



Employing PCA and *t*-statistical approach for feature extraction and classification of emotion from multichannel EEG signal

Md. Asadur Rahman^{a,*}, Md. Foisal Hossain^b, Mazhar Hossain^b, Rasel Ahmmed^c

^a Department of Biomedical Engineering, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh

^b Department of Electronics and Communication Engineering, Khulna University of Engineering & Technology (KUET), Khulna 9203, Bangladesh

^c Department of Electronics and Telecommunication Engineering, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

ARTICLE INFO

Article history:

Received 18 July 2019

Revised 22 September 2019

Accepted 13 October 2019

Available online 2 November 2019

Keywords:

Brain-computer interface (BCI)
Emotion recognition
Electroencephalogram (EEG)
SEED (SJTU Emotion EEG dataset)
Support vector machine (SVM)
Artificial neural network (ANN)
Linear discriminant analysis (LDA)
k-Nearest neighbor (kNN)
Classification accuracy

ABSTRACT

To achieve a highly efficient brain-computer interface (BCI) system regarding emotion recognition from electroencephalogram (EEG) signal, the most crucial issues are feature extractions and classifier selection. This work proposes an innovative method that hybridizes the principal component analysis (PCA) and *t*-statistics for feature extraction. This work contributes to successfully implement spatial PCA to reduce signal dimensionality and to select the suitable features based on the *t*-statistical inferences among the classes. The proposed method has been applied on the SEED dataset (SJTU Emotion EEG Dataset) that yielded significant channels and features for getting higher classification accuracy. With extracted features, four classifiers—support vector machine (SVM), artificial neural network (ANN), linear discriminant analysis (LDA), and k-nearest neighbor (kNN) method were applied to classify the emotional states. The classifiers showed slightly different classification accuracies compared to each other. ANN and SVM showed the highest classification accuracy (86.57 ± 4.08 and 85.85 ± 5.72) in case of subject dependent approach. On the other hand, the proposed method provides 84.3% and 77.1% classification accuracy with ANN and SVM, respectively in case of subject independent approach. Eventually, the proposed method and its outcomes demonstrate that this proposal is better than the several existing methods in emotion recognition.

© 2019 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Emotions are the psychological stages of feeling that can be intertwined through circumstances, temperament, relationships, motivation, dispositions, etc. Since emotion has a strong correlation with mental activity, it can be triggered by some external stimuli such as movie, pictures, scenario, sound, or some internal approaches like memorizing interesting, emotional, sad, inspiring memory, etc. Recently, research community has shown extensive interest to establish human emotion interaction with machine by

proper brain-computer interface (BCI) [1]. To develop a secure BCI system, emotion recognition plays a vital role. We can present a security system where emotion cannot be controlled forcefully suppose that an ATM machine will not dispense money when the user has scary emotion in the situation of robbery. Moreover, emotion recognition process has effective control over device such as computer and assistive applications [2]. Besides, emotion recognition has some other applications like music therapy [3,4], safe driving [5,6], artificial intelligence [7], etc.

Since all kinds of psychophysiological changes affect brain activities, Electroencephalogram (EEG) could be a potential way to read the human emotion from the neuro-electric signal. In recent researches [8–11], inferring the human emotions from the EEG signal becomes a very active research topic. To recognize the emotion from EEG signal, two major parts are needed to be taken into account: i) Significant feature extraction from emotional EEG signals and ii) Emotion classification [12]. Having a number of noises and artifacts in EEG signal, extracting the optimal features from multiple channel EEG signals is a challenging issue. However, a number of feature extraction methods have already been

* Corresponding author.

E-mail addresses: rahman1515701@stud.kuet.ac.bd (Md. Asadur Rahman), foisal@ece.kuet.ac.bd (Md. Foisal Hossain).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.



Production and hosting by Elsevier

proposed by different researchers for emotion recognition or classification. These features are broadly in time domain, frequency domain, time–frequency domain, or concurrent time and frequency domain approach. Another problem is the channel selection from the multichannel EEG signals because the reduced channel information reduces the feature dimension that helps to increase the classification accuracies [13–15].

There are a number of approaches to acquire emotional EEG data and there are several databases freely available like DEAP [16], SEED [17], MAHNOB [18], etc. The SEED dataset has three class emotional EEG data where emotional stimuli were given by three kinds of video clips: Neutral, Positive, and Negative. Several proposals [12,13,19–21] are available for feature extraction and classification of the emotions from the SEED dataset. The authors in [12] proposed a group sparse canonical correlation analysis (GSCCA) for channel selection and feature extraction. On the other hand, a deep neural network (DNN) has been proposed for channel selection based on the distribution of mean absolute weights in [13]. The researches proposed in [19,20] also used deep learning algorithm to classify the emotional EEG data from SEED dataset. Using deep learning for only three-class classification is like *using cannon to kill a mosquito*. In addition, some other works [14,22] utilized shallow or extreme learning method those got lower classification accuracy. Although deep learning can yield us high classification accuracy, its computational configuration will not be practical for the limited number of classes.

In this work, a novel feature extraction procedure has been proposed for the emotion classification utilizing the property of the principal component analysis (PCA) and *t*-statistics. In our proposal, PCA has been used on the selected EEG data channels to reduce information similarity instead of their extracted features. In our proposal, all the EEG channels were divided into several major groups and then we applied the PCA to reduce their dimensions. The total EEG channels are divided into five major groups: Prefrontal, Frontal, Central, Parietal, and Occipital. For example, the frontal group includes the channels: F7, F5, F3, F1, Fz, F4, F6, and F8. PCA is used on these channels and only the first principal component is considered for the further steps. Therefore, the representative nine channels of frontal lobe are linearly transformed into one channel information which conveys the information of the maximum variations. This procedure reserves the maximum frequency domain content inside the signal from all the channels. It is previously described that for emotion recognition, the most effective features are of frequency domain [21]. Since PCA converts and reduces the dimensions remaining the maximum variations from all the signals, it also preserves the required frequency content. This procedure reduces the 62 channels to only 5 EEG data channels. From these data channels, a number of renowned features of time and frequency domains were extracted. The significance ($p < 0.05$) of the features were tested by *t*-statistics and the inference was calculated compared to the neutral (neutral VS negative and neutral VS positive) data.

This work aims to use the shallow networks to get the equivalent classification accuracy of the deep neural networks utilizing the proposed feature extraction method. The significant features were selected from the results of *t*-test and these features were used to model the emotion by some well-known classifiers: Support Vector Machine (SVM) [23,24], Linear Discriminant Analysis (LDA) [24,25], Artificial Neural Network (ANN) [23,25–27], and k-Nearest Neighbor (kNN) [24,28]. The classification was performed considering subject depended and subject independent approach to observe the classification accuracy of the proposed model. The classification accuracy found by the proposed work suggests that the proposed method is an effective way to extract the features from the emotional EEG signals. The results are compared with the other existing methods applied in SEED database and we found

that the proposed method outperforms several conventional methods and comparable to the results of the deep neural network.

The paper is organized as follows: The data collection and the applied mathematical methods are described in Section 2. The results and discussions are presented with the necessary tables and graphs are in Section 3. Finally, we have concluded our research outcomes with limitations and future research perspectives in Section 4.

2. Materials and methods

From data collection to results include a number of steps according to the proposed methodology like EEG signal preprocessing, data reduction, different feature extraction, their significance test by *t*-test, classifications, etc. The steps of the proposed method have been presented in Fig. 1 and the corresponding elaborate descriptions of the analytical steps are discussed gradually in this section.

