Received 19 March 2023, accepted 12 April 2023, date of publication 25 April 2023, date of current version 10 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3270177

RESEARCH ARTICLE

Emotion Recognition Using Narrowband Spatial Features of Electroencephalography

IFFAT FARHANA[1], JUNGPIL SHIN[2], (Senior Member, IEEE), SHABBIR MAHMOOD[3], MD. RABIUL ISLAM[4], (Member, IEEE), AND MD. KHADEMUL ISLAM MOLLA[1], (Member, IEEE)

[1]Department of Computer Science and Engineering, University of Rajshahi, Rajshahi 6205, Bangladesh
[2]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Fukushima 965-8580, Japan
[3]Department of Computer Science and Engineering, Bangamata Sheikh Fojilatunnesa Mujib Science and Technology University, Jamalpur 2012, Bangladesh
[4]Department of Computer Science and Engineering, Varendra University, Rajshahi 6204, Bangladesh

Corresponding authors: Md. Khademul Islam Molla (khademul.cse@ru.ac.bd) and Jungpil Shin (jpshin@u-aizu.ac.jp)

This work was supported in part by the Information and Communication Technology Division (ICT Division), Ministry of Posts, Telecommunications and Information Technology, Bangladesh, through the ICT Fellowship; and in part by the Competitive Research Fund of The University of Aizu, Japan.

ABSTRACT Automatic recognition of human emotion has become an interesting topic among brain-computer interface (BCI) researchers. Emotion is one of the most fundamental features of a human subject. With proper analysis of emotion, the inner state of a human subject can be assessed directly. The human brain response can be competently represented by electroencephalography (EEG). The selection of potential features in EEG related to human emotion is a very important task for developing an effective emotion recognition system. In this paper, the discriminative features computed from rhythmic components of EEG are used to recognize human emotional states. The narrowband rhythmic components theta, alpha, beta, and gamma are extracted from multichannel EEG signals using filter bank implementation. The short-time entropy and energy features are extracted from each of the rhythmic components. The spatial filtering has been performed on the entropy-energy space by using common spatial pattern (CSP). Thus obtained spatial features are employed to recognize the emotion states using support vector machine (SVM) classifier. The publicly available two datasets DEAP and SEED are used to evaluate the performance of the proposed method. The experimental results reflect that higher recognition accuracy is obtained by using higher frequency subbands (beta and gamma) than that of the lower frequency subbands (theta and alpha). The combination of features from all subbands has better performance than the features obtained from individual subband signals. The performance of the proposed method outperforms the recently developed algorithms of emotion recognition.

INDEX TERMS Brain-computer interface (BCI), common spatial pattern (CSP), electroencephalography (EEG), emotion recognition, subband decomposition.

I. INTRODUCTION

Brain-computer interface (BCI) provides an alternative communication pathway between a brain and an external device. The brain-generated command is interpreted by the devices [1]. BCI has great significance in the biomedical engineering research fields because it aims to help physically disabled people like paralyzed patients [2]. Human emotion is effective for understanding the mental condition and it can be studied with BCI [3]. Real-time emotional assessment can make our life easier, enhance the interaction between physically disabled people and machines. It can be used for mental patients treatment.

Emotion is a complex psychological and physiological process that is related to the subject's moods, feelings, thoughts, and behaviors. People express their inner emotions verbally (emotional vocabulary) or non-verbally (intonation of voice, gesture, or facial expression). Its duration is short.

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

Researchers from a neuropsychological background have found a great relationship between emotions and the EEG signals. Mainly two areas of the human brain are associated with an emotional response. These are the pre-frontal cortex that covers part of the frontal lobe and the frontal portion of the temporal lobe (amygdala and hippocampus) [4]. Amygdala and hippocampus perform their functions independently. But when a human subject delivers emotional reactions to a stimulus, these two regions of the brain interact with each other for translating the emotion into a specific reaction. Amygdala senses emotions and processes them. Hippocampus is responsible for regulating episodic memory [5]. Emotion can be represented theoretically in two ways: categorical and dimensional emotion models. In the categorical model, emotions are tagged separately. Ekman and Friesen proposed six basic, universal, and distinct emotions [6]. These are fear, joy, anger, sadness, disgust, and surprise. In the dimensional model, emotions are mapped into the valence, arousal, and dominance dimension. Valence goes from negative to positive (unpleasant to pleasant), arousal goes from boring to excited (deactivation to activation), and dominance goes from submissive (being controlled) to a powerful (in control) feeling. Plutchik's emotion wheel [7] and Russell's circumplex model [8] are the representations of the dimensional model. Russell's scale of valence-arousal is mostly used by researchers and four categories of emotions are obtained from this model. These are high valence high arousal (HVHA), low valence low arousal (LVLA), high valence low arousal (HVLA), and low valence high arousal (LVHA). When the dataset for emotion recognition is prepared using a dimensional model, self-assessment manikin (SAM) is used for graphical representation [9]. SAM uses images for presenting valence, arousal, and dominance. It uses a rating scale ranging from 1 to 9 for each dimension. Subjects of any age can understand this figure and their level of emotion easily.

Emotion is recognized from non-physiological signals like speech, gesture, and facial expressions; physiological signals like electromyography (EMG), electroencephalography (EEG), galvanic skin response (GSR), electrocardiography (ECG), respiration rate (RR), and blood volume pressure [10]. Emotion detected from non-physiological signals may be fake [11]. That's why physiological signals particularly, EEG represents the brain's electrical activity and provide accurate information about the current emotional state. EEG is non-invasive, easy to use, inexpensive, fast, and portable. It consists of rhythmic components delta, theta, alpha, beta, and gamma. It has high time resolution but low spatial resolution [1]. EEG signal pattern varies from person to person. The model generated for a particular EEG dataset may not perform well while testing with another dataset. Most of the methods for emotion recognition have been developed using the standard dataset [4]. MAHNOB-HCI is a publicly available multimodal dataset for emotion analysis that includes both facial expressions and physiological signals. In recent years, the database for emotion analysis using physiological signals (DEAP) is mostly used [12] that contains both EEG and peripheral physiological signals. The Shanghai Jiao Tong University (SJTU) emotion EEG dataset (SEED) is also used for emotion analysis that contains only the EEG signals [13]. Finding appropriate features of EEG related to human emotion is important for better classification performance. Researchers have proposed time-domain, frequency-domain, and joint time-frequency domain techniques for this purpose. The joint time-frequency domain technique is the most effective for analyzing non-stationary signals like EEG [14].

Researchers have proposed several methods using machine learning techniques for classification of human emotion. Rhythmic component based features of EEG play important role to emotion classification [12]. For extracting rhythmic components of EEG discrete wavelet transform (DWT) [15], empirical mode decomposition (EMD) [11], [16], and multivariate empirical mode decomposition (MEMD) [17] are used. Besides, flexible analytic wavelet transform (FAWT) with information potential (IP) feature [18] and multivariate synchrosqueezing transform (MSST)-based methods [19] have been also discussed for emotion recognition. Higher frequency bands like gamma perform better for emotion classification and signal framing is also important for obtaining better performance [15], [16]. Cross-subject emotion recognition method was proposed with a combination of different methods [20], [21], [22]. These methods include transfer recursive feature elimination (TRFE) method [20], finite impulse response (FIR) bandpass filter [21], significance test/sequential backward selection and the support vector machine (ST-SBSSVM) using multiple features [22]. But there is difficulty in recognizing cross-subject emotions using EEG because of its poor generalization of features.

Several deep learning-based systems were proposed by the researchers for emotion recognition from EEG signals. Both subject-independent and dependent approaches were discussed in [23] using convolutional neural network (CNN). Capsule network (CapsNet) with multiband feature matrix (MFM) [24] for emotion recognition approach has been discussed. Another method based on stack autoencoder (SAE) with long short-term memory recurrent neural network (LSTM-RNN) was described in [25]. Asghar et al. [26] proposed multi-modal emotion recognition approach using AlexNet for extracting time and frequency domain features and bag of deep features (BoDF) for feature reduction. Besides, attention-based convolutional recurrent neural network (ACRNN) [27], channel-fused dense convolutional network (CDCN) [28], fusion model of long-short term memories neural networks (LSTM) and graph convolutional neural network (GCNN) named ECLGCNN [29], cascaded and parallel hybrid convolution recurrent neural networks [30], dynamical graph convolutional neural networks (DGCNN) [31], spatial-temporal recurrent neural network (STRNN) [32], deep convolutional neural network [33], four-dimensional convolutional recurrent neural network (4D-CRNN) [34], graph convolutional broad network (GCB-net) along with broad learning system (BLS) [35], regularized graph neural network (RGNN) [36], combination of convolutional neural network (CNN) and deep neural network (DNN) [37], spiking neural networks (SNNs) [38], optimized residual networks (ResNet) [39], combined deep neural network (DNN) models [40], [41], CNN-BiLSTM-MHSA model [42] consisting of a convolutional neural network (CNN), bi-directional long and short-term memory network (BiLSTM), and multi-head self-attention (MHSA) were also proposed by the researchers for emotion recognition from EEG signals. Identifying EEG features that are effective for proper emotion recognition is very important. In the previous works, the extracted features are not well discriminative for human emotion classification. That is why the emotion classification performance is not up to the mark for real-world application development.

This study has focused on the derivation of potential features closely related to represent human emotion. After preprocessing of raw EEG data, bandpass filtering approach is implemented on each channel of the trials to obtain the rhythmic components. Each component is a narrowband signal. Each channel of the trials of a rhythmic component is segmented into short-term frames. The entropy and energy of each frame are calculated. The channels are expressed as a sequence of short-term entropy and energy obtained from individual frames. Thus the trials are characterized by short time attribute (entropy-energy terms) instead of EEG samples. The CSP is then applied to the newly defined trials to generate spatial features. Further, the spatial features derived from an individual rhythmic component are used for emotion recognition using support vector machine (SVM). The proposed CSP-based method generates highly discriminative features that enhance the emotion recognition performance. The experimental results are compared with the existing methods.

