foundations (Nos. 16ZR1430500, 19411969100, and 19410710800).

Chao Jiang is with the School of Communication and Information Engineering, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444 China, with the College of Electronics and Information Engineering, Shanghai University of Electric Power, Shanghai 200090 China (e-mail: jiangchao@shiep.edu.cn).

*Yingjie Li is with Shanghai Institute for Advanced Communication and Data Science, Institute of Biomedical Engineering, School of Communication and Information Engineering, Shanghai University, Shanghai 200444 China (corresponding author phone: 086-21-66136902; e-mail: liyj@i.shu.edu.cn).

Yingying Tang is with Shanghai Mental Health Center, Shanghai 200030, China (e-mail: yytang0522@gmail.com).

*Cuntai Guan is with the School of Computer Science and Engineering, Nanyang Technological University, 639798 Singapore (corresponding author phone: (65) 6790 6205; e-mail: ctguan@ntu.edu.sg)

I. Introduction

DEPRESSION, as a common illness worldwide, is classified as a mood disorder and described as feelings of sadness or anger that interfere with a person's everyday activities [1]. According to the World Health Organization, it is likely to be the leading global disease by 2030 [2].

Depression disorder is a pathological process that causes many symptoms, resulting in limited mental and physical functionality [3]. It is often accompanied by cognitive impairments, which may increase the risk of Alzheimer's disease and suicide and accelerate cognitive decline [4]**.** The earlier depression is detected, the easier it is to treat. As a low-cost, noninvasive acquisition, and high temporal resolution technique, electroencephalography is widely used in neural systems and rehabilitation engineering [5][6]. Acharya *et al.* proposed a typical computer-aided system for electroencephalogram (EEG)-based diagnosis of depression, which primarily includes an offline and online system [7]. This paper is focused on the experimental paradigm, emotion feature extraction, feature selection, machine learning, and the dataset for training and testing, particularly on spatial information feature extraction and selection. This focus was chosen because many studies have shown that subjects with depression exhibit different spatial responses in neurophysiological signals compared to healthy controls, when they are exposed to stimuli [8–11].

Many studies have been conducted on depression; some studies focused on the resting-state [11–15], whereas others focused on tasks [16–18]**.** For example, Li *et al.* performed a study on the EEG-based brain electrical source of mildly depressed subjects, which suggested that depressed subjects spent more time viewing negative emotional faces, causing a dysregulation in temporal pole activity [19]. Liao *et al.* collected 54 resting-state EEG signals, 6 s in length, from 12 patients with depression and 12 healthy controls [11], and Yang *et al.* extracted 24 resting-state EEG signals, 8 s in length, from 17 depressed patients and 17 control subjects; both studies resulted in a classification accuracy above 80% [13]. Wu *et al.* recorded the EEG for a resting-state session, followed by an emotion-induction session, from 55 participants (24 with major depressive disorder (MDD) and 31 healthy controls), considering that the ability to distinguish MDD using resting-state EEG reaches a bottleneck, which provides a higher accuracy than emotion-induction EEG of above 83% [14]. Li *et al.* conducted an experiment on the facial expression viewing task (emotional and neutral blocks) involving 48 college students, 24 of whom were considered depressed and 24 healthy, which provided an accuracy of 85.62% for the detection of depressed and healthy students using a convolutional neural network [15]. The participants of our experiment were from the Shanghai Mental Health Center and included 16 patients with depression (Dep) and 14 healthy controls (HC); the participants were presented with the face-in-the-crowd task stimuli of six human faces.

EEG signals are nonstationary and nonlinear signals, similar to many other physiological signals [20]. To analyze these signals, linear and nonlinear features are typically used, such as the power spectrum density, Lempel-Ziv complexity, variance, mobility, fluctuations, Higuchi fractal, approximate entropy, Kolmogorov entropy, correlation dimension, Lyapunov exponent, and permutation entropy [2][8][9]. To analyze our hypothesis effectively, it was necessary to select optimal features, as some dimension features may mislead the classifiers. The BestFirst, GreedyStepwise (GSW), GeneticSearch, and RankSearch approaches, based on correlation feature selection, are typical data mining search methods, and the BayesNet, support vector machine (SVM), k-nearest neighbor (KNN), logistic regression (LR), linear discriminant analysis (LDA), and random forest approaches are widely used for discriminating classes [8–10][21]. Hosseinifard *et al.* extracted four nonlinear features from EEG signals and obtained the highest classification accuracy of depressed patients and controls by using the correlation dimension and LR approaches, among the KNN, LDA, and other nonlinear feature selection methods, with a genetic algorithm employed to select features [6]. Li *et al.* calculated eight linear features and nine nonlinear features from theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) waves, which yielded high accuracy using GSW and KNN for the beta frequency band [9]. As the electrode channels are located in different areas on the surface of the human's head, the channel dimension contains spatial information of EEG. When EEG channels are chosen, the optimal spatial information should be selected. The common spatial pattern (CSP) has been proven to be one of the most effective algorithms for a brain-computer interface (BCI) for the optimization of the spatial-spectral filter, and many novel approaches have been proposed accordingly [22][23]. This paper presents an effective EEG-based detection method for depression classification by employing spatial information, namely the task-related common spatial pattern (TCSP).

Subject-independent k-fold cross-validation (CV) [24][25] and leave-one-subject-out (LOSO) CV [26–28] are two widely used EEG classification strategies. In fact, when $k = 1$, the LOSO method is a special case of the k-fold technique. As the LOSO approach can enjoy more training data and adjust super-parameters on each subject, it will always achieve better results compared with the k-fold method. When detecting a potential depression patient, we chose the LOSO strategy to evaluate the model for detecting depression patients in this study, to make the best use of the existing data.

In this study, we conducted an emotion-induction experiment for the Dep group and HC group, who responded to the face-in-the-crowd task stimuli of six human facial expressions. We used two types of emotion stimuli: positive and negative. Differential entropy (DE) and genetic algorithm (GA) were used for feature extraction and feature selection, and SVM was used for classification, which has been recognized as making a significant contribution to EEG classification in previous studies [8][10][21][30][31]. TCSP was used to enhance the spatial differences before the feature extraction.

The remainder of this paper is organized as follows. Section II explains the experiment protocol, EEG signal processing, feature extraction and feature selection, machine learning, and assessment methods. Section III presents the results and Section IV provides a discussion. Finally, Section V summarizes the conclusions of the study.

II. Research Methodology

The objective of our experiment was to develop an effective EEG-based detection method for depression classification by employing spatial information. Fig. 1 shows the block diagram of EEG signal processing used in our study, which mainly consists of the experiment protocol, preprocessing, feature extraction, feature selection, machine learning, and statistical analysis.

A. Participants and Procedure

This study was approved by the Institutional Review Board of the Shanghai Mental Health Centre (SMHC) [32][33]. The Dep group included 16 right-handed diagnosed outpatients with depression (male/female = 6/10, 37.75 ± 14.19 years old, 12.06 ± 2.91 years of education), recruited from SMHC. The HC group included 14 right-handed healthy participants (male/female = 4/10, 40.86 ± 12.29 years old, 11.54 ± 3.75 years of education) with no personal history of neurological or psychiatric illness. Before the experiments were conducted, all participants underwent an interview in which the Hamilton rating scale for depression (HAMD, Dep: 24.5 ± 7.40, HC: 7.27 ± 6.94) was administered. The self-rating anxiety scale (SAS, Dep: 61.3 ± 9.74, HC: 35.5 ± 5.13) and self-rating depression scale (SDS, Dep: 0.89 ± 0.08, HC: 0.48 ± 0.09) were assessed by the subjects.