2.1. Data collection

The data we used in our research work is known as SJTU Emotion EEG Dataset (SEED). The data contains the three-class emotional EEG data of 15 subjects. The data were recorded by showing the participants some emotional film clips (positive, negative and neutral video clips) as given in Fig. 2. After watching the film during the data acquisition, the participants were requested to report their emotional reactions by completing the questionnaire for the feedback purpose. The EEG data acquiring device was ESI Neuro-Scan System with sampling rate of 1000 Hz from 62-channel active AgCl electrode cap according to the international 10–20 system [17]. The details about the dataset are described in [13].

2.2. Preprocessing of EEG data

The EEG data of the SEED dataset was primarily preprocessed by filtering with cutoff frequency 70 Hz and downsampled to 200 Hz from 1000 Hz acquisition sampling rate. In this work, all data were processed again. At first, the data was filtered by 50 Hz notch filter to remove the power line noise. Then the effect of eyeblink was removed by the Matlab based automatic EEG artifact rejection toolbox, EAWICA toolbox [29,30]. The eye blink effect removed EEG signal was bandpass filtered at 1–35 Hz. The effect of the preprocessing steps on the raw EEG signal has been shown in Fig. 3. Then, the filtered EEG signals were grouped into 5 different set to apply the PCA to reduce its dimensions. The processing steps can be compared with the block diagram given in Fig. 4.

2.3. Dimensionality reduction by PCA

PCA is a multivariate analytical method based on the linear transformation that is often used to reduce the dimensionality of the data, to extract significant information from big data, to analyze the variable structures, etc. [28,31]. The PCA method has been used for dimensionality reduction of the EEG signals in this paper. Since the spatial resolution of the EEG signal is poor, considering all channels for feature extraction is just increasing the burden. In this situation, we can use PCA because it can find the maximum variation from the input higher dimensions. Therefore, for reducing the number of channels of interest a group of the EEG signal can be transformed into a single signal using PCA.

In our proposed work, the frontal group was constructed by considering the following channels: F7, F5, F3, F1, Fz, F2, F4, F6,

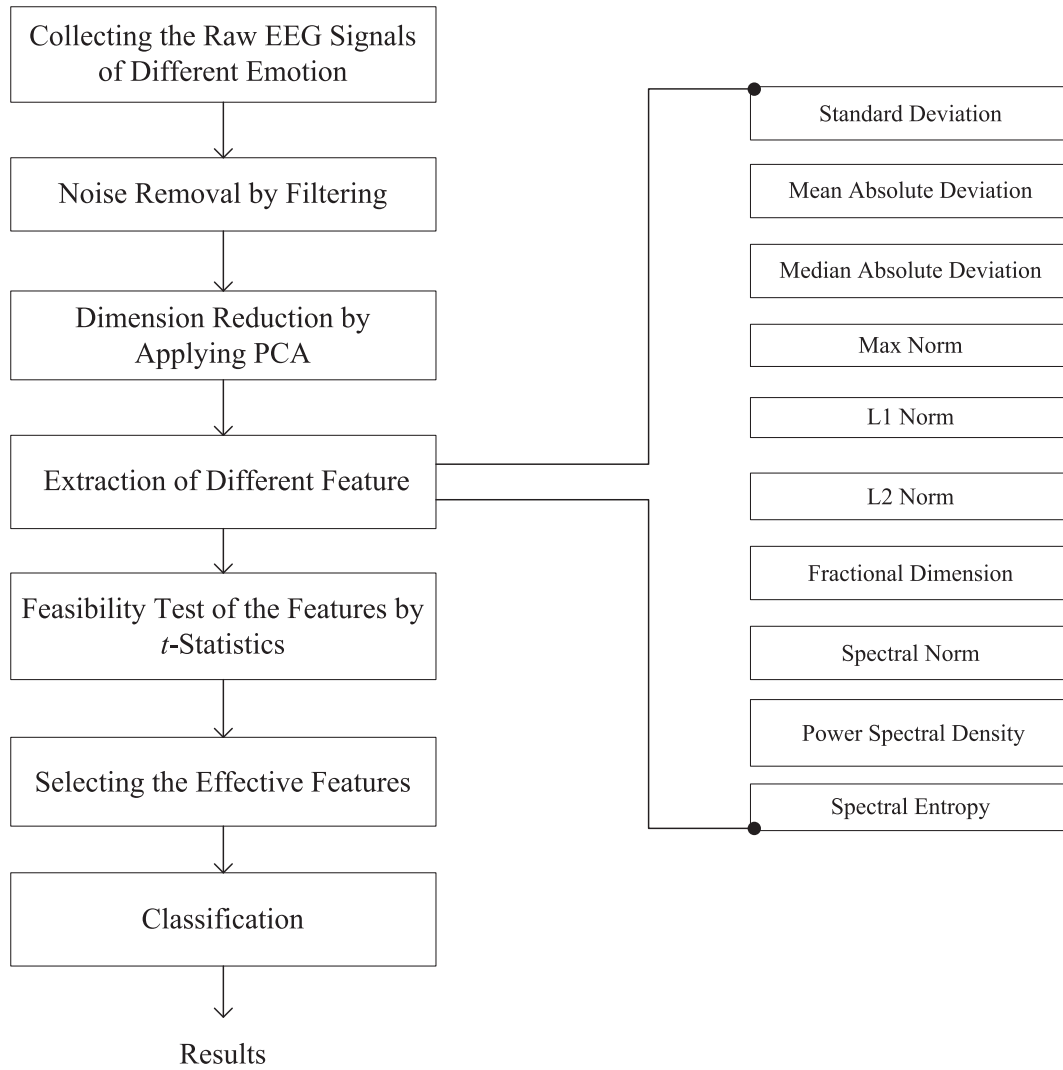


Fig. 1. Flow diagram of the proposed method that was applied to the multichannel EEG signal for finding the classification accuracy in emotion recognition from SEED dataset.



Fig. 2. EEG data acquisition arrangement showing the emotional video clips [13,17].

and F8. We applied PCA on these nine channels and took the 1st PCA. The similar procedure was applied to prefrontal, frontal, central, parietal, and occipital lobe groups and transformed all the signals to five signals of interest of every set of the input EEG signal. This procedure yielded 5 set of data instead of input 62 sets of data

(the EEG signal contains 62 channels). The channel number with their groups and the excluded channels are graphically presented in Fig. 5. If we have an n set of data channels, $X(= X_1, X_2, \dots, X_n)$ that is to be transformed by PCA. Applying PCA, the data vector X is transformed to Y obeying the following relation [32],

$$Y = U^T X \quad (1)$$

where U^T is the eigenvector of the covariance matrix of X . Furthermore, the eigenvectors of the matrix U^T is sorted as descending order. Now, if we consider first r rows of U^T to project the data, the data become of r dimensional instead of n dimensions. We know that any nonsingular matrix X can be decomposed as (2) using singular value decomposition.

$$X = \Omega \Gamma \Psi^T \quad (2)$$

In (2), Ω is an $m \times n$ matrix such that $\Omega^T \Omega = I$ with orthonormal columns, Ψ as an $n \times n$ matrix ($\Psi^T \Psi = I$). Here the samples are denoted by m and the channel number is denoted by n . Furthermore, Γ is an $n \times n$ diagonal matrix with the elements ≥ 0 which is called singular values. In addition, we know that the covariance matrix, Σ of X is found as (3) and consequently the transformed data, Y is evaluated as (4) [32].