The rest of this paper is organized as follows: Section II describes the datasets used in this experiment and data preprocessing, Section III describes the detailed methodology, Section IV illustrates the experimental results, Section V presents a general discussion of the proposed method, its feasibility for real-time BCI application development in comparison with the other recently developed algorithms, and finally Section VI presents the conclusions of this study.

II. DATASET DESCRIPTION AND DATA PREPROCESSING

A. DATASET DESCRIPTION

Two publicly available datasets termed as DEAP dataset [12] and SEED dataset [13] have been used to evaluate the performance of the proposed method. These datasets are considered as the benchmark datasets among the researchers for EEG based emotion recognition. A short description of these datasets is given below.

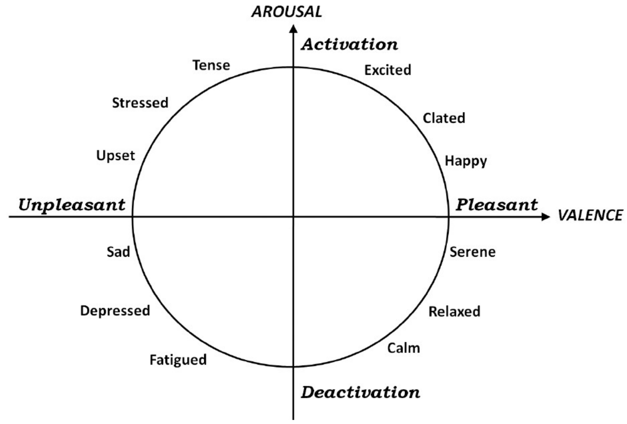

FIGURE 1. Russell's 2D valence-arousal scale. Valence and arousal are presented by the horizontal and vertical axis respectively. Valence ranges from unpleasant to pleasant and arousal ranges from deactivation to activation.

1) DEAP DATASET

The database for emotion analysis using physiological signals (DEAP) is a publicly accessible multimodal dataset for emotion analysis [12]. This dataset was recorded in collaboration with Queen Mary University of London, University of Twente, and Universitè De Genève. In this dataset, 40 music videos were used as stimuli for eliciting emotion from human subjects and length of each video was 1 minute. A grading scale ranging from 1 to 9 was associated with each music video. The total number of subjects is 32 (50% female) and each of them was shown 40 music videos in 40 trials. The average age of the participants was 26.9 years. All of the participants were requested to read about the details of the experiment and the rules of self-assessment. They signed a consent form before starting the signal recording. This dataset is based on the valence, arousal, dominance, and liking (level of preference) emotional model. It follows Russell's scale (Fig. 1).

The electroencephalogram (EEG) and peripheral physiological signals of each participant were recorded. After watching each music video all the participants were asked to give ratings based on their self-assessment according to the level of valence, arousal, dominance, and liking. They could click on any value within the range of 1 to 9. Another rating was given based on familiarity and its range was 1 to 5. The original dataset was recorded with 512 Hz sampling rate and 48 channels. Then the dataset was downsampled to 128 Hz to prepare the preprocessed dataset with 40 channels. Among 40 channels, 32 channels are used for recording EEG signals and the remaining 8 channels are for recording physiological signals (galvanic skin response, blood volume pressure, respiration rate, skin temperature, electromyography, and electrooculography). Electrodes were placed according to the international standard 10-20 system. The preprocessed dataset was bandpass filtered to the frequency range of 4 Hz to 45 Hz. The electrooculography (EOG) artifacts were removed and the values of the data were averaged to

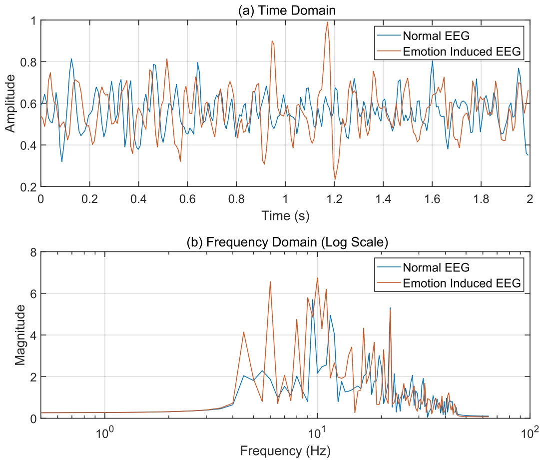

FIGURE 2. Comparison of normal EEG and emotion-induced EEG for the DEAP dataset where the subplot (a) and subplot (b) represent the time domain and frequency domain respectively.

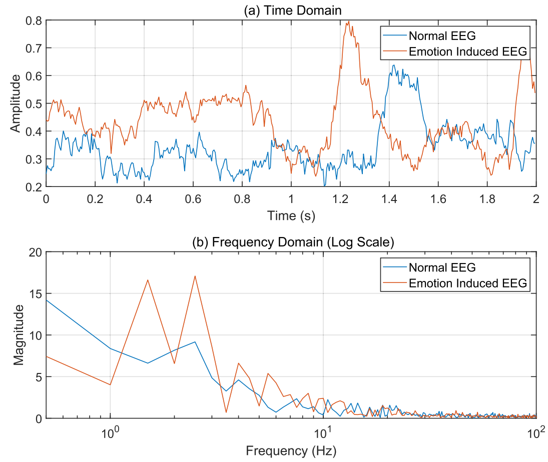

FIGURE 3. Comparison of normal EEG and emotion-induced EEG for the SEED dataset where the subplot (a) and subplot (b) represent the time domain and frequency domain respectively.

TABLE 1. DEAP dataset description.

Name	Dimension	Content name
data	40 × 40 × 8064	Video/trials × channels × data points
labels	40 × 4	Video/trials × label (valence, arousal, dominance, liking)

TABLE 2. SEED dataset description.

No. of subjects	15
No. of channels	62
No. of sessions	3
No. of trials/session	15
Length of video stimuli	4 minutes
Labels	+1, 0, -1 (Positive, Neutral, Negative)

the common reference. The preprocessed dataset is used in this experiment. The description of the DEAP dataset is given in Table 1. Among 40 trials, some trials have high valence and some trials have low valence. This categorization is the same for the arousal dimension. For binary classification, a threshold is used to categorize the labels into two classes for a grading scale of 1 to 9. If the grading scale is grade $>= 4.5$, then the label of valence/arousal is high and if it is grade < 4.5, the label of valence/arousal is considered low. The time domain and frequency domain representation of normal EEG and emotion-induced EEG has been illustrated in Fig. 2 for the DEAP dataset.

2) SEED DATASET

The Shanghai Jiao Tong University (SJTU) emotion EEG dataset (SEED) is a freely available emotion recognition dataset [13]. While recording this dataset, 15 Chinese film clips were used as stimuli for eliciting 3 types of emotions (positive, neutral, and negative). The duration of each video clip was 4 minutes long. Fifteen subjects (7 males and 8 females) participated in the EEG data recording and their average age was 23.27 years. All of the participants were students from Shanghai Jiao Tong University. Initially, this dataset was recorded using ESI NeuroScan System with 1000 Hz sampling rate from a 62-channel electrode cap according to the international standard 10-20 system. Then the dataset was downsampled to 200 Hz. There is a 5 s hint before each clip, 45 s for self-assessment, and 15 s for rest after each clip in one session. The order of presentation is arranged so that two film clips targeting the same emotion are not shown consecutively. Each subject performed the experiment three times with an interval of about one week. Each experiment includes 15 trials. For each subject, the total number of trials is $3 \times 15 = 45$. Among the 45 trials, 15 trials represent positive emotion experiments, 15 trials represent neutral emotion experiments, and 15 trials represent negative emotion experiments. In this experiment, the preprocessed dataset is used. This dataset is bandpass filtered from 0 Hz to 75 Hz. Positive, neutral, and negative emotions are labeled by +1, 0, and -1. For this experiment, only positive and negative emotions have been used for binary classification. The details of the SEED dataset are given in Table 2. The time domain and frequency domain representation of normal EEG and emotion-induced EEG has been illustrated in Fig. 3 for the SEED dataset.

B. DATA PREPROCESSING

1) SELECTION OF CHANNELS AND TRIAL LENGTH

There are 40 channels (32-EEG and 8-physiological) used in the recording of the DEAP dataset. The first 32 EEG channels are used in this study [12]. The 60 s of trial length produces $60 \times 128 = 7680$ data points for each trial and hence the data dimension becomes $32 \times 7680 \times 40$ (channels $\times$ data points $\times$ trials). Sixty-two (62) EEG channels are used in the SEED dataset. To maintain consistency with the DEAP dataset, the EEG signals of 60 s are extracted from the middle of 4 minutes long trial. Among the three types of emotions, only positive and negative emotions are considered for performing binary classification.

The dimension of the SEED dataset used in this study is $62 \times 7680 \times 30$ (channels $\times$ data points $\times$ trials).

2) DATA NORMALIZATION

Normalization is performed on samples to reduce individual differences and computational complexity. The average mean reference (AMR) method [15] is applied for data normalization to remove the noises due to all external interference. The mean value of all the data points is calculated for each channel of a trial. The obtained mean value is subtracted from the data points of i-th channel and j-th trial using the formula (1).

$$D_{i,j}(m) = O_{i,j}(m) - \frac{1}{N}\sum_{m=1}^{N} O_{i,j}(m) \tag{1}$$

Here, O and D represent the original EEG data and the new EEG data after applying the AMR method respectively, m is the index of the number of data points where $1 \leq m \leq N$. $N = 7680$ for the DEAP dataset and $N = 12000$ for the SEED dataset. The ranges of i and j are $1 \leq i \leq L$ (number of channels) and $1 \leq j \leq T$ (number of trials) respectively.

Then the min-max normalization is performed on the resultant data using equation (2). As a result, the values of all the data points of i-th channel and j-th trial will have in the range [0, 1].

$$E_{i,j}(m) = \frac{D_{i,j}(m) - \min\left(D_{i,j}\right)}{\max\left(D_{i,j}\right) - \min\left(D_{i,j}\right)} \tag{2}$$

Here, E represents the normalized EEG data.