The face-in-the-crowd task stimuli consisted of six human faces, which were selected from the Ekman emotion database [28]. There were three types of expressions (positive, negative, and neutral) without hair, glasses, beard, or other facial accessories. The experiment contained 4 blocks, 2 positive and 2 negative target blocks, and each block had 144 trials. During the positive blocks, 72 positive, 36 negative, and 36 neutral stimuli were presented to the participants, and during the negative blocks, 72 negative, 36 positive, and 36 neutral stimuli were presented to them. As shown in Fig. 1, each trial was displayed for 1500 ms against a black background. Then, an interstimulus interval of 1000 ms was presented, during which a fixation cross appeared alone in the center of the screen. The experiment followed a GO-NOGO paradigm [31–36].

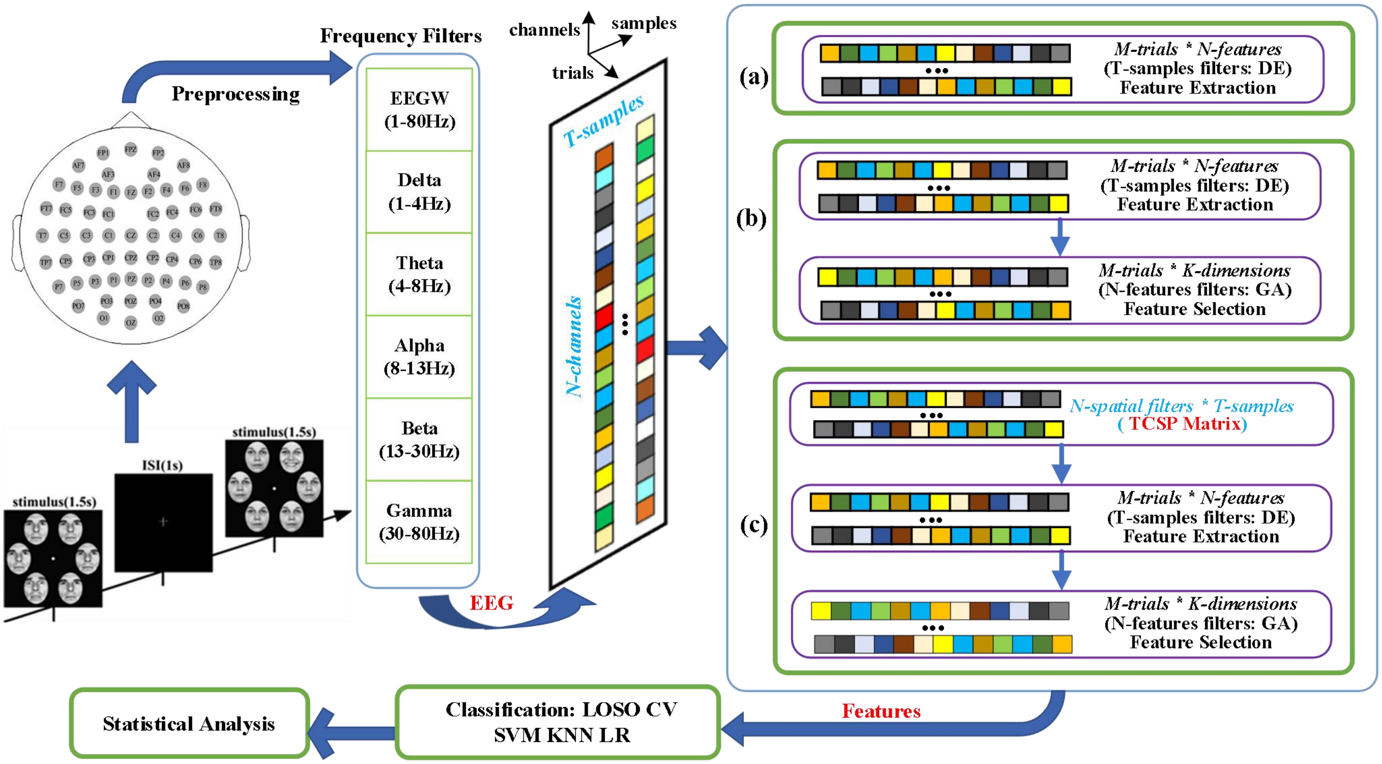

Fig. 1. Block diagram of EEG signal processing. After frequency filtering, feature extraction of the time information and feature selection of the channel information were applied to the EEG signals. There were three classification strategies: (a) a traditional method using all channels without feature selection, (b) a typical method using feature selection without employing the TCSP, and (c) our proposed method using a GA with the TCSP. The EEG signal of one trial consists of the number of channels times the number of samples (*N-channels* * *T-samples*). In case (a), *M-trials* represents the set of trials, and *N-features* represents the set of features. In case (b), *k-dimensions* represents the selected features subset of *N-features*. In case (c), *N-spatial* filters represents the projection of *N-channels* after using the TCSP matrix. By employing the TCSP matrix, it was expected that the EEG-based classification performance of depression patients could be enhanced.

The participants were comfortably seated 80 cm away from a 17-inch LCD-screen and were asked to judge whether the present image contained the target face during the stimulus onset asynchrony. During the positive block, the participants were asked to press button “1” if positive face stimuli were found, and during the negative blocks, the participants were asked to press button “5” if negative face stimuli were found. There was a break period of 1 min between blocks, and the whole experiment took approximately 30 min for each subject.

The EEG signals were recorded at a sampling frequency of 1000 Hz from 64 channel surface electrodes (QuickCap™, Brain Products Inc., Gilching, Bavaria, Germany), and the interelectrode impedance was maintained below 5 kΩ. Data recording was referenced to the tip of the nose. Artifacts from vertical and horizontal eye movements and blinks were removed offline by an ocular correction algorithm using a Brain Vision Analyzer (Brain Products Inc., Gilching, Bavaria, Germany). Fifty-nine electrodes were selected from the 64 electrodes that covered the whole scalp. As shown in Fig. 1, we preprocessed the EEG signals so that we could extract optimal features. The artifact-free data were band-pass-filtered between 0.05 and 100 Hz. The data were segmented from 200 ms before stimulus onset to 1000 ms post stimulus. Segmentations with artifacts (>±100 μV) or those leading to incorrect answers were excluded. In this study, we selected the EEG signal 200 ms before the stimulus onset as the baseline-EEG and the EEG signal 1000 ms post-stimulus as the task-EEG.

B. Task-related Common Spatial Pattern Matrix

Research over the past decade has shown that spatial information can effectively contribute to the detection of Dep [9–13]. The objective of this study was to find an effective method to process the detection between Dep and HC. The common spatial pattern is a mathematical procedure widely used in signal processing to separate a multivariate signal into additive subcomponents [30][31], with the effects of feature extraction [22][23]. Traditional CSP and its variants considerably contribute to the classification performance during motor imagery and emotion recognition [37][38][39–42]. Based on the filter bank common spatial pattern framework, Park *et al.* proposed a filter bank regularized CSP [22]. When calculating the estimated covariance, the trials of subjects that exclude the interested one are used to calculate the estimated covariance. In our study, there were two participant groups (Dep and HC) and two task stimuli (positive and negative emotions). For session 1 (positive stimuli blocks), we labeled the trials of Dep as class 1 and the trials of HC as class 0. Additionally, we labeled the trials of Dep as class 1 and the trials of HC as class 0 for session 2 (negative stimuli blocks). When calculating the estimated covariance, we used all the trials of the subjects that exclude the interested/test one (Dep as class 1 and health as class 0), which implies that the EEG trials of covariance 1 increase with depression, and the EEG trials of covariance 0 increase with health. As shown in Fig. 2, we obtained the training dataset and test dataset according to the LOSO strategy and calculated the TCSP projection matrix using the training dataset. We then put

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2021.3059429, IEEE Transactions on Neural Systems and Rehabilitation Engineering

the estimated TCSP projection matrix into the test dataset to differentiate Dep from HC.