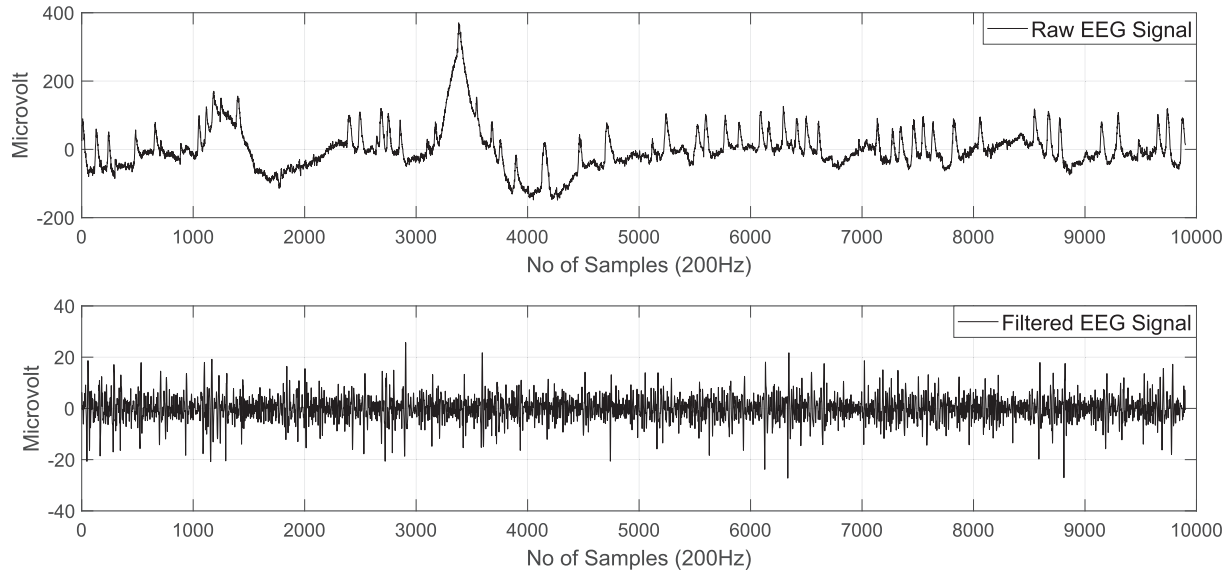


Fig. 3. Raw EEG signal and the filtered output by the proposed preprocessing steps (notch filtering + bandpass filtering + eye blink removal by EAWICA toolbox [28,29]).



Fig. 4. The processing methods of the EEG signal before feature extraction.

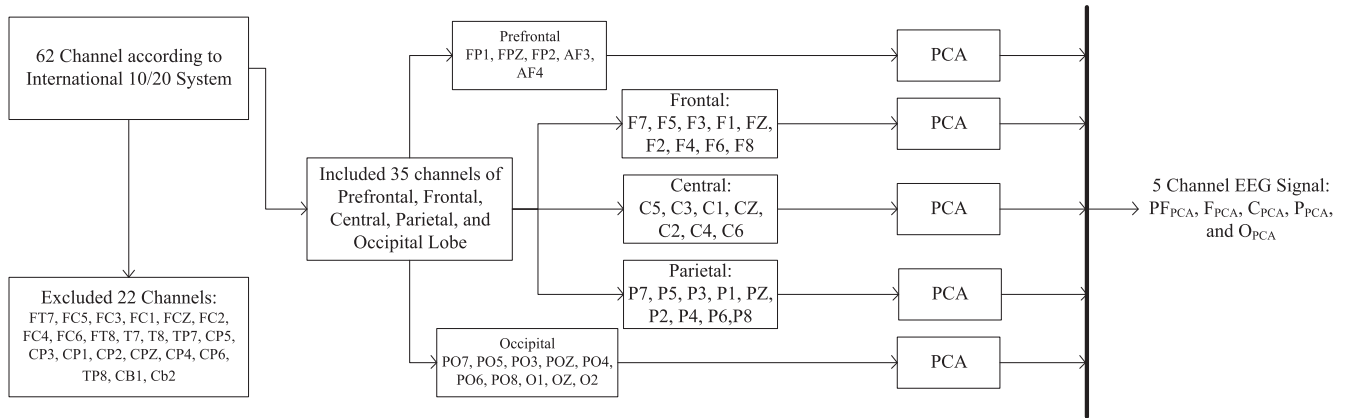


Fig. 5. This figure describes the included and excluded EEG channels along with the considered channels of each group to perform PCA to make the 42 (rest 22 excluded) channels EEG data to 5 data channels.

$$\Sigma = \frac{1}{n} \mathbf{X} \mathbf{X}^T = \frac{1}{n} (\mathbf{\Omega} \mathbf{\Gamma} \mathbf{\Psi}^T) (\mathbf{\Omega} \mathbf{\Gamma} \mathbf{\Psi}^T)^T = \frac{1}{n} \mathbf{\Omega} \mathbf{\Gamma}^2 \mathbf{\Omega}^T \quad (3)$$

$$\mathbf{Y} = \mathbf{U}^T \mathbf{X} = \mathbf{U}^T \mathbf{\Omega} \mathbf{\Gamma} \mathbf{\Psi}^T \quad (4)$$

We considered $r = 1$ for converting all channels of a group to one channel data by the applied PCA transformation. In practice, the channels of the frontal lobe of a randomly selected participant of neutral stimuli are shown in Fig. 6(a) where only 2000 samples are shown. These data are converted to one channel information applying proposed PCA which is shown in Fig. 6(b). This figure illustrates the effect of applying the proposed PCA algorithm for converting a group of channels to a single-channel EEG data.

2.4. Feature extraction

The features are the representatives of the signal. Therefore, appropriate feature extraction is an important task before the classification. Eventually, an appropriate feature extraction method converts one or several signals into a feature vector [32,33]. The selected features are expected to have the properties to represent the signal and to discriminate from the others. Most of the researchers regarding machine learning applications believe that proper feature extraction is the key to construct an efficient predictive model [34]. There are several kinds of features based on time domain and frequency domain of the EEG signal. Here, we have considered the widely used time and frequency domain features. The origin and mathematical clarifications are given gradually.