III. METHODOLOGY

A common spatial pattern (CSP)-based approach for emotion recognition from EEG signals is implemented in this study to enhance the classification accuracy of human emotion recognition. A block diagram of the proposed method is demonstrated in Fig. 4. In the block diagram of Fig. 4, the basic theme of the proposed method has been represented for the individual subband. According to the representation of Fig. 4, the performance analysis of the proposed method has been conducted on the individual subband. Besides, the proposed method's performance analysis has also been conducted on combined CSP-generated features for all subbands from which better performance has been obtained. Since Fig. 4 represents the basic theme of the proposed method, the process of combined features from all subbands has not been shown in the block diagram. To clarify the block diagram of Fig. 4, the algorithmic steps of the proposed method have been described below.

Step 1: Perform preprocessing on EEG trials.
Step 2: Perform subband decomposition on the preprocessed EEG signals.
Step 3: Representing each channel of subband Sb_b with short-time entropy (η) and energy (ε) features of EEG.
Step 4: Extracting CSP-based discriminative features for individual subbands.
Step 5: Emotion classification by SVM classifier.

The channel representation technique mentioned in **Step 3** with short time entropy and energy features is further described into five sub-steps in the following.

Step 3.1: Extracting subband Sb_b from all channels of a single trial.
Step 3.2: Decomposing each channel into multiple frames.
Step 3.3: Extracting entropy (η) and energy (ε) features from each frame.
Step 3.4: Horizontal concatenation of entropy (η) and energy (ε) features for individual frame.
Step 3.5: Finally, the concatenated entropy (η)-energy (ε) features for all frames and all trials are again concatenated horizontally to represent a single channel. The same procedure has been conducted for the representation of each channel.

The proposed method has been developed by implementing the following major steps. They are, (A) subband decomposition, (B) signal segmentation, (C) short-term characteristics, (D) features filtering, and (E) classification. The details of these steps are described in the following subsections.

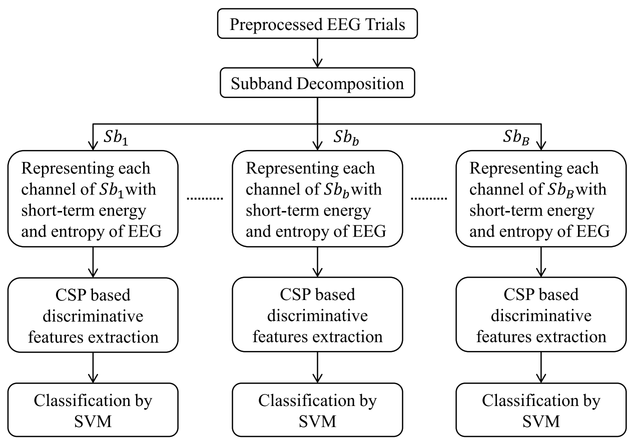

FIGURE 4. Block diagram of the proposed common spatial pattern (CSP)-based emotion recognition system, where $\mathrm{Sb}_1, \ldots \mathrm{Sb}_b, \ldots \mathrm{Sb}_B$ represent the different subbands.

A. SUBBAND DECOMPOSITION

Subband decomposition represents a wide band signal into a finite set of narrow band signals. The rhythmic components of EEG signals are used in the proposed method. The bandpass filters are used to extract rhythmic components from the EEG signal [43], [44]. In motor imagery-based BCI, filter bank in combination with common spatial pattern (CSP) is used for improving classification accuracy [44]. The preprocessed DEAP dataset has a frequency ranging from 4 Hz to 45 Hz. This dataset demonstrates that the rhythmic components theta, alpha, beta, and gamma are effective for emotion analysis [12]. For obtaining these rhythmic components, bandpass filtering is implemented. A 5^{th} order Butterworth bandpass filter has been applied to the preprocessed EEG data for rhythmic component extraction. For the SEED dataset, rhythmic components have been extracted in a similar way.

B. SIGNAL SEGMENTATION

Emotion doesn't stay for a long time, it changes rapidly. Its duration is generally within the range of 0.5 s to 4 s [15]. From the previous studies, it has been found that the researchers used a sliding window of 1 s, 2 s, 4 s, 5 s, and so on [25]. The nature of EEG is non-stationary and for this reason, the same model fails to perform well for all subjects. Signal framing with appropriate frame length is important during EEG signal processing for emotion recognition [25]. A sliding window of a few seconds is used for segmenting the EEG signal into multiple numbers of frames. In this experiment, signal framing has been performed on the preprocessed EEG data of each subband signal. For the DEAP dataset, 4 s and 2 s signals have been extracted from 60 s signal and 50% overlap has been used for both of them. The 60 s signal has been segmented into 29 frames and 59 frames for 4 s and 2 s frame lengths respectively. The same procedure has been followed for processing the SEED dataset.

C. SHORT-TERM CHARACTERISTICS

The short-term entropy and energy [15] have been calculated as features from each frame for each subband (theta, alpha, beta, gamma).

1) ENTROPY (η)

Entropy (η) is used to analyze EEG data. It can be computed by the following formula (3).

$$\eta_{i,j}^{b} = -\sum_{g=1}^{V}\left[\left\{\mathrm{Sb}_b^h(g)^2\right\}\log\left\{\mathrm{Sb}_b^h(g)^2\right\}\right] \tag{3}$$

2) ENERGY (ε)

Energy (ε) is computed by the following formula (4).

$$\varepsilon_{i,j}^{b} = \sum_{g=1}^{V}\left\{\mathrm{Sb}_b^h(g)^2\right\} \tag{4}$$

Here, $\eta_{i,j}^{b}$ and $\varepsilon_{i,j}^{b}$ represents the entropy (η) and energy (ε) features of each frame for i-th channel and j-th trial of a specific subband b respectively. Sb_b^h represents the h-th frame of the specific rhythmic or subband component for frequency band b. Here, b = theta or alpha or beta or gamma. In Fig. 4, B represents the total number of subbands where $B = 4$. h represents the index of the number of frames where $1 \leq h \leq F$. g is the index of the number of samples in each frame where $1 \leq g \leq V$. For DEAP dataset, $F = 29$ and 59, $V = 512$ and 256 (for 4 s and 2 s sliding windows respectively) and for SEED dataset, $F = 29$, $V = 800$ (for 4 s sliding window). Finally, the features obtained from the individual frame are horizontally concatenated for every trial and for each subband to represent a channel. Spatial filtering has been applied to this feature matrix for calculating spatial features.

D. FEATURES FILTERING

Spatial filtering has been applied on short-time attribute data using the common spatial pattern (CSP) to generate discriminative spatial features. The features are extracted from each frequency band. The dataset is first divided into the training subset and test subset. The CSP algorithm is applied to the training subset only for computing spatial filters. Then CSP features are calculated from both the training subset and the test subset by projecting the corresponding dataset onto the computed spatial filters. For the final CSP feature extraction, the log variance of the projected feature is calculated [45].

The CSP first proposed by H. Ramoser is a spatial filtering technique that has been successfully implemented in the EEG-based motor imagery classification [46]. This technique finds optimal spatial filters that are effective in discriminating EEG signals of two classes. Another potential use of CSP is to reduce the dimensionality of high dimensional EEG data. The output of the CSP is a projection matrix and its row represents the weights for channels. The basic idea is to project the multi-channel EEG data into a lower dimensional spatial subspace with the resultant projection matrix and a linear transformation. It maximizes the variance for one class while minimizing the variance for the other class yielding maximizing inter-class variance and minimizing intra-class variance. The CSP algorithm focuses on the simultaneous diagonalization of two covariance matrices. It is a supervised algorithm trained on the labeled data [45], [47], [48].

In this study, CSP is implemented for emotion classification using EEG data. It is used separately for the arousal and valence dimensions of human emotion. Both arousal and valence are divided into two classes - high and low. The details of the CSP algorithm are explained here for the arousal dimension. The same procedure is followed for the valence dimension.

The training EEG data represented by short-time attribute (entropy-energy terms) is denoted by X_q for class q. The dimension of X_q is $L \times N$, where L is the number of channels and N is the number of samples per channel [46], [47]. The process of generating X_q from initial training data has been explained in the last paragraph of this subsection in detail.

For high and low arousal, a single trial EEG data is denoted by $X_{q\in \mathrm{ha,la}}$. X_{ha} and X_{la} represent high arousal and low arousal data respectively. The normalized spatial covariance of the EEG data for high arousal, C_{ha} and low arousal, C_{la} is calculated as follows,

$$C_{\mathrm{ha}} = \frac{X_{\mathrm{ha}}X_{\mathrm{ha}}'}{\mathrm{trace}\left(X_{\mathrm{ha}}X_{\mathrm{ha}}'\right)}, \quad C_{\mathrm{la}} = \frac{X_{\mathrm{la}}X_{\mathrm{la}}'}{\mathrm{trace}\left(X_{\mathrm{la}}X_{\mathrm{la}}'\right)} \tag{5}$$

$(.)'$ is the transpose operator and trace (.) computes the sum of the diagonal elements of a given matrix. The next step is to calculate the composite spatial covariance, C. It can be computed as,

$$C = \overline{C_{\mathrm{ha}}} + \overline{C_{\mathrm{la}}} = U_0 \lambda U_0' \tag{6}$$

$\overline{C_{\mathrm{ha}}}$ and $\overline{C_{\mathrm{la}}}$ are the averaged normalized covariance matrix. $\overline{C_{\mathrm{ha}}}$ and $\overline{C_{\mathrm{la}}}$ are calculated by averaging over all the trials of each class. U_0 is the matrix of eigenvectors and λ is the diagonal matrix of eigenvalues. It is assumed that the eigenvalues are sorted in descending order.

$$W = \lambda^{\frac{-1}{2}} U_0' \tag{7}$$

Here, W is the whitening transformation matrix. $\overline{C_{\text{ha}}}$ and $\overline{C_{\text{la}}}$ are transformed by the W as

$$S_{\text{ha}} = W\overline{C_{\text{ha}}}W', \quad S_{\text{la}} = W\overline{C_{\text{la}}}W' \tag{8}$$