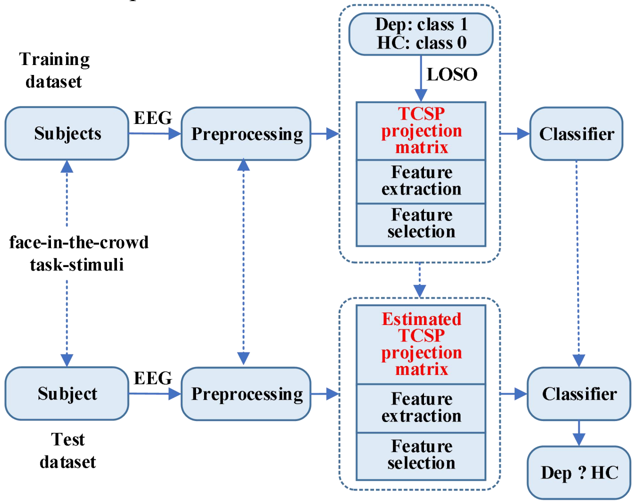

Fig. 2. Use of TCSP projection matrix to enhance the classification.

As shown in Fig. 1, after frequency filtering, the EEG signal X of one trial consists of channels and time sampling points (*N-channels* * *T-samples*). Based on the CSP, our objective was to find the TCSP projection matrix, which could transform X in the original sensor space into a new space, where the time series can contain more discriminative information.

The averaged spatial covariance matrices of class 1 (X_1) and class 0 (X_2) were calculated respectively as shown below:

$$R_1=\frac{1}{m_1}\sum_{i=1}^{m_1}\frac{x_{1i}x_{1i}^T}{trace\left(x_{1i}x_{1i}^T\right)} \tag{1}$$

$$R_2=\frac{1}{m_2}\sum_{i=1}^{m_2}\frac{x_{2i}x_{2i}^T}{trace\left(x_{2i}x_{2i}^T\right)} \tag{2}$$

where T denotes the transpose of a matrix, trace(.) denotes the sum of the diagonal elements of the matrix, X_{1i} (i=1, … m_1) and X_{2i} (i=1, … m_2) are the EEG training data, and m_1 and m_2 are the numbers of the EEG training data in class 1 and class 0, respectively. The composite covariance in class 1 and class 0 can be diagonalized as,

$$R=R_1+R_2 \tag{3}$$

$$R=U\lambda U^T \tag{4}$$

where λ is an *N-channels* * *N-channels* diagonal matrix, in which the diagonal elements are the eigenvalues of R, sorted in descending order, and U is the eigenvector matrix of R.

The whitening matrix and eigen decomposition were then calculated as

$$P=\sqrt{\lambda^{-1}}U^T \tag{5}$$

$$S_1=PR_1P^T, S_2=PR_2P^T \tag{6}$$

and diagonalized as

$$S_1=B_1\lambda_1B_1^T, S_2=B_2\lambda_2B_2^T \tag{7}$$

where S_1 and S_2 share the same eigenvalues. Thus, we obtained B_1=B_2=B and λ_1+λ_2=I, i.e., the larger the eigenvalues for S_1, the smaller the eigenvalues for S_2.

$$W=B^TP \tag{8}$$

Then we obtained the TCSP projection matrix, which was a spatial filter. We could obtain X' (*N-spatial filters** *T-samples*) by using the TCSP projection matrix to transform X in the original sensor space into a new space, enhancing the discriminative information.

C. Feature Extraction and Feature Selection

As is well known, entropy can be utilized to describe the degree of signal irregularity in dynamical systems. Previous studies have shown that changes in entropy may reflect changes in brain activation when conducting cognitive tasks [41]. DE has been proven to be an effective feature in emotion recognition [36–39]. In this study, delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–80 Hz) waves and wideband EEG (EEGW) were extracted by wavelet packet decomposition.

Fig. 1 shows three classification strategies: (a) a traditional approach using all channels without feature selection, (b) a typical strategy using a GA without employing the TCSP, and (c) our proposed method using a GA with the TCSP. In strategies (a) and (b), DE is extracted from delta, theta, alpha, beta, and gamma as well as EEGW. In strategy (c), the TCSP is used to enhance the feature performance; thus, DE is extracted from the new delta, theta, alpha, beta, and gamma waves as well as EEGW, as constructed using the TCSP projection matrix. The formula for calculating DE can be expressed as

$$DE=\int_{-\infty}^{\infty}P(x)\log\left(P(x)\right)dx \tag{9}$$

Here, x denotes a random variable, and $P(x)$ is the probability density function of x. We assume that the EEG signals obey a Gaussian distribution: $x \sim N(\mu, \sigma^2)$. Then, the DE calculation can be simplified as

$$DE=\frac{1}{2}\log 2\pi e\delta^2 \tag{10}$$

For a segment of EEG, the DE estimation is equivalent to the logarithm energy spectrum in a particular frequency band [43]. In addition, the logarithm energy spectrum can effectively eliminate the problem of low-frequency energy typically having a relatively higher magnitude than high-frequency energy in EEG [44].

Feature selection is widely used to improve performance because some dimension features will mislead the classifiers [30][31]. This method will also reduce the computational complexity compared with that when all the features are used.

The GA is a metaheuristic inspired by the process of natural selection in evolutionary algorithms, which is a global optimization solution. Using the GA, chromosomes were coded in this study as follows.

For each trial, the number of features was FD. A vector of length FD defined one chromosome F_{GA}= {f_1, …, f_i, …, f_{FD}}, $f_i \in$ (0 1). Each bit in the chromosome corresponded to one of the features and indicated whether the corresponding feature was selected or not. We used a learning curve to obtain the optimal number of features FS within the range of 2–59 and initialized FS as 2. When feature selection began, a random candidate chromosomes population was first generated. We initialized the population size as P_n = 100, indicating that there were 100 random candidate chromosomes in the optimization problem. Furthermore, a fitness function was used as the optimization objective [8][10]. Here, we used the corresponding classifier as the fitness function, and the fitness value of each chromosome was the classification during training. After calculating the fitness value of each candidate chromosomes, we used a roulette wheel selection algorithm to select the potentially useful chromosomes for recombination, with a selection rate of 0.5. We randomly (crossover rate 0.7) chose the parent chromosomes among the potentially useful chromosomes to create new chromosomes by recombination. According to the selection rate, we needed 50 kid chromosomes. Then, the kid chromosomes mutated randomly with a mutation rate of 0.001, producing a new population of 100 chromosomes. The chromosome selection, crossover, mutation, and population were updated and iterated until the function no longer produced improved results when given a certain *FS*. Then, we obtained a potentially optimal subset of chromosomes $\{F_{GA1}, F_{GA2}, \ldots, F_{GA100}\}$ and chose the optimal chromosome with the maximum fitness value, $max\ (fitness\ (\{F_{GA1}, F_{GA2}, \ldots, F_{GA100}\})$. By calculating the learning curve of *FS*, we obtained the final optimal chromosome, $max\ (max\ (fitness\ (\{F_{(GA1,\ j)}, F_{(GA2,\ j)}, \ldots, F_{(GA100,\ j)}\})),\ j \in (2\ 59)$. The features of this chromosome could produce the best classification performance in the training dataset; therefore, when a new subject EEG was given to the classifier, we could choose the corresponding selected features.