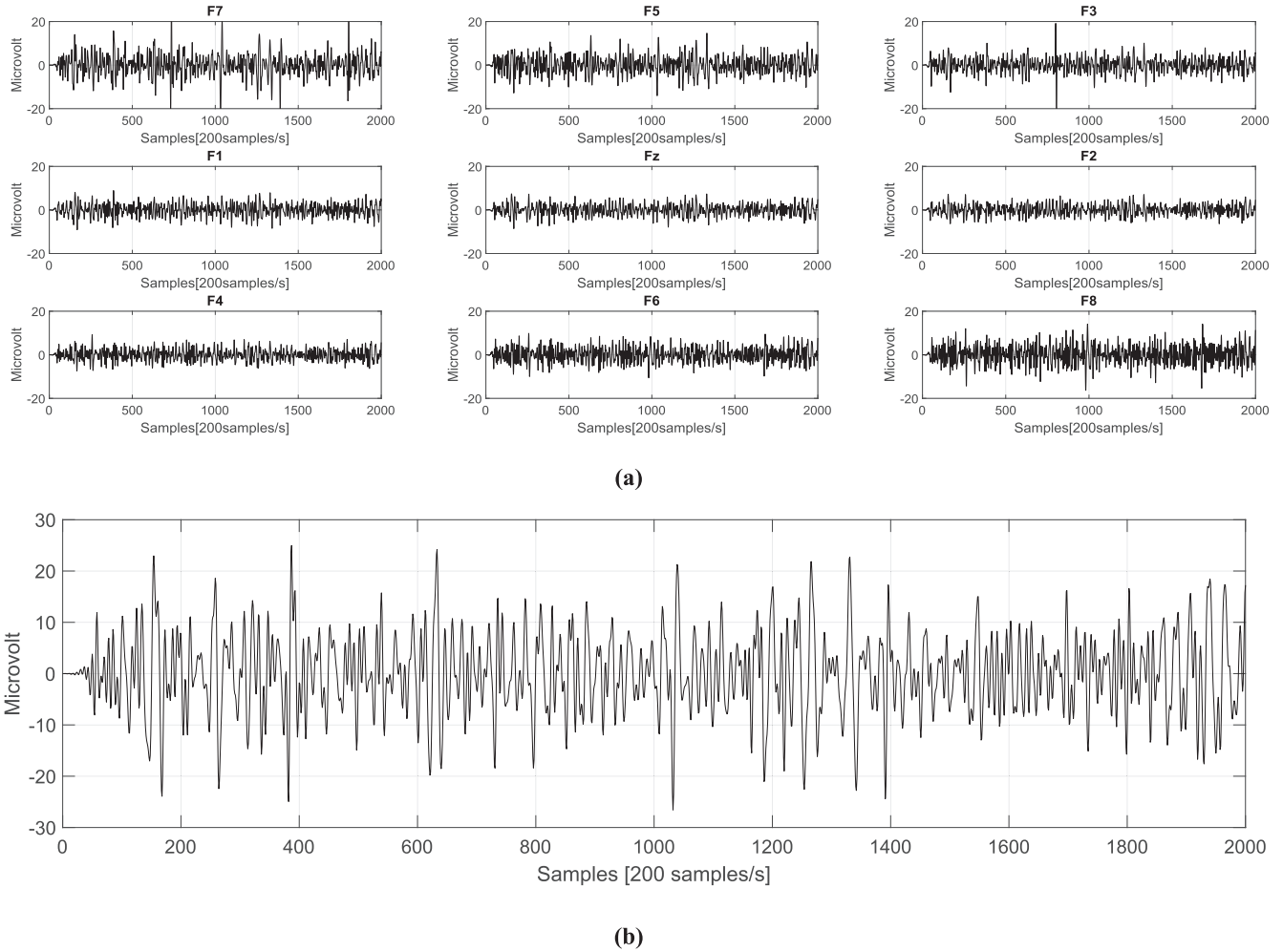


Fig. 6. Graphical representation of a single channel converted EEG data from 9 channel EEG signal using PCA: **(a)** The filtered selected frontal nine-channel EEG signal and **(b)** the converted one channel EEG signal from the previous nine.

2.4.1. Standard Deviation

This feature quantifies the deviation of the signal points with respect to its mean value. Mathematically it can be denoted as,

$$SD = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

2.4.2. Mean Absolute Deviation

Suppose, in a data set there exist n number of observations like $x_1, x_2, x_3, \dots, x_n$. The average or mean of the data set can be expressed as, $\bar{x} = 1/n(\sum_{k=1}^n x_k)$. Therefore, the absolute distances from the mean of the data (\bar{x}) is $(\sum_{k=1}^n |x_k - \bar{x}|)$. From this viewpoint, the definition of mean absolute deviation is average of absolute distances from the mean of the data and that is why MAD is calculated as given,

$$MAD = \frac{\sum_{k=1}^n |x_k - \bar{x}|}{n} \quad (6)$$

MAD is also termed as mean absolute deviation. The formula for mean absolute deviation for an ungrouped data is (6) but for grouped data the formula for mean absolute deviation is

$$MAD = \frac{\sum_{k=1}^n |x_k - \bar{x}| f_k}{n} \quad (7)$$

where, f_k depicts the frequencies corresponding to the given values or observations.

2.4.3. Median Absolute Deviation (MedAD)

MedAD is a quantity of a data set that is quite similar to MAD except the absolute deviation is measured from the viewpoint of median value instead of mean of that data. Therefore, MedAD is calculated as,

$$MedAD = median(|x_k - median(x)|) \quad (8)$$

2.4.4. L1-Norm

L1-norm also known as mean norm or least absolute deviations. It is calculated from the sum of the absolute values of the dataset.

$$L1 - norm = \sum_{k=1}^n |x_k| \quad (9)$$

2.4.5. L2-Norm

L2-norm also represented as mean-square norm or least-squares norm and illustrates as the square root of the sum of the absolute values of the dataset.

$$L2 - norm = \sqrt{\sum_{k=1}^n |x_k|^2} \quad (10)$$

2.4.6. Max Norm

Max norm is also represented as infinity norm or uniform norm and described as the maximum of the absolute values of the dataset.

$$L_{\infty} - \text{norm} = \max |x_i| \quad (11)$$

2.4.7. Fractional Dimension

A fractal dimension is a ratio that yields a numerical index of complexity. It also presents a measure of the space-filling capacity of a pattern that describes the procedure of how a fractal can differ from the space. Higuchi's algorithm is a prominent technique to calculate the fractal dimension, D of a time series. According to the Higuchi algorithm [35], a finite set of observations with a regular interval is considered as,

$$X(1), X(2), \dots, X(N) \quad (12)$$

From the series given in (12), a new member X_k^m , can be introduced as,

$$X_k^m; X(m), X(m, k), X(m, 2k), \dots, X(m, \lceil \frac{N-m}{k} \rceil k) \quad (13)$$

Here, $m = 1, 2, \dots, k$; and where the notation $\lceil * \rceil$ presents the Gauss notation, which is the larger integer, and both k and m are integers. In addition, m indicates the initial time and k presents the interval time. For a k time interval, we get k set of new time series. As for example, suppose that $k = 4$ and $n = 50$, four new time series will be found as,

$$\left. \begin{aligned} X_4^1 &= X(1), X(5), X(9), \dots, X(47) \\ X_4^2 &= X(2), X(6), X(10), \dots, X(48) \\ X_4^3 &= X(3), X(7), X(11), \dots, X(49) \\ X_4^4 &= X(4), X(8), X(12), \dots, X(50) \end{aligned} \right\} \quad (14)$$

This algorithm defines the length of the curve associated with each time series, X_k^m as follows:

$$L_m(k) = \frac{1}{k} \left(\sum_{i=1}^{\lceil \frac{N-m}{k} \rceil} (X(m + ik) - X(m + (i-1)k)) \right) \left(\frac{N-1}{\lceil \frac{N-m}{k} \rceil k} \right) \quad (15)$$

where the term $\lceil \frac{N-m}{k} \rceil k$ represents a normalization factor. The average value $\langle L(k) \rangle$ is generally considered for the time series as given in (15). If the average value of $\langle L(k) \rangle$ follows a power law $\langle L(k) \rangle \propto k^{-D}$, then the curve is considered to be fractal with a dimension D . The described Higuchi's algorithm is applicable for non-stationary time series, as well.