The transformed terms S_{ha} and S_{la} share common eigenvectors and the sum of the corresponding eigenvalues for the two matrices will always be 1.

$$S_{\text{ha}} = U\lambda_{\text{ha}}U', \quad S_{\text{la}} = U\lambda_{\text{la}}U', \ \lambda_{\text{ha}} + \lambda_{\text{la}} = I \tag{9}$$

Here, I is the identity matrix. The eigenvector that has the largest eigenvalue for S_{ha} has the smallest eigenvalue for S_{la} and vice versa because the sum of the corresponding eigenvalues is always 1. A high eigenvalue for S_{ha} means a high variance for EEG in high arousal and a low eigenvalue for S_{la} means a low variance for EEG in low arousal and vice versa. Classification is performed based on this property. The projection of whitened EEG onto the first and last eigenvectors will give feature vectors that are optimal for discriminating two classes of EEG data [46], [47]. The projection matrix is calculated by

$$P = U'W \tag{10}$$

With the projection matrix P, the original EEG data can be transformed into uncorrelated components Z according to,

$$Z = PX \tag{11}$$

The original EEG data matrix can be reconstructed by

$$X = P^{-1}Z \tag{12}$$

P^{-1} is the inverse of P. The columns of P^{-1} are the spatial patterns. These can be considered time-invariant EEG source distribution vectors. The most discriminatory spatial patterns are the first and last columns of P^{-1}. It describes the highest variance of one class and the lowest variance of the other class.

Here, the initial training EEG data is represented by short-time attribute (entropy-energy terms) which is denoted by X for a specific subband. The dimension of X is $L \times N_f \times N_{\text{tr}}$, where L is the number of channels, N_f is the number of samples ($N_f = 2$) per channel, and N_{tr} is the number of training trials. Then, X_q training EEG data has been extracted from the initial training EEG data X for a specific class q. Now, the dimension of X_q is $L \times N_f \times N_{\text{tr}}^q$. Here, N_{tr}^q represents the training trials of class q. Each channel is constructed by horizontally placing the two features (entropy and energy) in a single row for a specific class q (high/low). This has been performed for all channels L of each training trial. Then, the X_q three-dimensional (3D) matrix has been reshaped into a two-dimensional (2D) matrix by horizontally placing the features (entropy-energy) from all trials into a single row of a specific channel. These have been done for every channel L of each class q. Then CSP has been performed on the training data X_q. Here, the dimension of X_q is $L \times \left(N_f \times N_{\text{tr}}^q\right) = L \times N$. After applying CSP, L numbers of features are generated in 2D matrix format. From L numbers of features, seven pairs (number of selected features, $N_{\text{sf}} = 7$ pairs= 14) of features have been selected for training the SVM classifier. For each band, a new projected training feature matrix has been generated from X and N_{sf} numbers of CSP-generated features by considering the log variance of the projected features. The dimension of the projected feature matrix is $N_{\text{tr}} \times N_{\text{sf}}$. Finally, the SVM classifier has been trained with this data to show individual subband performance. To show the final performance of the proposed method, the feature matrix of all subbands has been combined to produce the final feature matrix of dimension $N_{\text{tr}} \times (N_{\text{sf}} \times B)$ which has been used to train SVM classifier. Here, B represents the total number of subbands.

E. CLASSIFICATION

The spatial feature vector of a single subband as well as the combination of multiple subbands is used for emotion classification using support vector machine (SVM). The SVM classifier is trained with the labeled training dataset and binary classification is performed using the test dataset. The polynomial kernel of order 3 is used with k (=5) fold cross-validation technique.

IV. EXPERIMENTAL RESULTS

The well-known publicly available datasets named DEAP and SEED are used to evaluate the performance of the proposed method for emotion recognition. The binary classification is considered here. For the DEAP dataset, classification is performed on both the valence (high/low) and arousal (high/low) dimensions. A threshold of 4.5 is used for dividing the rating scale 1 to 9 into high and low. For the SEED dataset, the positive and negative emotions are classified. The preprocessed EEG data have been decomposed into four subband signals using Butterworth bandpass filtering technique. Then framing has been applied on each subband signal for each channel and trial. The entropy and energy of each frame are calculated for each subband signal. Each subband is represented as a sequence of short-term entropy and energy. The spatial features are calculated from the resultant representation by applying CSP for each subband signal. The spatial features of multiple subbands are also combined. Then the obtained features are used for training the SVM classifier with cubic polynomial kernel function, followed by the evaluation of the classification performance with the test data. The details of this procedure are described in section III. Four different subbands (theta, alpha, beta, gamma) of a 4 s window taken from a single channel and a single trial signal of the subject 's01' are shown in Fig. 5.

A k-fold (here, $k = 5$) cross-validation technique is used to evaluate the performance of the proposed method on an

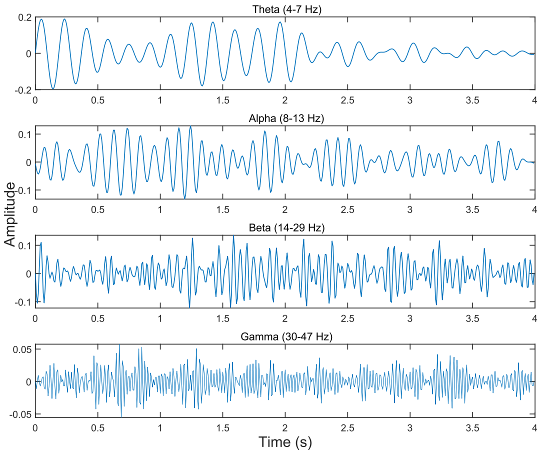

FIGURE 5. **Four different subbands (Theta 4-7 Hz, Alpha 8-13 Hz, Beta 14-29 Hz, and Gamma 30-47 Hz) of a window obtained from channel FC5 of subject 's01' selected from DEAP dataset.**

individual subject basis. The trials of each subject are divided into k equal groups. Among the k groups, (k - 1) groups are used for training and the remaining one is used for testing. This process is repeated k times and each time the data are permuted differently. The results obtained from the k repetitions are averaged to derive the classification accuracy. The classification performance is evaluated by accuracy calculated as Accuracy $= 100 \times (\frac{N_C}{N_T})$. Here, N_T is the total number of trials in the test dataset and N_C is the total number of trials correctly recognized out of N_T.

In the proposed method, each EEG trial of 60 s is divided into 29 frames using 4 s frame length with 2 s overlap for the DEAP dataset. Then 40 trials produce the total of $29 \times 40 = 1160$ frames and each of the five folds contains 232 segments. The 2 s frame length with 1 s overlap is also used. It produces 59 frames from each trial and obtained $59 \times 40 = 2360$ for 40 trials. The results obtained from the five folds are averaged to calculate the accuracy. This process is followed for each subject and the average result is computed over $N_{\text{sub}} = 32$ subjects. There are 30 trials in the SEED dataset. The 4 s and 2 s frames are also implemented for the SEED dataset and a five-fold cross-validation approach is implemented using a similar process as implemented in the DEAP dataset. The average accuracy is considered the final accuracy. The accuracy of the individual subject is calculated by using the formula (13).

$$\text{Accuracy}_{\text{sub}} = \frac{1}{k}\sum_{i=1}^{k}\text{Accuracy}(i) \tag{13}$$

The average accuracy of all subjects is calculated by using the formula (14) for both datasets (DEAP and SEED).

$$\text{Accuracy}_{\text{mean}} = \frac{1}{N_{\text{sub}}}\sum_{j=1}^{N_{\text{sub}}}\text{Accuracy}_{\text{sub}}(j) \tag{14}$$

The performances of emotion recognition in the valence and arousal dimensions with the DEAP dataset for frame length of 4 s and 2 s are presented in Fig. 6 and Fig. 7 respectively. The features of different subbands (theta, alpha, beta, gamma) for individual subjects are compared in these figures. It is observed that the higher-frequency subbands (gamma and beta) perform better than the lower-frequency subbands (alpha and theta). The performance of the combined features obtained from multiple subbands is higher than that of the single subband. The emotion recognition accuracy of the combined features extracted from all the subbands is also presented. It is observed from these figures that accuracy is gradually enhancing with the increasing number of subbands from which the obtained features are combined. The effects of frame length for emotion recognition are illustrated in Fig. 8 considering 4 s and 2 s frame length for valence and arousal dimensions. This figure shows that the 4 s frame length has higher classification accuracy in both cases. The frame length of 4 s is considered to evaluate the performances of the proposed method with the SEED dataset. The classification of positive and negative emotions results for the SEED dataset (with 4 s frame length) are presented in Fig. 9. The average accuracy over the subjects for different subbands with DEAP and SEED datasets are shown in Fig. 10 and Fig. 11 respectively. The highest accuracy is obtained for the combined features of all the subbands for both datasets. From Fig. 6 and Fig. 7, it has been found that the contribution of theta and alpha bands is lower than the others for the DEAP dataset. On the other hand, the contribution of theta and alpha bands for the SEED dataset is comparatively higher than the DEAP dataset. But it is still lower than the other subbands in the SEED dataset which is evident in Fig. 9 and Fig. 11. We can see that the performance of the SEED dataset is higher than the DEAP dataset. For the SEED dataset, the theta and alpha bands contribute more than the theta and alpha bands of the DEAP dataset. That is why the overall performance of the SEED dataset is higher than the DEAP dataset. Besides, the contribution of different subbands of EEG signals mostly depends on the mental state of the subjects as well as the signal acquisition environments. It could be that these effects have been reflected in the SEED dataset for lower bands (theta, alpha) whereas it has not been reflected in the lower bands of the DEAP dataset. In addition to, it is clearly shown that in some of the existing methods like differential entropy spatial-temporal recurrent neural network (DE-STRNN) proposed by Zhang et al. [32], the performance of the theta and alpha bands is better than the other bands for the SEED dataset.