D. Machine Learning and Assessment Methods

The SVM has been employed widely in different classification and regression problems. The performance of the SVM is affected by the kernel function, which may be a linear, radial basis, sigmoid, or polynomial function [25][45]. A library for support vector machines was used for classification, using the SVM-SVC (support vector classification, SVC) model with a linear function, C-SVC of cost 1. Additionally, we used the KNN and LR approaches [46], which are widely used in BCIs, to make our results more robust.

The performance accuracy (ACC), precision, recall, and F1 score were calculated in this study. As the LOSO classification strategy was used, ACC was equal to the recall, the precision was 1, and F1 was always larger than ACC.

In addition, the Wilcoxon signed-rank test was used to calculate the statistically significant difference of our experiment results, e.g., the statistical significance of the accuracy enhancement using the TCSP.

III. Results

Based on the face-in-the-crowd task stimuli, we recorded and preprocessed the EEG signals; the frame flow chart is shown in Fig. 1. We selected the EEG signal 200 ms before the stimulus onset as the baseline-EEG and the EEG signal 1000 ms post-stimulus as the task-EEG. We evaluated the TCSP performance with the two EEG signals.

A. TCSP Performance under Task-EEG

To observe the performance improvement with the TCSP, we utilized three classification strategies: (a) a traditional method using all channels without feature selection; (b) a typical method using feature selection without employing the TCSP, where we used a GA; and (c) our proposed method using a GA with the TCSP. Table I presents the results of the task-EEG under positive and negative stimuli, respectively. The average ACC and standard deviation (SD) are used here.

We achieved a LOSO CV classification result of 84% and 85.7% for positive and negative stimuli, respectively, by using the TCSP, and a classification result of 81.7% and 83.2%,

TABLE I. LOSO CV Classification Performance for Task-EEG

Bands	ACC ± SD (100%)					
	Task-EEG under Positive-Stimuli			Task-EEG under Negative-Stimuli		
	Channels without Feature Selection	Feature Selection with GA		Channels without Feature Selection	Feature Selection with GA	
		Without TCSP	With TCSP		Without TCSP	With TCSP
EEGW	0.603±0.133	0.698±0.089	0.757±0.079	0.615±0.128	0.736±0.082	0.768±0.089
Delta	0.562±0.126	0.615±0.083	0.632±0.083	0.593±0.127	0.627±0.081	0.638±0.076
Theta	0.568±0.139	0.612±0.072	0.627±0.081	0.596±0.131	0.632±0.078	0.653±0.075
Alpha	0.583±0.126	0.628±0.079	0.652±0.077	0.609±0.118	0.646±0.081	0.681±0.083
Beta	0.595±0.138	0.697±0.082	0.738±0.082	0.612±0.129	0.728±0.080	0.776±0.087
Gamma	0.627±0.119	0.739±0.091	0.789±0.075	0.643±0.122	0.789±0.089	0.807±0.089
6-Bands	**0.665±0.121**	**0.817±0.079**	**0.840±0.064**	**0.686±0.126**	**0.832±0.080**	**0.857±0.071**

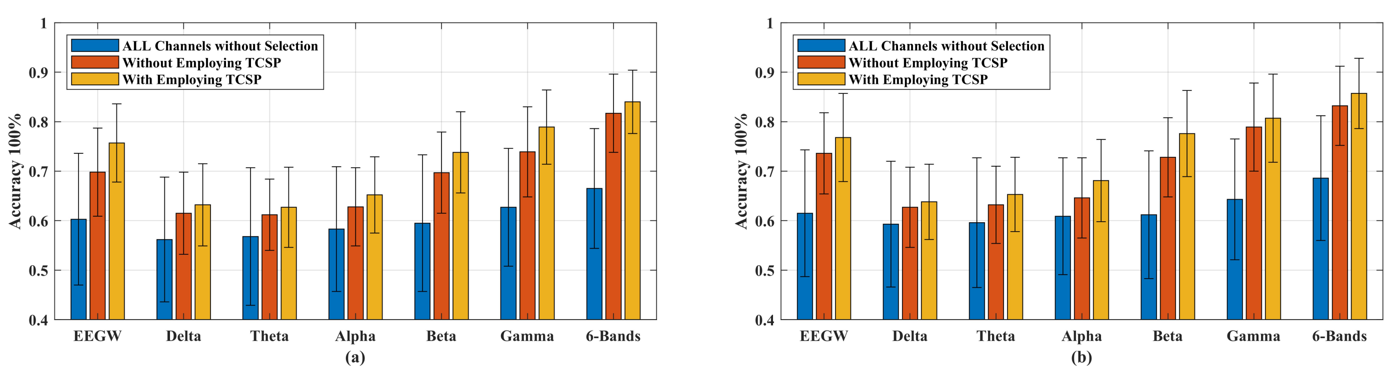

Fig. 3. LOSO classification performance bar chart of different bands for (a) task-EEG under positive-stimuli and (b) task-EEG under negative-stimuli. The spatial information contributes to the classification performance, and the TCSP can enhance the spatial differences. ACC is higher when the TCSP is used than when it is not used. ACC in the 6-Bands case achieves the best classification performance.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2021.3059429, IEEE Transactions on Neural Systems and Rehabilitation Engineering

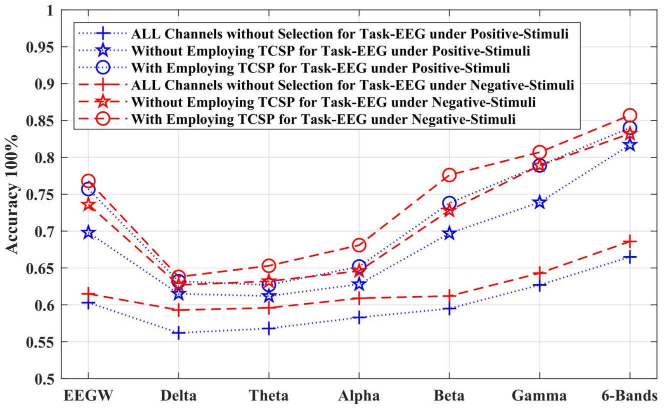

Fig. 4. LOSO classification performance line chart for task-EEG under positive and negative stimuli

respectively, without using the TCSP. When using all the channels without feature selection, the classification results were only 66.5% and 68.6% for the positive and negative stimuli, respectively, which are significantly lower than the values obtained when using feature selection.

From Fig. 3 and Fig. 4, we can observe that the classification performance with TCSP is better than that without TCSP in the EEGW, delta, theta, alpha, beta, gamma, and 6-Bands cases. We obtained the highest ACC when employing TCSP in 6-Bands as it provides more frequency information and spatial information to the feature selection mechanism and classifiers. Furthermore, we observe that the gamma band predominantly contributes to the classification performance, which is consistent with previous studies [4][33].