2.4.8. Spectral Norm

We know that the natural norm comes from the L2 norm. The spectral norm is the highest singular value of a matrix. In a word, we can express it as the maximum 'scale', by the means of a matrix can 'stretch' a vector. Suppose that, the conjugate transpose of X is X^H , so that $(\bar{a}_{ji})^H = (\bar{a}_{ji})$, then our desired spectral norm is estimated as the square root of the maximum eigenvalue of $X^H X$ [35–37], i.e.,

$$\|X\|_2 = (\text{maximum eigenvalue of } X^H X)^{1/2} = \max_{|x|_2 \neq 0} \frac{|Xx|_2}{|x|_2} \quad (16)$$

2.4.9. Power Spectral Density

Suppose, P is the average power of a signal $x(t)$, then we can write for the total time period T as,

$$P = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |x(t)|^2 dt \quad (17)$$

If the frequency content of the $x(t)$ is $\hat{x}(\omega)$ by Fourier transformation, the power spectral density can be calculated as [37–39],

$$S_{xx}(\omega) = \lim_{T \rightarrow \infty} \bar{E} [|\hat{x}(\omega)|^2] \quad (18)$$

2.4.10. Spectral Entropy

The spectral entropy represents the spectral power distribution of a signal. This concept is constructed based on the Shannon entropy. This feature of the medical signal is widely used in biomedical signal classification [40]. Suppose, the considered signal is $x(n)$ and its power spectrum is $s(m) = |X(m)|^2$, $X(m)$ where is the discrete Fourier transform of $x(n)$. Then, the probability distribution $P(m)$ is calculated as,

$$P(m) = \frac{S(m)}{\sum_i S(i)} \quad (19)$$

Consequently, the spectral entropy H is presented as,

$$H = - \sum_{m=1}^N P(m) \log_2 P(m) \quad (20)$$

2.5. Classification

A predictive machine learning-based model is needed to classify the EEG data utilizing the features. In this consequence, we train some supervised network to construct the predictive model. A number of supervised machine learning based algorithms exist. Among them, the extensively used classifiers are ANN, SVM, LDA, and k-NN. In this work, we applied these classifiers for emotional EEG data classification. A short description of these algorithms is presented as follows.

2.5.1. ANN

ANN utilizes a special technique of artificial intelligence that can able to mimic the behavior of the human brain [40]. In a feed-forward ANN, there are three types of layer: input layer, output layer, and hidden layer those are called multilayer perceptrons. The input layer buffers the distribution of the input signals $x_n (n = 1, 2, 3, \dots)$ and sends it to the hidden layer neurons. The hidden layer neurons add the input signals, x_n after weighing them with the respective connectivity strengths, W_{nl} from the input layer. Finally, the output Y is calculated as [26,41],

$$Y_l = f \left(\sum_{n=1}^n W_{nl} x_n \right) \quad (21)$$

Here, the numbers of neuron are l . Furthermore, the function $f(\cdot)$ can be of different functional forms like sigmoidal, radial basis function, linear, hyperbolic tangent, etc. With a similar approach, the outputs of the neurons are computed. The gradient descent algorithm is the widely used propagation algorithm to train the multilayer perceptron training algorithm. It yields the change, ΔW_{nl} which is the weight of a connection between neurons n and l as,

$$\Delta W_{nl} = \eta \delta_l x_n \quad (22)$$

where the learning parameter has been denoted by the notation, η and δ_j is a factor which value rely on the decision of the neuron is of input layer or hidden layer.

2.5.2. SVM

SVM is a supervised machine learning classification algorithm which has good scalability on the high dimensional features and explicit control on the errors. SVM initiates necessary hyperplanes among the features of the different classes during its training

session to maximize the margins among the classes [42]. The points are nearest to the hyperplanes are called support vectors. In a two-dimensional feature space, the discriminating hyperplane can be found as [43–45],

$$\phi(x) = d \cdot x + c_1 \quad (23)$$

Here, $x, d \in R^2$ and $c \in R$ (R is the set of real number). In (23), the margin distance, d between the support vectors is maximized by the algorithms of SVM by minimizing a cost function [43–45] that can be equated as,

$$\phi(d, \xi) = \frac{1}{2} \|d\|^2 + \Gamma \sum_{z=1}^Z \xi_z \quad (24)$$

In (24), $d^T d = \|d\|^2$ and Γ is a regularization parameter chosen by the users considering the penalty factor of classification errors. Again, ξ_n represents the measurement of error during the training period and Z represents the number of misclassified samples. In this work, the SVM structure was defined with polynomial kernel function with default order value 3. The polynomial kernel SVM can be equated as,

$$K(x, x_i) = 1 + \sum (x, x_i)^d \quad (25)$$

Kernel function, denoted by $K(x, y)$, maps the data to another region of much higher dimensionality.

2.5.3. LDA

LDA is the mostly-used classifier in data classification where the features can be discriminated based on the Fisher's distance. LDA finds a hyperplane that creates linear discrimination between features of two classes. The dimension of the hyperplane is less than one of the number of features representing each class. In the case of classifying two classes, the feature matrix $X_{M \times N}$ is taken. The main target of LDA is to minimize the ratio of within-class variance and maximize the ratio of between-class variance. The between-class matrix, S_B and within-class matrix, S_w of n samples are calculated as [46–47],

$$S_B = \sum_{i=1}^C n_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (26)$$

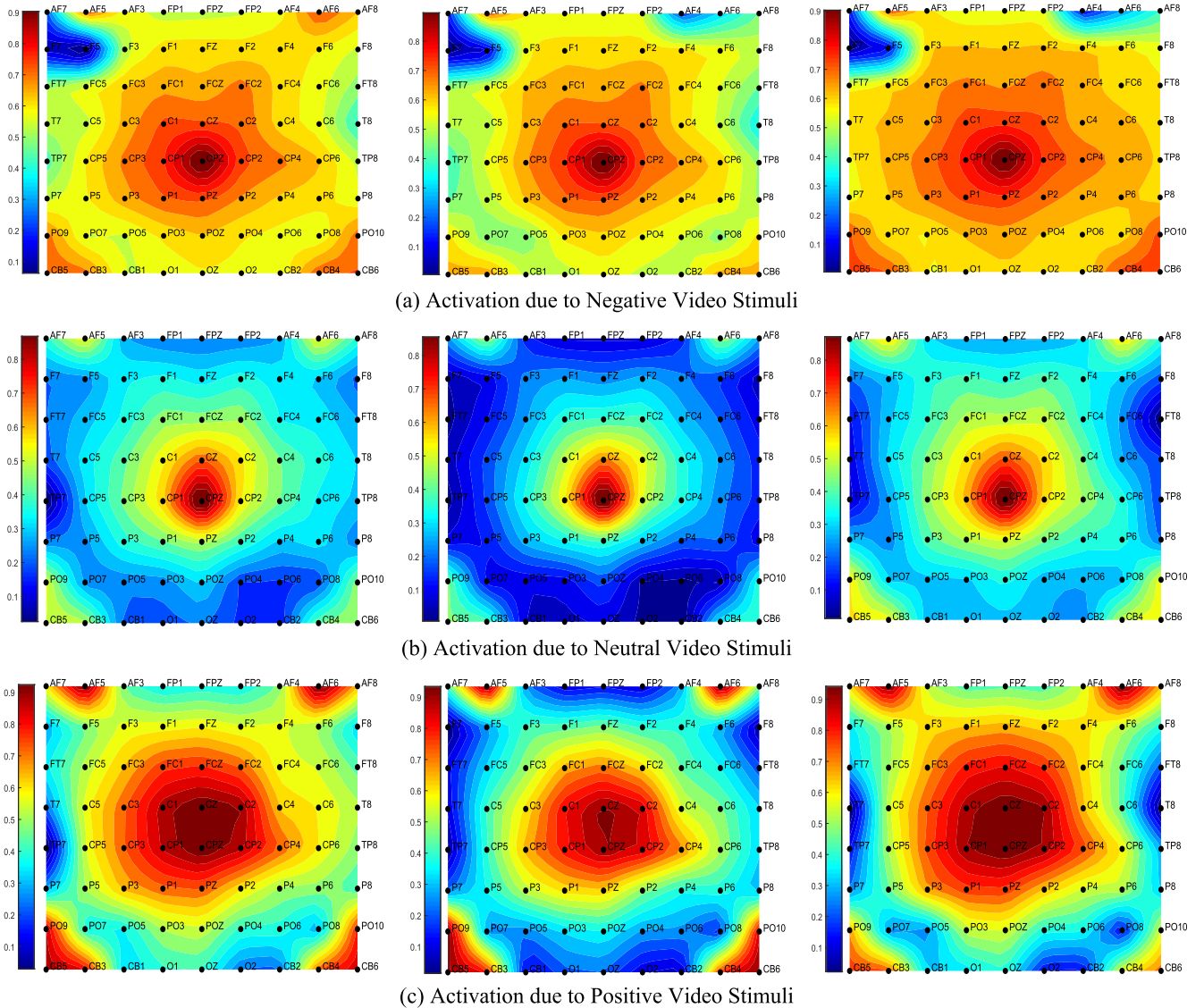


Fig. 7. The activation pattern of the EEG signal regarding negative (a), neutral (b), and (c) positive video stimuli. These topographic images were constructed obeying the method described in [50] with necessary modification where the activation pattern is based on the PSD which was calculated by the Welch method considering 200 samples window along with 50% overlapping.