It is observed that the spatial features extracted by CSP from the short term energy-entropy representation of EEG samples are very effective for the binary classification of emotion. The results obtained from different pairs of spatial features for the DEAP dataset are illustrated in Fig. 12 in which the average accuracy over the subjects as a function of the number of the selected spatial features are presented for the valence and arousal dimensions. The classification performance is enhancing with the increasing number of spatial features gradually. After including a certain number

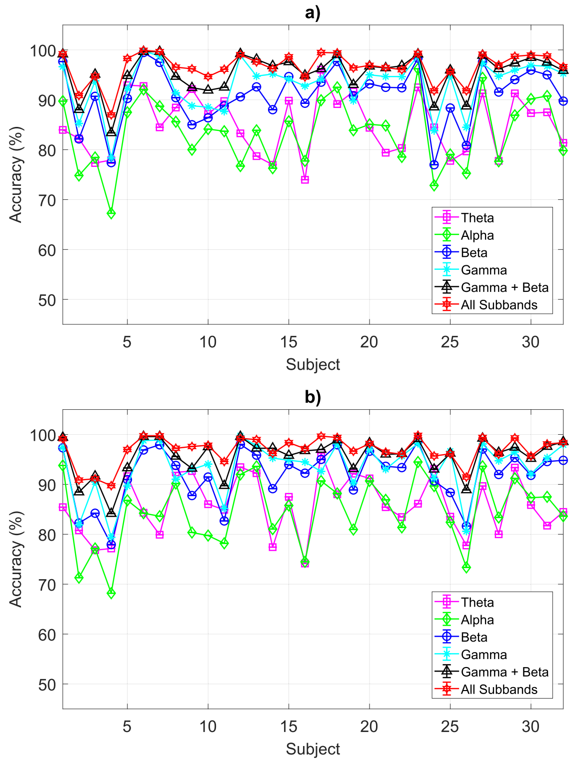

FIGURE 6. Classification performance of individual subjects of the DEAP dataset for different subband features for a) Valence dimension and b) Arousal dimension for 4 s frame length with 2 s overlap.

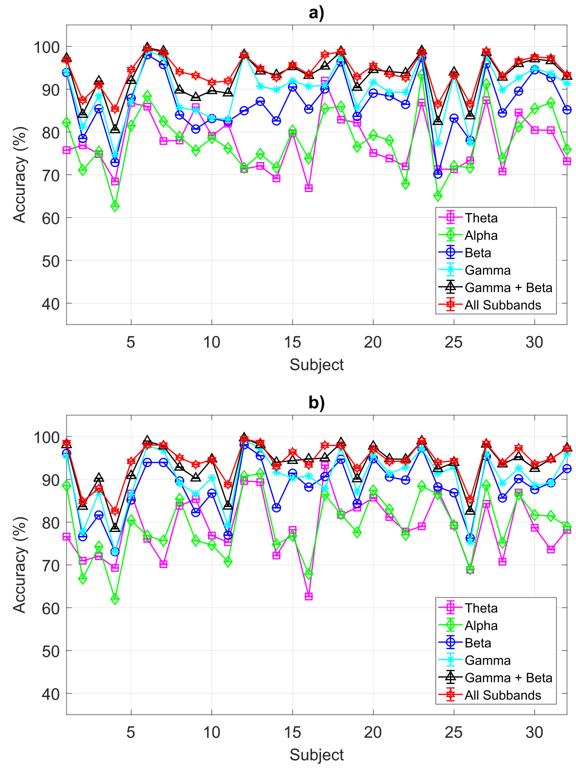

FIGURE 7. Classification performance of individual subjects of the DEAP dataset for different subband features for a) Valence dimension and b) Arousal dimension for 2 s frame length with 1 s overlap.

of spatial features, the performance is decreased and then increases again. The average accuracy becomes maximum (96.15% $\pm$ 2.98 for valence and 96.47% $\pm$ 2.90 for arousal) for seven pairs of spatial features resulting in fourteen spatial features and it falls for sixteen spatial features (93.70% $\pm$ 4.07 for valence and 94.29% $\pm$ 4.26 for arousal). After that, the performance increases and the resulting change is very small from eighteen spatial features (96.39% $\pm$ 2.85 for valence and 96.57% $\pm$ 2.98 for arousal) to thirty-two spatial features (96.62% $\pm$ 2.95 for valence and 96.90% $\pm$ 2.76 for arousal). For the SEED dataset, the average classification performance of all subjects as a function of the number of selected spatial features is presented in Fig. 13. The result is maximum (99.98% $\pm$ 0.06) for six pairs of spatial features. The accuracy for seven pairs of spatial features is 99.95%$\pm$0.10 which is very close to the maximum accuracy. Considering the consistency, the result of seven pairs has been taken for the SEED dataset.

V. DISCUSSION

Recently, emotion recognition from EEG signals has gained much attention among researchers in the BCI field. The reason is the use of emotion recognition result in medical application. From the previous studies, it is found that recognition accuracy is not enough for real-world application development. The researchers are still working to develop an efficient method to improve classification accuracy. In existing methods, the time domain, frequency domain, and joint time-frequency domain analysis were employed. The EEG signals are decomposed to obtain different rhythmic components. Several techniques such as empirical mode decomposition (EMD), multivariate empirical mode decomposition

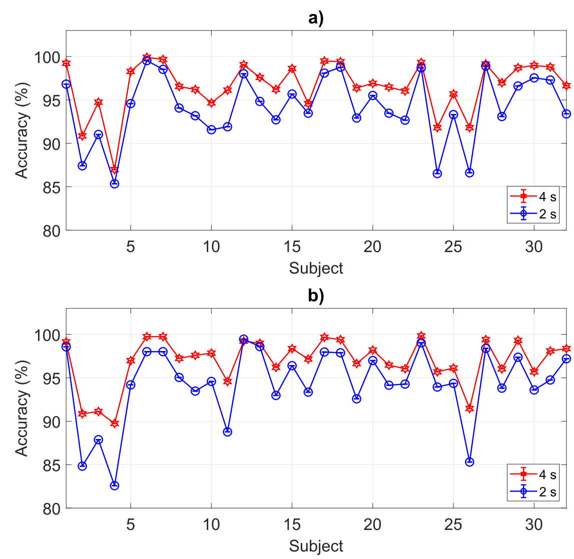

FIGURE 8. Comparison of classification performance of individual subjects of the DEAP dataset for a) Valence dimension and b) Arousal dimension for 4 s and 2 s frame length with 50% overlap. 4 s frame length gives better results.

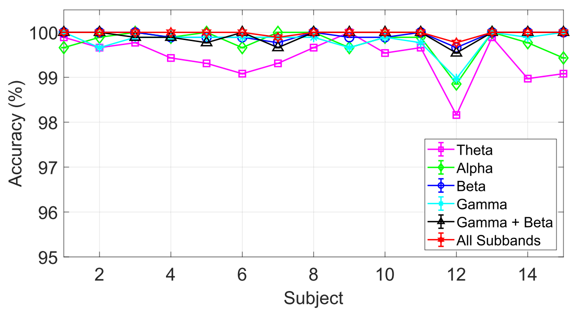

FIGURE 9. **Classification performance of individual subjects of the SEED dataset for different subbands for classifying positive and negative emotions for 4 s frame length with 2 s overlap.**

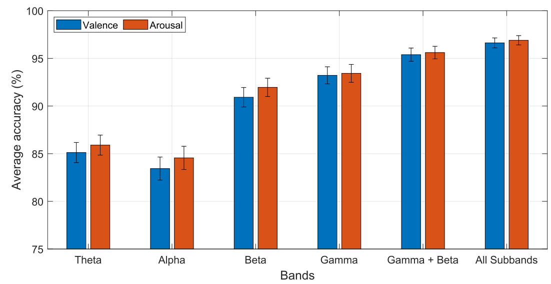

FIGURE 10. **Average classification performance across 32 subjects of the DEAP dataset for 4 s frame length with 2 s overlap. Accuracy is increasing from the lower frequency band to the higher frequency band. Combination of features from all subbands have the highest accuracy.**

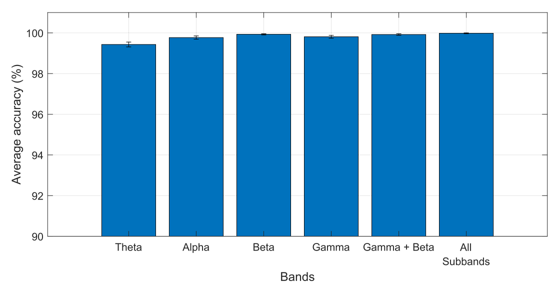

FIGURE 11. **Average classification performance across 15 subjects of the SEED dataset for 4 s frame length with 2 s overlap. Accuracy is increasing from the lower frequency band to the higher frequency band. Combination of features from all subbands have the highest accuracy.**

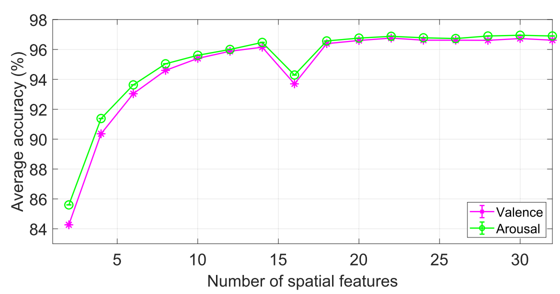

FIGURE 12. **Average classification performance of all subjects of the DEAP dataset as a function of the number of selected spatial features for valence and arousal dimensions.**

(MEMD), flexible analytic wavelet transform (FAWT), and discrete wavelet transform (DWT) were used for EEG signal