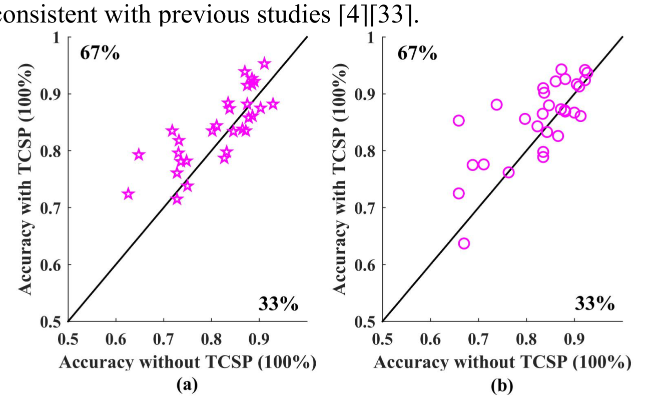

Fig. 5. Employing TCSP vs. not employing TCSP per subject for task-EEG under (a) positive and (b) negative stimuli in the 6-Bands case. The markers in the coordinate system are mostly located in the upper left part, which indicates that the performance with the TCSP is better than that without the TCSP. There is a similar trend for the EEGW, delta, theta, alpha, and beta cases.

Moreover, we plotted Fig. 5 and Fig. 6, which intuitively illustrate the classification performance with the TCSP versus without the TCSP for each subject detection. Fig. 6 is a scatter diagram, which indicates that for the majority classification of subjects, the ACC of employing TCSP is higher than that of not employing TCSP for task-EEG under positive-stimuli and under negative-stimuli, respectively, in 6-Bands. Fig. 6 shows three types of lines obtained by using the GA and by employing or not employing the TCSP for the task-EEG under positive and negative stimuli. When we sorted the accuracy values acquired by using the GA without the TCSP for 30 subjects, we obtained the first line, called condition 1. According to the sequence of subjects of condition 1, we acquired the paired accuracy value using the GA when the TCSP was employed, which is the second line, called condition 2. When we sort the accuracy values of condition 2, we obtained the third accuracy value line, called condition 3. It can be observed that the overall performances of conditions 2 and 3 are better than that of condition 1. If we obtained a larger dataset, the graph of condition 2 would tend to that of condition 3.

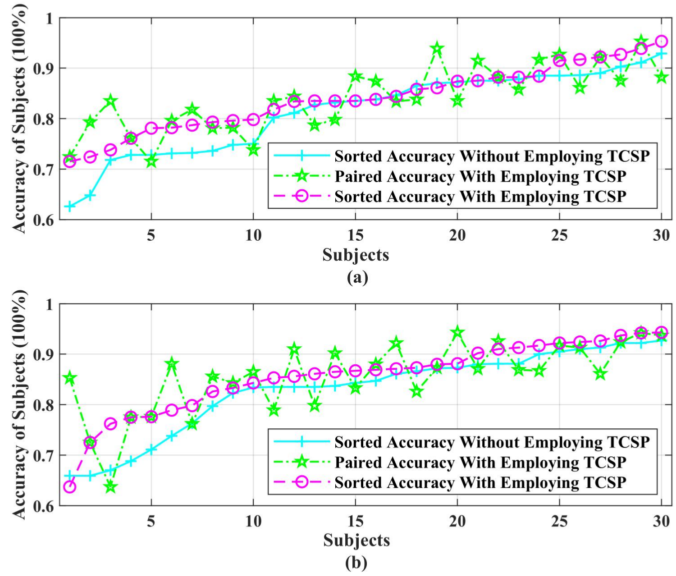

Fig. 6. Classification performance per subject for task-EEG under (a) positive and (b) negative stimuli in the 6-Bands case. The sorted accuracy without employing the TCSP of subjects (named condition 1) is the accuracy of using the GA without employing the TCSP. By employing the TCSP, we obtained two other types of graphs: the paired accuracy of using the GA and the TCSP, corresponding to the subject sequence of condition 1 and called condition 2, and the sorted accuracy of condition 2 for the subjects, called condition 3. The performance of condition 2 is better than that of condition 1, which is also proved in Fig. 6 and is consistent with the results in Table I, where the average accuracies of condition 2, 84% and 85.7%, are higher than those of condition 1, 81.7% and 83.2%. For larger numbers of subjects, the graph of condition 2 will tend to that of condition 3 and the performance of the TCSP.

B. *TCSP Performance under Baseline-EEG*

As is well known, when presented with task stimuli, participants will respond to the stimuli, and the brain activation will change with psychological and physiological activities. In Subsection III.A, we evaluated the TCSP performance under task-EEG. In this subsection, we will discuss the TCSP performance under the baseline-EEG. There are three classification strategies: (a) a traditional method using all channels without feature selection, (b) a typical method using a GA without employing the TCSP, and (c) our proposed method using a GA with the TCSP.

Fig. 7 shows the results of the baseline-EEG under positive and negative stimuli. We achieve classification results of 72.8% and 73.6% for baseline-EEG under positive and negative stimuli, respectively, using the TCSP, and the classification results of 70.9% and 71.5%, respectively, without the TCSP, in the 6-Bands case. When using all channels without feature selection, we obtain the classification results of 63.5% and 63.7% in 6-Bands, respectively, which are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2021.3059429, IEEE Transactions on Neural Systems and Rehabilitation Engineering

significantly lower than the values obtained with feature selection, similar to the task-EEG results. Further, the classification performance with the TCSP is better than that without the TCSP. The highest ACC was also obtained in the 6-Bands case, which was a combination of all six traditional frequency bands, and the gamma frequency band contributes more to the classification performance, which is consistent with the task-EEG results under positive and negative stimuli.

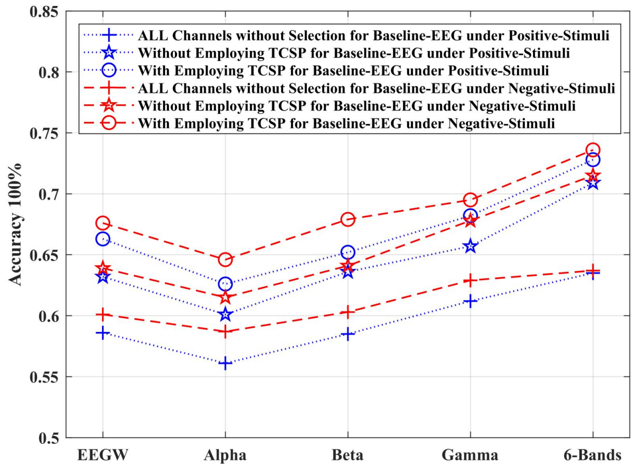

Fig. 7. LOSO classification performance line graph for baseline-EEG under positive-stimuli and baseline-EEG under negative-stimuli. ACC is higher with TCSP than without TCSP, and it is highest in the 6-Bands case.

C. Statistical Results and Significant Improvement

Fig. 8 shows the statistical results for baseline-EEG and task-EEG under positive-stimuli, and negative-stimuli, respectively, for the 6-Bands frequency case (* (P<0.05), ** (P<0.01), and *** (P<0.001)). This section focuses on the difference between the strategies and the difference between the task-EEG and baseline-EEG.