$$S_w = \sum_{j=1}^C \sum_{i=1}^{n_j} (x_{ij} - \mu_j)(x_{ij} - \mu_j)^T \quad (27)$$

Here x_{ij} is i^{th} sample of j^{th} class and μ_j & μ represents the class mean and total mean, respectively. C is the number of classes.

2.5.4. kNN

To classify the feature set of emotional EEG signals, kNN has been used in this work. The kNN algorithm is a nonparametric estimation method that implements a refinement where the feature environment is high resolution in regions with dense training and low resolution in variance. The method works based on the following steps. Suppose, $Z(z) \subset N$ is a hypersphere of volume, v and center, z . Let us consider that the training set S_k consists of N_k number of samples. Therefore, according to the binomial distribution, the probability of including precisely n samples within the hypersphere, $Z(z)$ can be equated by the following expectation [48–49].

$$E[n] = N_k \int_{y \in Z(z)} p(y|w_k) \approx N_k v p(z|w_k) \quad (28)$$

Here, w_k represents the number of classes. The radius, z of the sphere Z is selected in a way that Z contains exactly κ samples. In this sense, the volume becomes the function of z and it can be presented as $v(z)$. So, the estimation of density based on the number of κ can be calculated as,

$$\hat{p}(z|w_k) = \frac{\kappa}{N_k v(z)} \quad (29)$$

The parameter κ can control the stability between the variance and bias. To classify any vector, \hat{z} the radius of the sphere is selected such a way that the sphere contains exactly κ samples occupied from S_k . These samples are termed as the κ -nearest neighbor.

3. Results & discussions

In a common analysis, it is found that the patterns of activation due to visual emotional stimuli in the EEG signal are different from each other. The activation level is the smallest in case of a neutral emotional evokes like river, birds signing, natural beauty, garden of flower, etc. On the contrary with neutral, the positive and negative emotional videos impact the neural activation in different form. To present such a result, relative power spectral density of the three different emotional EEG signals are given in Fig. 7. From the figure, we can observe that the neutral emotion affects the EEG brain signal with the lowest activation power. On the other hand, the positive and negative emotions activate the brain EEG signal in different pattern.

According to the proposed methodology, after filtering and data dimension reduction some well-established features were extracted from the EEG signals. The features are commonly used in EEG signal classifications [26,34,41,51]. The extracted features from three trials of randomly chosen ten participants were taken into account to find the effective features using the t -statistics. The statistical inference of the positive emotion and negative emotion were tested with neutral emotion. Separately, all features were individually tested using the t -statistics, where the features of individual emotions of all the trials of the selected participants were used to form groups. Therefore, three groups were formed: positive, negative, and neutral. The statistical inference between positive VS neutral and negative VS neutral are given in Tables 1–5.

The resulting statistical inferences of Tables 1–5 suggest that all features are not significant to create statistically discrimination between classes. In addition, all spatial positions of the brain scalp are also not statistically inferable. From the results of Tables 1–5, we observe that in case of Tables 3 and 4, the inferences of the features are not statistically significant. Therefore, the spatial EEG information of central and parietal are not significant to classify the emotional EEG signal. This result suggests excluding the chan-

Table 1

Results of the statistical inference test by students- t distribution for significant feature selection regarding the signal PF_{PCA}.

Serial No	Feature Names	t and p -value (Positive VS Neutral)	t and p -value (Negative VS Neutral)
1	Standard Deviation	$t = 1.0652$; $p = 0.1477$	$t = 1.1462$; $p = 0.1305$
2	Mean Absolute Deviation	$t = 0.9487$; $p = 0.1753$	$t = 2.2127$; $p = 0.0174$
3	Median Absolute Deviation	$t = 1.7451$; $p = 0.0457$	$t = 2.1851$; $p = 0.0185$
4	Max Norm	$t = 2.4172$; $p = 0.0173$	$t = 2.7354$; $p = 0.0052$
5	L1 Norm	$t = 1.3498$; $p = 0.0937$	$t = 1.4788$; $p = 0.0749$
6	L2 Norm	$t = 3.4157$; $p = 0.0009$	$t = 3.7458$; $p = 0.0003$
7	Fractional Dimension	$t = 2.9478$; $p = 0.0031$	$t = 3.0124$; $p = 0.0026$
8	Spectral Norm	$t = 2.7454$; $p = 0.0045$	$t = 2.9841$; $p = 0.0028$
9	Power Spectral Density	$t = 3.0145$; $p = 0.0026$	$t = 3.2451$; $p = 0.0014$
10	Spectral Entropy	$t = 2.7664$; $p = 0.0048$	$t = 3.0041$; $p = 0.0027$

*Degrees of Freedom = 29.

Bold values indicate that features could discriminate between the mental states significantly.

Table 2

Results of the statistical inference test by students- t distribution for significant feature selection regarding the signal F_{PCA}.

Serial No	Feature Names	t and p -value (Positive VS Neutral)	t and p -value (Negative VS Neutral)
1	Standard Deviation	$t = 1.1245$; $p = 0.1350$	$t = 0.8478$; $p = 0.2017$
2	Mean Absolute Deviation	$t = 1.4517$; $p = 0.0786$	$t = 1.1254$; $p = 0.1348$
3	Median Absolute Deviation	$t = 1.6457$; $p = 0.0553$	$t = 2.2115$; $p = 0.0175$
4	Max Norm	$t = 2.0154$; $p = 0.0266$	$t = 2.4187$; $p = 0.0110$
5	L1 Norm	$t = 1.2415$; $p = 0.1121$	$t = 1.2014$; $p = 0.1196$
6	L2 Norm	$t = 2.8415$; $p = 0.0040$	$t = 3.0415$; $p = 0.0024$
7	Fractional Dimension	$t = 2.6415$; $p = 0.0065$	$t = 2.4170$; $p = 0.0110$
8	Spectral Norm	$t = 2.1145$; $p = 0.0215$	$t = 2.4012$; $p = 0.0114$
9	Power Spectral Density	$t = 2.6547$; $p = 0.0063$	$t = 2.2104$; $p = 0.0175$
10	Spectral Entropy	$t = 1.9451$; $p = 0.0307$	$t = 2.1024$; $p = 0.0221$

Bold values indicate that features could discriminate between the mental states significantly.

Table 3Results of the statistical inference test by students-*t* distribution for significant feature selection regarding the signal C_{PCA} .