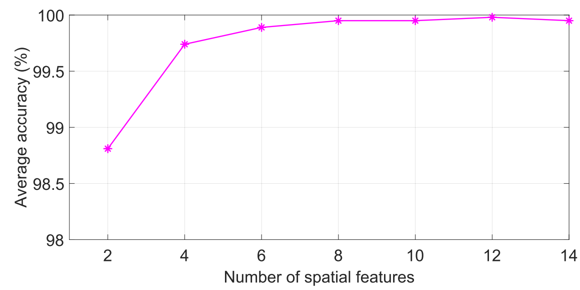

FIGURE 13. **Average classification performance of all subjects of the SEED dataset as a function of the number of selected spatial features for positive and negative emotion classification.**

decomposition. The classification accuracy was presented for individual frequency bands. The higher frequency bands provide better accuracy than the lower frequency bands. Signal framing has gained much attention for improving recognition accuracy because of eliminating the non-stationary problem of the EEG. It is assumed that the signal within this short time frame behaves like a stationary signal [25]. The DWT-based subband decomposition leading to extraction of entropy and energy features for emotion classification was implemented using KNN classifier with the DEAP dataset in [15]. The recognition accuracies were 74%, 77%, 81%, and 86.75% for the rhythmic components theta, alpha, beta, and gamma band respectively for the valence dimension; 71%, 73%, 79%, and 84.05% using theta, alpha, beta, and gamma band respectively for the arousal dimension. The classification accuracy was maximum for 4 s frame length with 2 s overlap with the gamma band in the valence and arousal dimensions. In the proposed approach, the classification has been performed for the individual rhythmic components theta, alpha, beta, and gamma bands respectively. The combination of different subbands (combination of gamma and beta subbands, combination of all subbands) has also been also used to evaluate the classification performance. A better classification performance has been achieved with the combination of subbands rather than the individual subband. Zhuang et al. [16] applied empirical mode decomposition (EMD) for decomposing the EEG signals into intrinsic mode functions (IMFs). The first IMF (IMF1) demonstrated maximum performance. They used frames of 5 s length with the DEAP dataset. The emotion classification is performed with SVM classifier using the features first difference of time series, the first difference of phase, and the normalized energy. The accuracy was 70.41% for valence and 72.10% for arousal dimension. The multivariate empirical mode decomposition (MEMD) was used for signal decomposition in [17]. The power ratio, power spectral density, entropy, Hjorth parameters, and correlation of IMFs were used as features. The independent component analysis (ICA) was employed for dimensionality reduction. The classification accuracy was 67.00% for valence and 51.01% for arousal with K-NN classifier; whereas, 72.87% for valence and 75% for arousal with artificial neural network (ANN) classifier. In the proposed method, the energy and entropy are extracted as features from short-time EEG signals of length 2 s and 4 s of both datasets. It performs better than the EMD-based methods [16], [17]. To improve the classification performance, various methods are used to reduce the dimension of the feature matrix including ICA. Different spatial filtering techniques are also used for the reduction of feature dimension. The CSP-based spatial filtering technique is employed here for dimensionality reduction. It gives excellent performance compared to other existing methods. Various deep learning-based approaches are proposed to classify emotions along with ANN. For instance, Chao et al. proposed a deep learning framework where the multiband feature matrix (MFM) was constructed by combining the features from frequency-domain, spatial characteristics, and frequency band characteristics of the multi-channel EEG signals [24]. A capsule network (CapsNet) was introduced for classifying emotion states based on the input of MFM. The classification accuracy was 67% for the valence dimension and 69% for the arousal dimension for 3 s frame length. Xing et al. proposed a novel framework that contains a linear EEG mixing model built using stack autoencoder (SAE) and the emotion timing model with long short-term memory recurrent neural network (LSTM-RNN) [25]. Short time frames were used for segmenting the EEG signals. The maximum classification performance was obtained using a 1 s frame length with 50% overlap. Therefore, the power spectral density (PSD) with Welch's method was calculated as a feature from each frame and classification was performed with LSTM. The accuracy was 81.10% for valence and 74.38% for arousal. The same framing concept has been used in our proposed approach. However, 2 s and 4 s frame length have been used in this proposed approach and the highest performance has been achieved for 4 s frame length. The channel-wise attention and self-attention based emotion recognition methods were also introduced [27]. The attention-based convolutional recurrent neural network (ACRNN) was used for enhancing emotion recognition performance by extracting more discriminative features from EEG signals. The channel-wise attention mechanism adaptively assigns the weights of different channels. The spatial information was extracted from the EEG signals by a CNN model. Then, the extended self-attention was integrated into an RNN for exploring the temporal information of EEG signals. The accuracy was 93.72% for the valence dimension and 93.38% for the arousal dimension for the DEAP dataset. Gao et al. [28] developed a deep architecture named channel-fused dense convolutional network (CDCN) for EEG-based emotion recognition using differential entropy (DE) as a feature with frame length 1 s. Both DEAP and SEED datasets were used for evaluating the performances of the proposed method. The accuracy for the DEAP dataset was 92.24% for the valence dimension and 92.92% for the arousal dimension. The performance of SEED dataset was 90.63% for classifying positive and negative emotions. A deep learning fusion model of long-short term memories neural networks (LSTM) and graph convolutional neural network (GCNN) named ECLGCNN was used for EEG-based emotion recognition with differential entropy (DE) feature with frame length 6 s [29]. The accuracy was 90.45% for the valence dimension and 90.60% for the arousal dimension with the DEAP dataset in a subject-dependent experiment. A spatial-temporal EEG data representation model with both cascaded and parallel hybrid convolution recurrent neural networks were used for recognizing emotional categories more accurately [30]. The maximum classification accuracy was for cascaded convolution recurrent neural network (CASC-CNN-LSTM for short) with 93.64% for the valence dimension and 93.26% for the arousal dimension with the DEAP dataset for 1 s frame length. Zheng et. al. used different features such as differential entropy (DE), power spectral density (PSD), differential asymmetry (DASM), and rational asymmetry (RASM) for EEG-based emotion classification [13]. The classification was performed with support vector machine (SVM) and deep belief networks (DBNs) on the SEED dataset. Accuracy was presented for different frequency bands (delta, theta, alpha, beta, gamma) with different features and classifiers. The highest accuracy of 86.08% was obtained for DE features with DBN. The dynamical graph convolutional neural networks (DGCNN) with differential entropy (DE) feature was proposed and 90.40% accuracy was obtained for the SEED dataset [31]. The spatial-temporal recurrent neural network (STRNN) was proposed for integrating the feature learning from both spatial and temporal information for EEG-based emotion recognition and 89.50% accuracy was obtained for the SEED dataset for 1 s frame length [32]. A novel concept of electrode-frequency distribution maps (EFDMs) with short-time Fourier transform (STFT) with deep convolutional neural network (CNN) was proposed for automatic feature extraction [33]. The accuracy was 90.59% for the SEED dataset for 1 s frame length. A four-dimensional convolutional recurrent neural network (4D-CRNN) was proposed for emotion recognition from EEG signal in [34] using DEAP and SEED datasets. The accuracy was 94.22% for valence dimension and 94.58% arousal dimension for DEAP dataset. The accuracy was 94.74% for SEED dataset. The frame length was 2 s. Graph convolutional broad network (GCB-net) along with broad learning system (BLS)-based emotion recognition method was proposed in [35] using SEED dataset. The accuracy was 94.24% for classifying positive and negative emotions. Regularized graph neural network (RGNN) was proposed for EEG-based emotion recognition in [36] using SEED dataset and accuracy was 94.24% for classifying positive and negative emotions for 1 s frame length. A novel deep neural network was proposed for emotion classification using EEG systems in [37], which combines the convolutional neural network (CNN), sparse autoencoder (SAE), and deep neural network (DNN) together. By using pearson correlation coefficient (PCC)-based features accuracy was 96.77% for SEED dataset. The accuracy was 89.49% and 92.86% for valence and arousal dimensions respectively for DEAP dataset. The frame length was 8 s. A model based on spiking neural networks (SNNs) architecture named NeuCube was used to classify emotions with the variance feature extraction method in [38]. The accuracy was 96.67% for SEED dataset for 3 s frame length. Optimized residual networks (ResNet) was used for emotion classification in [39] and the accuracy was 93.42% for SEED dataset for 2 s frame length. A multi-model intensive multivariate empirical mode decomposition (iMEMD)-based emotion recognition approach with spatial feature selection method was proposed in [40]. The accuracy was 96.30% for SEED dataset. A combined deep neural network (DNN) model for EEG-based emotion recognition was proposed in [41] where characteristics of multiple neural networks were combined by using deep feature clustering (DFC) method for selecting top EEG features. The accuracy was 97.50% for SEED dataset. The CNN-BiLSTM-MHSA model [42] was proposed for emotion recognition. Here, 'The proposed study 1' which did not use CSP-generated features has 95.12% and 94.62% accuracy for valence and arousal dimensions respectively. 'The proposed study 2' used CSP-generated features but it showed the average performance (binary classification) of valence, arousal, dominance, and liking dimensions.

The emotion classification performance of the proposed method is 96.15% and 96.47% for valence and arousal dimensions respectively with the DEAP dataset. The accuracy with the SEED dataset is 99.95%. The performance of the proposed method is better than other existing methods with both DEAP and SEED datasets.

In the proposed method, the Butterworth bandpass filter has been implemented to preprocess EEG data for extracting different subbands (theta, alpha, beta, and gamma). Then framing has been performed on the individual subbands. Further spatial features have been extracted by applying CSP after calculating entropy and energy-based features from each subband. All the spatial features from each of the subbands have been used for generating the final feature matrix. The inclusion of all subband features has a vital role in the discrimination of emotion tasks. Finally, the support vector machine (SVM) has been used for classifying emotions.

The CSP has been performed on the short-time entropy-energy feature space. Then SVM classifier has been applied on the CSP-generated feature space. It has been found that the emotional state can be recognized from the energy distribution of different subbands of EEG [16]. Besides, entropy features represent the amount of information that is carried by an individual subband [17]. Moreover, the CSP-generated features are more discriminative for emotion classification. That is why the performance of the proposed method is better than the other existing methods.