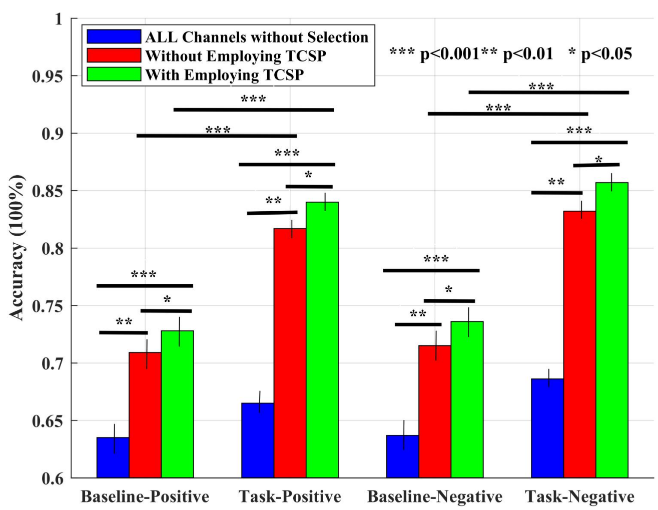

Fig. 8. Significant results for baseline-EEG and task-EEG under positive-stimuli and negative-stimuli in the 6-Bands case. Here, Baseline-Positive represents the baseline-EEG under positive-stimuli, Task-Positive represents the task-EEG under positive-stimuli, Baseline-Negative represents the baseline-EEG under negative-stimuli, and Task-Negative represents the task-EEG under negative-stimuli.

For the three strategies of using all channels without feature selection, using a GA without the TCSP, and using a GA with the TCSP, there are similar statistical results for the baseline-EEG and task-EEG under positive and negative stimuli in the 6-Bands case (using all channels without feature selection vs. using a GA without the TCSP, using all channels without feature selection vs. using a GA with the TCSP, using a GA without the TCSP vs. using a GA with the TCSP). Based on the spatial information, there is a significant performance improvement (using a GA without the TCSP vs. using all channels without feature selection and using a GA with the TCSP vs. using all channels without feature selection), which shows that spatial information contributes to the classification performance. Employing the TCSP also yields significant improvement (using a GA with the TCSP vs. using a GA without the TCSP), which reveals that employing the TCSP can enhance the spatial differences. Additionally, there is a consistent classification performance trend for the baseline-EEG and task-EEG. Further, significant classification improvements are evident for the task-EEG compared to the baseline-EEG and with the use of feature selection, with or without employing the TCSP, which may occur because of the amount of data in the task-EEG or because brain activation may change more in the task-EEG than in the baseline-EEG.

IV. Discussion

A. Improvement of Employing TCSP

Many studies have focused on depression and the CSP separately. Li *et al.* [9], Liao *et al.* [11], and Li *et al.* [21] achieved high classification of depression patients and HC by selecting optimal spatial information features. The CSP has been widely used to transform original sensor spaces into new spaces in which the time series can contain more discriminative information. In addition, a previous study showed that entropy can be utilized to describe cognitive performance and will decrease as cognitive performance decreases. Further, the entropy of depression patients was lower than that of healthy subjects when performing cognitive tasks [41].

Our previous study showed that the attention model for HC balances the forward input and feedback; however, there is increased forward input (from the posterior cortex to the central and anterior cortices and from the central cortex to the anterior cortex) and decreased feedback (from the anterior cortex to the central and posterior cortices and from the central cortex to the posterior cortex) in Dep [32], which may induce the difference in DE in different brain regions between Dep and HC. Additionally, we observed that the brain networks of both Dep and HC in gamma oscillation presented regular network characteristics during emotional processing; however, Dep showed randomization trends in [33], which may emphasize the difference in DE for different bands between Dep and HC.

By employing the TCSP, we achieved classification results of 84% and 85.7% for the task-EEG under positive and negative stimuli, respectively, showing a statistically significant difference, compared to 81.7% and 83.2%, respectively, without the TCSP. Furthermore, these results are significantly higher statistically than the values of 66.5% and 68.6%, when using all channels without feature selection. There is a similar trend for the baseline-EEG. The spatial information contributes to the classification performance, and the TCSP can enhance the spatial differences before feature

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2021.3059429, IEEE Transactions on Neural Systems and Rehabilitation Engineering

extraction to yield higher classification performance.

B. Brain Region Selection

In this study, during the training phase, we used the GA to select features in different bands. The optimal numbers of features of the different bands were not identical and ranged from 8 to 12. There was one feature for each channel, and the electrode channels were located in different areas on the surface of the human head. Thus, feature selection by the GA is also a brain region selection process, and the brain includes many different areas that have particular functions. The frontal area is primarily used for concentration and emotional reactions; the central and temporal areas are used for emotion and motion; and the parietal and occipital areas are used for recognition, attention, and sight.

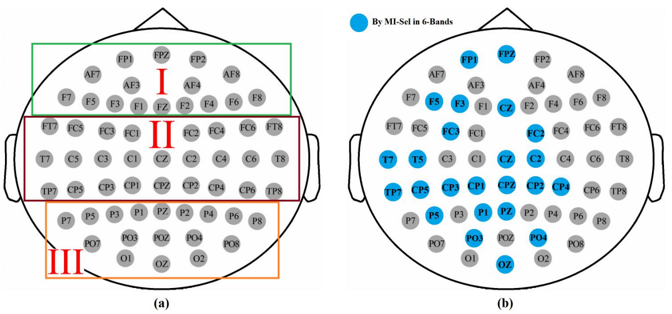

Fig. 9. Brain Regions based on frontal, central and temporal, and parietal and occipital regions, and MI-Sel.

As shown in Fig. 9 (a), we divided all the channels into three parts according to the brain regions: frontal (I) with 16 channels, central and temporal (II) with 26 channels, and parietal and occipital (III) with 17 channels. In addition, we used mutual information (MI) to select 24 channels that were distributed in the frontal, central, temporal, parietal, and occipital areas, called MI-Sel [47][48], as shown in Fig. 9 (b). We also used TCSP and GA in the four areas: I, II, III, and MI-Sel. The classification performance bar graph of the different regions is shown in Fig. 10.

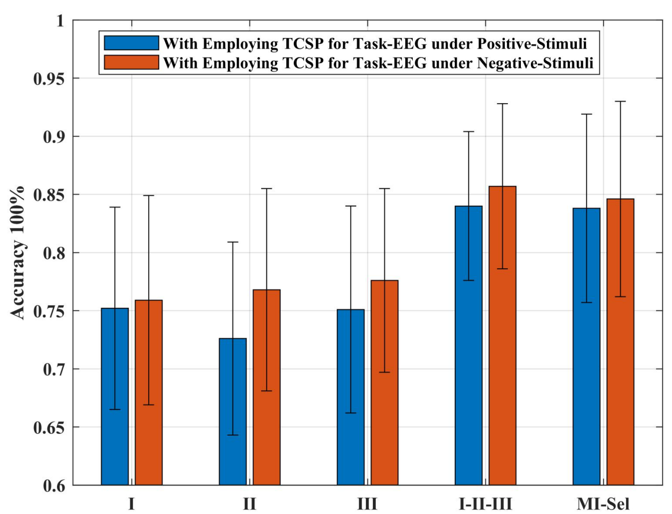

Fig. 10. TCSP classification performance bar graph for different brain regions for task-EEG under positive and negative stimuli in 6-Bands.

The results yield two primary findings. First, the ACCs in the three original regions, I, II, and III, are lower than those in the I-II-III and MI-Sel regions, which indicates that adequate original spatial information is necessary for classification performance. Second, the ACC in the I-II-III region is higher than that in the MI-Sel region, which indicates that in our study more original spatial information can enhance the performance of TCSP effectively to yield high classification performance.