Serial No	Feature Names	<i>t</i> and <i>p</i> -value (Positive VS Neutral)	<i>t</i> and <i>p</i> -value (Negative VS Neutral)
1	Standard Deviation	<i>t</i> = 1.0652; <i>p</i> = 0.1477	<i>t</i> = 1.1262; <i>p</i> = 0.1405
2	Mean Absolute Deviation	<i>t</i> = 0.8491; <i>p</i> = 0.2013	<i>t</i> = 1.0127; <i>p</i> = 0.1597
3	Median Absolute Deviation	<i>t</i> = 1.2451; <i>p</i> = 0.1115	<i>t</i> = 0.9551; <i>p</i> = 0.1737
4	Max Norm	<i>t</i> = 1.2272; <i>p</i> = 0.1148	<i>t</i> = 2.0454; <i>p</i> = 0.0249
5	L1 Norm	<i>t</i> = 1.4498; <i>p</i> = 0.0789	<i>t</i> = 1.5008; <i>p</i> = 0.0721
6	L2 Norm	<i>t</i> = 2.1157; <i>p</i> = 0.0215	<i>t</i> = 1.7008; <i>p</i> = 0.0498
7	Fractional Dimension	<i>t</i> = 1.9478; <i>p</i> = 0.0305	<i>t</i> = 1.0104; <i>p</i> = 0.1603
8	Spectral Norm	<i>t</i> = 2.5474; <i>p</i> = 0.0082	<i>t</i> = 1.1801; <i>p</i> = 0.1237
9	Power Spectral Density	<i>t</i> = 1.0145; <i>p</i> = 0.1593	<i>t</i> = 1.2401; <i>p</i> = 0.1124
10	Spectral Entropy	<i>t</i> = 2.2632; <i>p</i> = 0.0156	<i>t</i> = 1.0141; <i>p</i> = 0.1594

Bold values indicate that features could discriminate between the mental states significantly.

Table 4Results of the statistical inference test by students-*t* distribution for significant feature selection regarding the signal P_{PCA} .

Serial No	Feature Names	<i>t</i> and <i>p</i> -value (Positive VS Neutral)	<i>t</i> and <i>p</i> -value (Negative VS Neutral)
1	Standard Deviation	<i>t</i> = 1.1152; <i>p</i> = 0.1369	<i>t</i> = 1.1082; <i>p</i> = 0.1384
2	Mean Absolute Deviation	<i>t</i> = 1.2287; <i>p</i> = 0.1145	<i>t</i> = 2.0027; <i>p</i> = 0.0273
3	Median Absolute Deviation	<i>t</i> = 1.2401; <i>p</i> = 0.1124	<i>t</i> = 1.0451; <i>p</i> = 0.1523
4	Max Norm	<i>t</i> = 1.1144; <i>p</i> = 0.1371	<i>t</i> = 0.9584; <i>p</i> = 0.1728
5	L1 Norm	<i>t</i> = 1.2418; <i>p</i> = 0.1121	<i>t</i> = 0.8688; <i>p</i> = 0.1960
6	L2 Norm	<i>t</i> = 2.4157; <i>p</i> = 0.0111	<i>t</i> = 1.7458; <i>p</i> = 0.0457
7	Fractional Dimension	<i>t</i> = 1.9478; <i>p</i> = 0.0305	<i>t</i> = 2.0124; <i>p</i> = 0.0267
8	Spectral Norm	<i>t</i> = 1.3314; <i>p</i> = 0.0967	<i>t</i> = 1.2411; <i>p</i> = 0.1122
9	Power Spectral Density	<i>t</i> = 1.0173; <i>p</i> = 0.1587	<i>t</i> = 1.0101; <i>p</i> = 0.1604
10	Spectral Entropy	<i>t</i> = 2.7664; <i>p</i> = 0.0048	<i>t</i> = 1.2141; <i>p</i> = 0.1172

Bold values indicate that features could discriminate between the mental states significantly.

Table 5Results of the statistical inference test by students-*t* distribution for significant feature selection regarding the signal O_{PCA} .

Serial No	Feature Names	<i>t</i> and <i>p</i> -value (Positive VS Neutral)	<i>t</i> and <i>p</i> -value (Negative VS Neutral)
1	Standard Deviation	<i>t</i> = 1.1748; <i>p</i> = 0.1248	<i>t</i> = 1.0462; <i>p</i> = 0.1520
2	Mean Absolute Deviation	<i>t</i> = 1.0124; <i>p</i> = 0.1598	<i>t</i> = 1.0127; <i>p</i> = 0.1597
3	Median Absolute Deviation	<i>t</i> = 1.8804; <i>p</i> = 0.0350	<i>t</i> = 2.3114; <i>p</i> = 0.0140
4	Max Norm	<i>t</i> = 2.2554; <i>p</i> = 0.0159	<i>t</i> = 2.4781; <i>p</i> = 0.0096
5	L1 Norm	<i>t</i> = 1.4517; <i>p</i> = 0.0786	<i>t</i> = 1.0254; <i>p</i> = 0.1568
6	L2 Norm	<i>t</i> = 1.6614; <i>p</i> = 0.0537	<i>t</i> = 1.3124; <i>p</i> = 0.0998
7	Fractional Dimension	<i>t</i> = 2.8441; <i>p</i> = 0.0040	<i>t</i> = 3.2001; <i>p</i> = 0.0016
8	Spectral Norm	<i>t</i> = 2.3554; <i>p</i> = 0.0127	<i>t</i> = 2.7433; <i>p</i> = 0.0051
9	Power Spectral Density	<i>t</i> = 2.4187; <i>p</i> = 0.0110	<i>t</i> = 2.9541; <i>p</i> = 0.0030
10	Spectral Entropy	<i>t</i> = 2.3544; <i>p</i> = 0.0127	<i>t</i> = 2.4133; <i>p</i> = 0.0111

Bold values indicate that features could discriminate between the mental states significantly.

nels of the central and parietal positions *i.e.*, C_{PCA} and P_{PCA} according to our proposed scheme. In addition, all features are not significant in statistical inference based on the results of *t*-statistics. From the outcomes of Tables 1, 2, and 5, following features were excluded: Standard deviation, Mean absolute deviation, and L1 norm. The rest features were used for training the applied machine learning algorithms. Since four different classification methods were used to classify the emotional states, different results of accuracies were found. In addition, different training and testing ratios were also considered for observing classification performances. In case of training and testing, it is important to consider the appropriate ratio of the training and testing features. To find this limitation, a survey was conducted on the features of randomly selected five subjects. Since there are 15 trials of each subject, we trained the network with 2, 4, 6, 8, 10, and 12 trials and test with the rest trials, *i.e.*, 13, 11, 9, 7, 5, and 3. The resulting classification accuracies regarding different ratio from the selected participants were averaged to observe the effect of the train-test ratio on the classification accuracies. The results are presented in Fig. 8.

From the results in Fig. 8, we find that the arrangement of 10 set features for training and 5 set features for testing yield the highest classification accuracies. Therefore, it can be hypothesized that the 2/3 training data and 1/3 testing data is the best choice for finding the classification accuracies. Inspiring from the results, the predictive model for classification by SVM, ANN, LDA, and kNN were trained and test based on the 3-fold cross-validation approach. The classification accuracies of 14 participants based on the afore-said approach are calculated through the individual classifiers which have been presented in Fig. 9. The average classification accuracies of the different classifiers are given in Fig. 10. From the results, we get that the classification performance of the SVM and ANN are superior to LDA and kNN. Since kNN is a classifier that has no training stage actually, the classification accuracies at the satisfactory level (average > 70% in the proposed method) claim that the proposed method is a robust feature extraction method. This is because the proposed method is able to create sufficient discriminate among the classes which helps to classify at the satisfactory level by a statistical classifier like kNN.