For this experiment linear, 2nd, 3rd, 4th order polynomial kernels, as well as RBF kernels of SVM have been tried for classification of emotions for the DEAP dataset. The performance comparison of these kernels is exhibited in Fig. 14. The performance is higher for 3rd and 4th order polynomial kernel functions which are almost the same for valence and arousal dimensions (valence dimension: 96.15% (order 3) and 96.36% (order 4); arousal dimension: 96.47% (order 3) and 96.45% (order 4)). Here, the use of the order 4 polynomial kernel is computationally more costly than the order 3 polynomial kernel but contributes a little. In this proposed work, the 3rd order polynomial kernel of SVM has been used for the final performance demonstration. For similarity with the DEAP dataset, the same kernel function has been used for the SEED dataset. The classification performance improves because of generating more discriminative features after applying spatial filtering with CSP. The discrimination performance of the proposed method for the topmost two (one pair) CSP-generated features of gamma (30-47 Hz)

TABLE 3. Performance comparison of the proposed method with other recently developed algorithms based on DEAP dataset.

Methods	Accuracy (%)	
	Valence	Arousal
DWT- KNN [15]	86.75	84.05
EMD - SVM [16]	70.41	72.10
MEMD - ANN [17]	72.87	75.00
CapsNet [24]	67.00	69.00
LSTM-RNN [25]	81.10	74.38
ACRNN [27]	93.72	93.38
CDCN [28]	92.24	92.92
ECLGCNN [29]	90.45	90.60
Casc-CNN-LSTM [30]	93.64	93.26
4D-CRNN [34]	94.22	94.58
PCC [37]	89.49	92.86
CNN-BiLSTM-MHSA [42]	95.12	94.62
Proposed Method	**96.15**	**96.47**

TABLE 4. Performance comparison of the proposed method with other recently developed algorithms based on SEED dataset.

Methods	Accuracy (%)
DE-SVM [13]	83.99
DE-DBN [13]	86.08
CDCN [28]	90.63
DGCNN [31]	90.40
STRNN [32]	89.50
EFDM-CNN [33]	90.59
GCB-net + BLS [35]	94.24
RGNN [36]	94.24
PCC [37]	96.77
NeuCube + variance [38]	96.67
ResNet18 [39]	93.42
Multi-model iMEMD [40]	96.30
4D-CRNN [34]	94.74
Combined DNN [41]	97.50
Proposed Method	**99.95**

TABLE 5. Performance comparison of the proposed method with other recently developed algorithms for individual subbands (theta, alpha, beta, and gamma) based on SEED dataset.

Methods	Accuracy (%)			
	Theta	Alpha	Beta	Gamma
DE-SVM [13]	60.95	66.64	80.76	79.56
DE-DBN [13]	60.77	64.01	78.92	79.19
DE-CDCN [28]	69.84	72.16	80.83	82.63
DE-DGCNN [31]	71.52	74.43	83.65	85.73
DE-STRNN [32]	83.35	82.69	83.41	69.61
Proposed Method	**99.43**	**99.77**	**99.93**	**99.81**

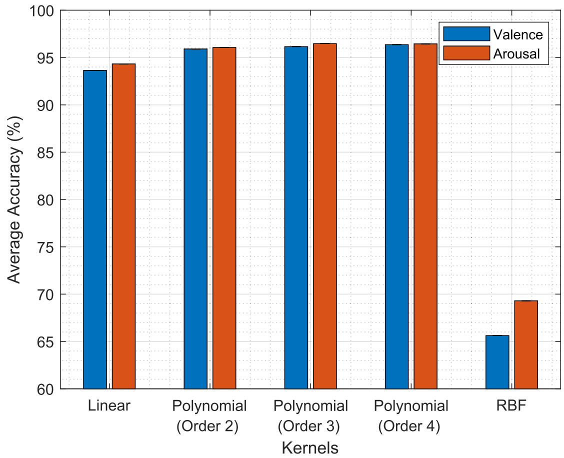

FIGURE 14. The performance comparison of the proposed method with different kernels of SVM classifier conducted on DEAP dataset.

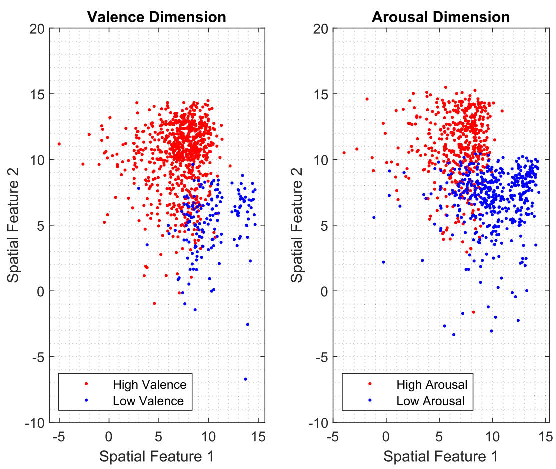

FIGURE 15. Discrimination performance of topmost two (one pair) CSP-generated features of gamma (30-47 Hz) subband of subject 's06' selected from DEAP dataset. Here, the leftmost figure represents the discrimination performance between high and low class in valence dimension whereas the rightmost figure represents the discrimination performance between high and low class in arousal dimension.

subband of subject 's06' selected arbitrarily from the DEAP dataset has been plotted in Fig. 15. Here, the discrimination performance has been represented between high and low classes both for the valence dimension as well as the arousal dimension. The optimal number of spatial features has been selected for generating the final result. For both of the datasets, seven pairs of spatial features resulting in fourteen features have been selected from each subband. The CSP features obtained from each of the four bands have been combined to comprise the final feature vector for each trial. The result of the proposed method has been compared with the result of the other existing methods based on the DEAP dataset and SEED dataset as illustrated in Table 3 and Table 4 respectively. From the comparative study, it is visible that the proposed method gives better classification accuracy than the other methods. The experimental results of the individual subbands obtained from the proposed method have been also compared with the other existing methods in Table 5 for the SEED dataset.

In the proposed method, only high or positive(+) and low or negative(−) emotions (binary classification) are classified in terms of valence and arousal dimensions for the DEAP dataset. On the other hand, in the case of the SEED dataset, the high or positive(+) and low or negative(−) emotions (binary classification) are classified in terms of the valence dimension only. These impressive results have been achieved for the proposed method only for the binary classification (high/low) of a single dimension (valence/arousal). The result would be decreased for the multiclass classification in the combined valence-arousal dimension (++, +−, −+, −− cases), which is not studied in this current study. In the future, it is planned to extend the proposed method to classify more than two classes of emotions that are generated by taking different combinations of valence and arousal values. Besides, different feature selection algorithms will be incorporated in the future works to train the classifier with more discriminating features for higher performance which has not been studied in this current study.

VI. CONCLUSION

The CSP-based feature extraction method has been implemented in this study for emotion classification using short-term EEG signals from multiple subbands. The experimental evaluation has been performed by publicly available DEAP and SEED datasets. In DEAP dataset, 32 EEG channels are used to present two-class (high arousal and low arousal/ high valence and low valence). The second dataset SEED contains 62 EEG channels to present two-class (positive and negative emotions) EEG-based emotion classification in the BCI paradigm. The bandpass filtering has been implemented on the preprocessed EEG data for extracting rhythmic components theta, alpha, beta, and gamma within the frequency range of 4 Hz to 45 Hz. Afterwards signal framing (4 s and 2 s with 50% overlap) has been performed on each subband. Every trial of each subband are represented by short-time attributes (energy-entropy terms) in place of whole EEG samples. The spatial features have been extracted by using CSP technique on the feature matrix of each subband. The features generated by CSP spatial filtering technique are more discriminative for emotion classification. The spatial features of individual subband and the combination of multiple subbands have been used for classification by SVM with cubic polynomial kernel function. The experimental analysis have been also performed for different number of spatial features. The CSP generates discriminative features that enhance the classification performance. The selection of the optimal number of CSP features gives better classification accuracy. The CSP-based emotion recognition approach has given higher performance that outperforms other existing methods. The subject independent emotion classification from EEG signal has been considered as the future extension of this study.

ACKNOWLEDGMENT

The authors thank the reviewers for their insightful comments and helpful suggestions. They also thank the governing body of the Signal Processing and Computational Neuroscience (SiPCoN) Laboratory, Department of Computer Science and Engineering, University of Rajshahi, Bangladesh for the beautiful, clean environment and laboratory facilities provided to them.

REFERENCES

[1] A. Roman-Gonzalez, "EEG signal processing for BCI applications," in *Human—Computer Systems Interaction: Backgrounds and Applications 2*. Berlin, Germany: Springer, 2012, pp. 571–591.

[2] S. Mahmood, J. Shin, I. Farhana, M. R. Islam, and M. K. I. Molla, "Frequency recognition of short-time SSVEP signal using CORRCA-based spatio-spectral feature fusion framework," *IEEE Access*, vol. 9, pp. 167744–167755, 2021.

[3] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *Sci. World J.*, vol. 2014, pp. 1–10, Sep. 2014.

[4] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul. 2019.

[5] Y. Yang and J.-Z. Wang, "From structure to behavior in basolateral amygdala-hippocampus circuits," *Frontiers Neural Circuits*, vol. 11, p. 86, Oct. 2017.

[6] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, and K. Scherer, "Universals cultural differences judgments facial expressions emotion," *J. Personality Social Psychol.*, vol. 53, no. 4, p. 712, 1987.

[7] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Amer. Scientist*, vol. 89, no. 4, pp. 344–350, 2001.

[8] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, p. 1161, Dec. 1980.

[9] J. D. Morris, "Observations: SAM: The self-assessment manikin; An efficient cross-cultural measurement of emotional response," *J. Advertising Res.*, vol. 35, no. 6, pp. 63–68, 1995.

[10] M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel EEG signals using K-nearest neighbor classification," *Technol. Health Care*, vol. 26, no. S1, pp. 509–519, Jul. 2018.

[11] Y. Zhang, S. Zhang, and X. Ji, "EEG-based classification of emotions using empirical mode decomposition and autoregressive model," *Multimedia Tools Appl.*, vol. 77, no. 20, pp. 26697–26710, Oct. 2018.

[12] S. Koelstra, C. Muhl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jun. 2011.

[13] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[14] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[15] Z. Mohammadi, J. Frounchi, and M. Amiri, "Wavelet-based emotion recognition system using EEG signal," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 1985–1990, Aug. 2017.