C. Contribution of Gamma Waves

From Fig. 3, it can be observed that the gamma frequency band yielded the highest ACC among the individual frequency bands. Our previous study revealed abnormal functional connectivity of the EEG gamma band in patients with depression during emotional face processing [33]. Malik *et al.* observed that gamma brainwave activity will assist with the diagnosis of depression more than other individual frequency bands [4]. According to neuroscientists, a low level of gamma brainwaves is linked to depression and depressed people are typically considered to be less focused than healthy people. Therefore, gamma brainwaves may predominantly contribute to the classification performance.

D. Stimuli vs. Baseline and Positive vs. Negative

To evaluate the TCSP performance in different datasets, the EEG signals were segmented into two types of EEG signal: baseline- and task-EEG signals. There is a similar classification performance trend between baseline-EEG and task-EEG. When presented task stimuli, participants will respond to the stimulus with psychological activities [49]. Moreover, the classification ACC during the task-EEG is significantly higher statistically than that during the baseline-EEG. Unfortunately, for this experiment, the resting EEG was not recorded. This measurement may have provided a more significant result.

To observe the TCSP performance for different emotion-induction experiments, we provided two task stimuli of positive and negative emotions. Additionally, there was a similar classification performance trend between the positive and negative stimuli [50][51]. As is well known, negative bias is typical for people with depression and they can be highly sensitive to negative events. In our results, the ACC during the negative stimuli was slightly higher than that during the positive stimuli; however, there was no statistically significant difference, which may be due to the limited number of samples. Some studies have been focused on the classification of neutral stimuli, which is a potential extension of our research.

E. Other Cases using KNN and LR

We also evaluated different classifiers, including KNN and LR classifiers. The KNN classification has at least two key points: the similarity measurement (such as Euclidean distance and Mahalanobis distance) between two datasets and the selection of the k value. In this study, we chose the Euclidean distance and $k = 5$. LR is one of the most widely used statistical models. It primarily refers to a logistic function, which is a common "S" shape (sigmoid curve), loss function, regularization, and probability distribution. L2 regularization and $\lambda = 1$ were selected in our study. The results showed a trend of improvement similar to that obtained by employing TCSP.

In comparison to traditional machine learning, deep learning has achieved great success in many fields, particularly in computer vision. Recently, many researchers have obtained

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2021.3059429, IEEE Transactions on Neural Systems and Rehabilitation Engineering

high classification performance in BCIs using convolutional neural networks, which demonstrate an advantage over traditional machine learning [52]. However, an increasing number of researchers have utilized deep learning owing to its interpretability because it is always presented in the form of a black box. If we effectively combine the brainwave mechanism and deep learning, we will achieve more interesting results.

F. LOSO and K-Fold

LOSO CV is a special case of k-fold CV in which k = 1. We also separated the two categories of participants into 4 or 5 groups, Dep 12-4 and HC 11-3 for training and testing, and 4- or 5-fold CV was performed. We achieved classification results of 82.3% and 83.9% using 4- or 5-fold validation, which are lower than the values of 84% and 85.7% obtained using LOSO. However, in the cases in which 4 or 5-fold CV was used, higher classification values occurred occasionally. If we chose some optimal datasets by using unsupervised learning or other optimal algorithms, the classification may result in greater improvement. We will focus on transfer learning and deep learning in the future.

V. CONCLUSION

As a mood disease, depression is affecting an increasing number of people. As a face-in-the-crowd task stimulus experiment based on frequency information filtering, time information feature extraction, and spatial information feature selection, we developed an improved EEG-based feature classification method employing spatial information, which is useful for the detection of patients with depression. By employing the TCSP, the classification performance was significantly improved, which indicates that the TCSP can enhance the spatial differences before feature extraction; however, we should be aware of the limitation of the datasets. In the future, we will continue to focus on correlation studies to obtain more detailed results.

REFERENCES

[1] S. Avenevoli, J. Swendsen, J. He, et al., Major Depression in the National Comorbidity Survey-Adolescent Supplement: Prevalence, Correlates, and Treatment, Journal of the American Academy of Child & Adolescent Psychiatry, vol. 1(54), 37-44 (2015).

[2] M. Bachmann, L. Paeske, K. Kalev, et al., Methods for classifying depression in single channel EEG employing linear and nonlinear signal analysis, Computer Methods and Programs in Biomedicine, vol. 155, 11-17 (2016).

[3] I. Spyrou, C. Fantzids, C. Bratsas, et al., Geriatric depression symptoms coexisting with cognitive decline: A comparison of classification methodologies, Biomedical Signal Processing and Control, vol.25, 118-129 (2016).

[4] J. Malik, M. Dahiya, N. Kumari. Brain Wave Frequency Measurement in Gamma Wave Range for Accurate and Early Detection of Depression, International Journal of Advance Research and Innovation, vol. 6(1), 21-24 (2018).

[5] Y. Liu, H. Zhang, M. Chen, et al., A Boosting-Based Spatial-Spectral Model for Stroke Patients' EEG Analysis in Rehabilitation Training. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol.24(1),169-179 (2016).

[6] A. Ozcan and S. Erturk, Seizure Prediction in Scalp EEG Using 3D Convolutional Neural Networks with an Image-Based Approach, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27(11), 2284-2293 (2019).

[7] U. Acharya, V. Sudarshan, H. Adeli, et al., Computer-Aided Diagnosis of Depression Employing EEG Signals, European Neurology, vol.73, 329-336 (2015).

[8] B. Hosseinifard, M. Moradi, R. Rostami, Classifying depression patients and normal subjects employing machine learning techniques and nonlinear features from EEG signal, Computer Methods and Programs in Biomedicine, vol. 109(3), 339-345 (2013).

[9] X. Li, B. Hu, S. Sun, EEG-based mild depressive detection employing feature selection methods and classifiers, Computer Methods and Programs in Biomedicine, vol. 136, 151-161 (2016).

[10] T. Erguzel, S. Ozekes, O. Tan, et al., Feature Selection and Classification of Electroencephalographic Signals: An Artificial Neural Network and Genetic Algorithm Based Approach, Clinical EEG and Neuroscience, vol. 46(4), 321-326 (2015).

[11] S. Liao, C. Wu, H. Huang, et al., Major depression detection from EEG signals employing kernel eigen-filterbank common spatial patterns, Sensors, vol. 17(6), 1385-1404 (2017).

[12] X. Li, B. Hu, J. Shen, et al., Mild Depression Detection of College Students: an EEG-Based Solution with Free Viewing Tasks, Journal of Medical Systems, vol. 39,187 (2015).

[13] J. Yang, J. Niu, S, Zeng, et al., Resting state EEG based depression recognition research employing voting strategy method, 2018 IEEE International Conference on Bioinformatics and Biomedicine, pp. 2666-2673. IEEE, Spain (2018).

[14] C. Wu, D, Dillon, H. Hsu, et al., Depression Detection Employing Relative EEG Power Induced by Emotionally Positive Images and a Conformal Kernel Support Vector Machine, Applied Sciences, vol. 8, 1244 (2018)

[15] X. Li, R. La, Y. Wanng, et al., EEG-based mild depression recognition employing convolutional neural network, Medical & Biological Engineering & Computing, vol. 57, 1341-1352 (2019).

[16] B. Guntekin, E. Basar, Review of evoked and event-related delta responses in the human brain, International Journal of Psychophysiology, vol. 43-52 (2016).