[16] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, "Emotion recognition from EEG signals using multidimensional information in EMD domain," *BioMed Res. Int.*, vol. 2017, pp. 1–9, Aug. 2017.

[17] A. Mert and A. Akan, "Emotion recognition from EEG signals by using multivariate empirical mode decomposition," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 81–89, Feb. 2018.

[18] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2266–2274, Nov. 2018.

[19] P. Ozel, A. Akan, and B. Yilmaz, "Synchrosqueezing transform based feature extraction from EEG signals for emotional state prediction," *Biomed. Signal Process. Control*, vol. 52, pp. 152–161, Jul. 2019.

[20] Z. Yin, Y. Wang, L. Liu, W. Zhang, and J. Zhang, "Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination," *Frontiers Neurorobot.*, vol. 11, p. 19, Apr. 2017.

[21] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Frontiers Neurosci.*, vol. 12, p. 162, May 2018.

[22] F. Yang, X. Zhao, W. Jiang, P. Gao, and G. Liu, "Multi-method fusion of cross-subject emotion recognition based on high-dimensional EEG features," *Frontiers Comput. Neurosci.*, vol. 13, p. 53, Aug. 2019.

[23] F. Al Machot, A. Elmachot, M. Ali, E. Al Machot, and K. Kyamakya, "A deep-learning model for subject-independent human emotion recognition using electrodermal activity sensors," *Sensors*, vol. 19, no. 7, p. 1659, Apr. 2019.

[24] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion recognition from multiband EEG signals using CapsNet," *Sensors*, vol. 19, no. 9, p. 2212, May 2019.

[25] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurorobotics*, vol. 13, p. 37, Jun. 2019.

[26] M. A. Asghar, M. J. Khan, Y. Amin, M. Rizwan, M. Rahman, S. Badnava, and S. S. Mirjavadi, "EEG-based multi-modal emotion recognition using bag of deep features: An optimal feature selection approach," *Sensors*, vol. 19, no. 23, p. 5218, Nov. 2019.

[27] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "EEG-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 382–393, Jan. 2023.

[28] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 4, pp. 945–954, Dec. 2021.

[29] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106954.

[30] J. Chen, D. Jiang, Y. Zhang, and P. Zhang, "Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset," *Comput. Commun.*, vol. 154, pp. 58–65, Mar. 2020.

[31] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul. 2020.

[32] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2018.

[33] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, Sep. 2020, Art. no. 107506.

[34] F. Shen, G. Dai, G. Lin, J. Zhang, W. Kong, and H. Zeng, "EEG-based emotion recognition using 4D convolutional recurrent neural network," *Cognit. Neurodynamics*, vol. 14, no. 6, pp. 815–828, Dec. 2020.

[35] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-Net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 379–388, Jan. 2022.

[36] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1290–1301, Jul. 2022.

[37] J. Liu, G. Wu, Y. Luo, S. Qiu, S. Yang, W. Li, and Y. Bi, "EEG-based emotion classification using a deep neural network and sparse autoencoder," *Frontiers Syst. Neurosci.*, vol. 14, p. 43, Sep. 2020.

[38] Y. Luo, Q. Fu, J. Xie, Y. Qin, G. Wu, J. Liu, F. Jiang, Y. Cao, and X. Ding, "EEG-based emotion classification using spiking neural networks," *IEEE Access*, vol. 8, pp. 46007–46016, 2020.

[39] K. H. Cheah, H. Nisar, V. V. Yap, C.-Y. Lee, and G. R. Sinha, "Optimizing residual networks and VGG for classification of EEG signals: Identifying ideal channels for emotion recognition," *J. Healthcare Eng.*, vol. 2021, pp. 1–14, Mar. 2021.

[40] M. A. Asghar, M. J. Khan, M. Rizwan, M. Shorfuzzaman, and R. M. Mehmood, "AI inspired EEG-based spatial feature selection method using multivariate empirical mode decomposition for emotion classification," *Multimedia Syst.*, vol. 28, no. 4, pp. 1275–1288, 2021.

[41] M. A. Asghar, M. J. Khan, M. Rizwan, R. M. Mehmood, and S.-H. Kim, "An innovative multi-model neural network approach for feature selection in emotion recognition using deep feature clustering," *Sensors*, vol. 20, no. 13, p. 3765, Jul. 2020.

[42] Z. Hu, L. Chen, Y. Luo, and J. Zhou, "EEG-based emotion recognition using convolutional recurrent neural network with multi-head self-attention," *Appl. Sci.*, vol. 12, no. 21, p. 11255, Nov. 2022.
[43] M. Vetterli and C. Herley, "Wavelets and filter banks: Theory and design," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2207–2232, Sep. 1992.
[44] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, no. 1, p. 39, 2012.
[45] S. Ali, J. Ferdous, E. Hamid, and K. I. Molla, "A novel features selection approach with common spatial pattern for EEG based brain-computer interface implementation," *IETE J. Res.*, vol. 68, no. 3, pp. 1757–1771, 2019.
[46] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
[47] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Jan. 2005, pp. 5392–5395.
[48] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Optimizing spatio-temporal filters for improving brain-computer interfacing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 315–322.

IFFAT FARHANA received the B.Sc. and M.Sc. degrees in computer science and engineering from the University of Rajshahi, Rajshahi, Bangladesh, in 2018 and 2020, respectively.

She was a Research Fellow with the Information and Communication Technology Division (ICT Division), Ministry of Posts, Telecommunications and Information Technology, Bangladesh, from 2019 to 2020. She was a former Lecturer with the Department of Computer Science and Engineering, University of Information Technology and Sciences (UITS), Dhaka, Bangladesh, and the Green University of Bangladesh, Dhaka. She is currently a Research Fellow with the Signal Processing and Computational Neuroscience Laboratory (SiPCoN), University of Rajshahi. Her research interests include brain signal (EEG) processing, brain–computer interface (BCI), emotion recognition, and biomedical engineering.

JUNGPIL SHIN (Senior Member, IEEE) received the B.Sc. degree in computer science and statistics and the M.Sc. degree in computer science from Pusan National University, South Korea, in 1990 and 1994, respectively, and the Ph.D. degree in computer science and communication engineering from Kyushu University, Japan, in 1999, under a scholarship from the Japanese Government (MEXT). He was an Associate Professor, a Senior Associate Professor, and a Full Professor with the School of Computer Science and Engineering, The University of Aizu, Japan, in 1999, 2004, and 2019, respectively. He has coauthored more than 300 published papers for widely cited journals and conferences. His research interests include pattern recognition, image processing, computer vision, machine learning, human–computer interaction, non-touch interfaces, human gesture recognition, automatic control, Parkinson's disease diagnosis, ADHD diagnosis, user authentication, machine intelligence, as well as handwriting analysis, recognition, and synthesis. He is a member of ACM, IEICE, IPSJ, KISS, and KIPS. He served as the program chair and as a program committee member for numerous international conferences. He serves as the Editor for IEEE journals and *Sensors* (MDPI). He serves as a reviewer for several major IEEE and SCI journals.

SHABBIR MAHMOOD received the B.Sc. and M.Sc. degrees in computer science and engineering from the University of Rajshahi, Rajshahi, Bangladesh, in 2018 and 2020, respectively.

He was a Research Fellow with the Information and Communication Technology Division (ICT Division), Ministry of Posts, Telecommunications and Information Technology, Bangladesh, from 2019 to 2020. He was a former Lecturer with the Department of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh. He is currently a Lecturer with the Department of Computer Science and Engineering, Bangamata Sheikh Fojilatunnesa Mujib Science and Technology University, Jamalpur, Bangladesh. He is also a Research Fellow with the Signal Processing and Computational Neuroscience Laboratory (SiPCoN), University of Rajshahi. His research interests include brain signal (EEG/SSVEP) processing, brain–computer interface (BCI), and biomedical engineering.

MD. RABIUL ISLAM (Member, IEEE) received the B.Sc. and M.Sc. degrees in computer science and engineering from the University of Rajshahi, Bangladesh, in 2008 and 2009, respectively, and the Ph.D. degree from the Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan, in September 2018. He was a Research Fellow with the National Science, Information and Communication Technology (NSICT), Bangladesh, from 2008 to 2009. From October 2014 to September 2015, he was a Research Student with the Department of Electrical and Information Engineering, TUAT. He was an Assistant Professor (research) with TUAT. From April 2021 to June 2021, he was a Visiting Researcher with The University of Aizu, Fukushima, Japan. He is currently a Postdoctoral Research Fellow with the Department of Medicine/Division of Nephrology, The University of Texas Health Science Center at San Antonio, USA. His research interests include brain signal (EEG/iEEG and SSVEP) processing, bioinformatics, brain–computer interface, and epilepsy.

MD. KHADEMUL ISLAM MOLLA (Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics and computer science from the Shahjalal University of Science and Technology, Bangladesh, in 1995 and 1997, respectively, and the Ph.D. degree from The University of Tokyo, Tokyo, Japan, in 2006.

In 2006, he joined as an Associate Professor with the Department of Computer Science and Engineering, University of Rajshahi, Bangladesh, where he is currently a Professor. He was a JSPS Postdoctoral Fellow with the Department of Information and Communication Engineering, The University of Tokyo, from September 2006 to September 2008. He was a Research Fellow with the University of Alberta, Edmonton, AB, Canada, from November 2010 to October 2011. From January 2012 to June 2012, he was an Associate Professor with King Khalid University, Saudi Arabia. He was an NICT Fellow with The University of Tokyo, from January 2013 to July 2013. He was a JSPS Invited Researcher and a Visiting Professor with The University of Tokyo, the Tokyo University of Agriculture and Technology, Tokyo, and the Kyushu Institute of Technology, Fukuoka, Japan, several times. His research interests include the application of signal processing, brain–computer interface, biomedical engineering, and time series analysis. He is a member of APSIPA. He was the Organizing Chair of the Asia-Pacific Signal and Information Processing Association (APSIPA) BioSiPS 2014.