[17] A. Wolff, S. Salle, A. Sorgini, et al., Atypical Temporal Dynamics of Resting State Shapes Stimulus-Evoked Activity in Depression-An EEG Study on Rest-Stimulus Interaction, Frontiers in Psychiatry, vol.10, 719-793 (2019).

[18] Y. Li, C. Kang, Z. Wei, et al., Beta oscillations in major depression-signaling a new cortical circuit for central executive function, Scientific Reports, vol. 7, 21-36 (2017).

[19] X. Li, B. Hu, T. Xu, et al., A study on EEG-based brain electrical source of mild depressed subjects, Computer Methods and Programs in Biomedicine, vol. 120(3), 135-141 (2015).

[20] M. Arslan, S. Eraldemir, E. Yıldırım, Subject-Dependent and Subject-Independent Classification of Mental Arithmetic and Silent Reading Tasks, International Journal of Engineering Research and Development, vol. 9 (3), 186-195 (2017).

[21] Y. Li, B Li, X. Zheng, et al., EEG-Based Mild Depressive Detection Employing Differential Evolution, IEEE Access, vol. 7, 7814-7822 (2018).

[22] S. Park, D. Lee and S. Lee, Filter Bank Regularized Common Spatial Pattern Ensemble for Small Sample Motor Imagery Classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26(2), 498-505 (2018).

[23] L. Ko, O. Komarov, S. Lin. Enhancing the Hybrid BCI Performance with the Common Frequency Pattern in Dual-Channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27(7), 1360-1369 (2019).

[24] Y. Liu, C. Wu, W. Cheng, et al., Emotion Recognition from Single-Trial EEG Based on Kernel Fisher's Emotion Pattern and Imbalanced Quasiconformal Kernel Support Vector Machine, Sensors, vol. 14(8), 13361-13388 (2014).

[25] P. Bilgin, K. Agres, N. Robinson, et al., A Comparative Study of Mental States in 2D and 3D Virtual Environments Employing EEG, IEEE International Conference on Systems, Man and Cybernetics, pp. 2833-2838. IEEE, Italy (2019).

[26] S. Chu, C. Lenglet, M. Schreiner, et al., Classifying Treated vs. Untreated MDD Adolescents from Anatomical Connectivity employing Nonlinear SVM, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4780-4783. IEEE, USA (2018).

[27] B. Ay, O. Yildirim, M. Talo, et al., Automated Depression Detection Employing Deep Representation and Sequence Learning with EEG Signals. Journal of Medical Systems, vol, 43, 205 (2019).

[28] K. Qian, H. Kuromiya, Z. Zhang, et al., Teaching Machines to Know Your Depressive State: On Physical Activity in Health and Major

Depressive Disorder, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3592-3595. IEEE, Germany (2019).

[29] H. Cai, J. Han, Y. Chen, et al., A pervasive approach to EEG based depression detection, Complexity, 118-129 (2018).

[30] K. Ang, Z. Chin, H. Zhang, et al., Filter bank common spatial pattern (FBCSP) in brain-computer interface, IEEE International Joint Conference on Neural Networks, pp. 2391-2398. IEEE, China (2008).

[31] K. Ang, Z. Chin, H. Zhang, et al., Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs, Pattern Recognition, vol. 45(6), 2137-2144 (2012).

[32] Y. Tang, Y. Li, N. Wang, et al., The altered cortical connectivity during spatial search for facial expressions in major depressive disorder, Progress in Neuro-Psychopharmacology and Biological Psychiatry, vol. 35(8), 1891-1900 (2011).

[33] Y. Li, D. Cao, L. Wei, et al., Abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing, Clinical Neurophysiology, vol. 126(11), 2078-2089 (2015).

[34] H. Jia, H. Li, D. Yu, The relationship between ERP components and EEG spatial complexity in a visual Go/Nogo task, Journal of Neurophysiology, vol. 117, 275-283 (2017).

[35] D. Karamacoska, R. Barry, G. Steiner, et al., Intrinsic EEG and task-related changes in EEG affect Go/NoGo task performance, International Journal of Psychophysiology, vol. 125, 17-28 (2018).

[36] B. Berg, L. Appelbaum, K. Clark, et al., Visual search performance is predicted by both prestimulus and poststimulus electrical brain activity, Scientific Reports, vol. 10, 718-790 (2016).

[37] Y. Liu, H. Zhang, M. Chen, et al., A Boosting-Based Spatial-Spectral Model for Stroke Patients' EEG Analysis in Rehabilitation Training. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol.24(1),169-179 (2016).

[38] A. Ozcan and S. Erturk, Seizure Prediction in Scalp EEG Using 3D Convolutional Neural Networks with an Image-Based Approach, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27(11), 2284-2293 (2019).

[39] S. Goh, H. Abbass, K. Tan, et al., Spatio-Spectral Representation Learning for Electroencephalographic Gait-Pattern Classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26(9), 1858-1867(2018).

[40] X. Zhao, H. Zhang, G. Zhu, et al., A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27(10), 2164-2177 (2019).

[41] Y. Tian, H. Zhang, Y. Jiang, et al., A Fusion Feature for Enhancing the Performance of Classification in Working Memory Load With Single-Trial Detection, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27(10), 1985-1993(2019).

[42] A. Jiang, J. Shang, X. Liu, et al., Efficient CSP Algorithm with Spatio-Temporal Filtering for Motor Imagery Classification, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28(4), 1006-1016(2020).

[43] W. Zheng, B. Lu, Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks, IEEE Transactions on Autonomous Mental Development, vol. 7(3), 162-175 (2015).

[44] L. Shi, Y. Jiao, B. Lu, Differential entropy feature for EEG-based vigilance estimation, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6627-6630. IEEE, Japan (2013).

[45] G. Tavares, R. San-Martin, J. N. Ianof, et al., Improvement in the automatic classification of Alzheimer’s disease employing EEG after feature selection, 2019 IEEE International Conference on Systems, Man and Cybernetics, pp. 1264-1269. IEEE, Italy (2019)

[46] S. Zhang, X. Li, M. Zeng, et al., Efficient KNN Classification with Different Numbers of Nearest Neighbors, IEEE Transactions on Neural Networks and Learning System, vol. 5(29), 1774-1785 (2018).

[47] J. Vergara, P. Estevez, A review of feature selection methods based on mutual information, Neural Computing & Applications, vol. 24 (1), 175-186 (2014).

[48] B. Remeseiro, V. Bolon-Canedo, A review of feature selection methods in medical application, Computers in Biology and Medicine, vol. 112, 375-383 (2019).

[49] C. Benwell, C. Tagliabue. D. Veniero, et al., Prestimulus EEG Power Predicts Conscious Awareness but Not Objective Visual Performance, eNeuro, vol. 4(6), 182-198 (2017).

[50] M. Soroush, K. Maghooli, S. Setarehdan, et al., A Review on EEG Signals Based Emotion Recognition, International Clinical Neuroscience Journal, vol. 4(4), 118-129 (2017).

[51] W. Wu, Y. Zhang, J. Jiang, et al., An Electroencephalographic Signature Predicts Antidepressant Response in Major Depression, Nature Biotechnology, vol.38, 439–447 (2020).

[52] V. Lawhern, A. Solon, N. Waytowich, et al., EEGNet: A Compact Convolutional Neural Network for EEG-based Brain-Computer Interfaces, Journal of Neural Engineering, vol. 15(5), 1-30 (2018).