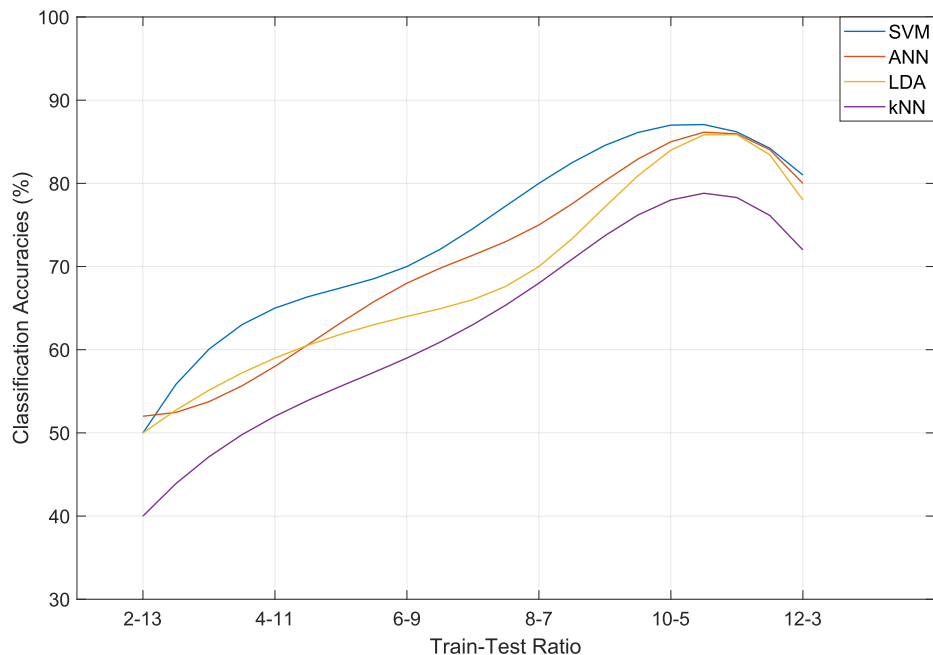


Fig. 8. The variation of the classification accuracies with respect to the ratio of the training and testing data set. The resulting points of accuracies in the figure are connected by spline interpolation. The results are prepared based on the average accuracies regarding different training and the testing ration of the randomly selected five subjects.

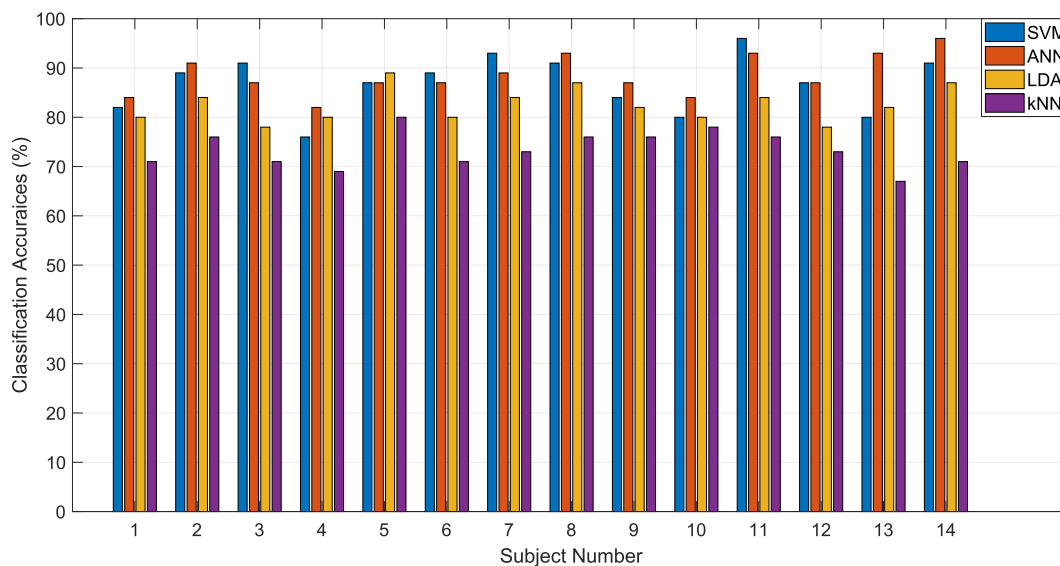


Fig. 9. The classification accuracies of emotional EEG data by the proposed method. Here subject independent classification accuracies are presented for 14 participants with 4 different classifiers.

In addition, the performances of the proposed method have been compared with the other existing method which is given in Table 6. From the comparisons, it has been found that the proposed method yields almost similar result the deep neural networks. In our proposed method, we have used the shallow networks on the extracted features those are outcomes of the combined PCA and *t*-statistics. The outcome of the proposed method is comparable to the outcomes of the existing state of the arts and in some cases, the proposed method outperforms some existing methods.

Since the proposed method is an effective feature extraction way, a similar approach can be used to model the subject independent emotional recognition machine. In this dataset, there are 14 participants and 15 trials of each emotion. There are three classes' emotions- positive, negative, and neutral. According to subject independent approach, there are $14 \times 15 \times 3 = 630$ feature vectors of three emotional EEG data where 310 sets are of each emotion. We have used 2/3 feature vectors to train the classifiers to prepare a model to predict the rest of the feature vectors. So, 140 trials of

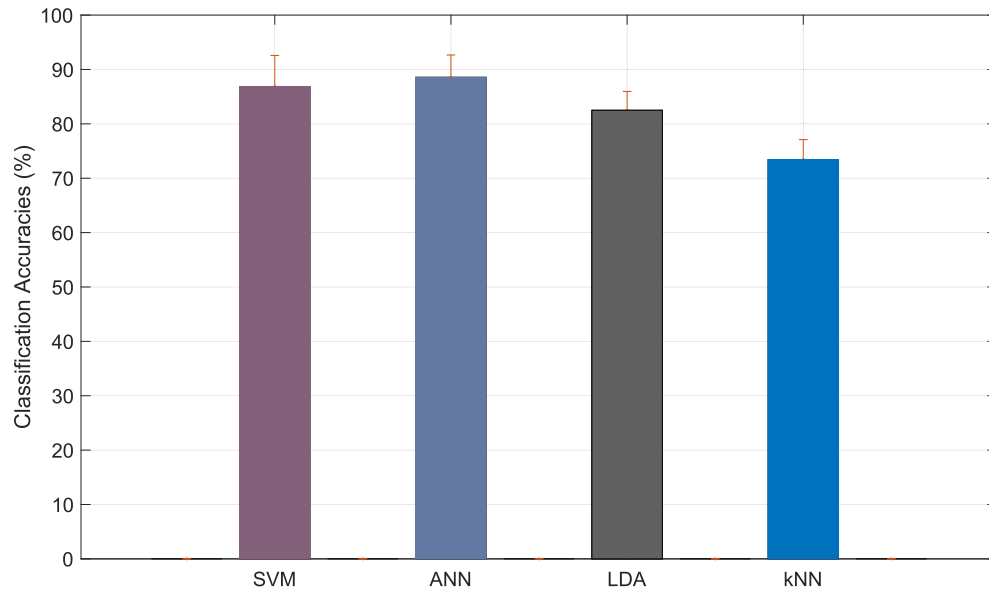


Fig. 10. The average performance of the classification accuracies by different classifiers.

Table 6

Comparison with other methods with the same dataset

Author & Study	Method	Classifier	Average Accuracy
W. Zheng et al. [12]	Group Sparse Canonical Correlation Analysis		86.65%
W. L. Zheng et al. [13]	Critical Frequency Band Investigation	Deep Belief Network	86.08 ± 8.34
W. L. Zheng et al. [14]	Differential Entropy as Features	Discriminative Graph regularized Extreme Learning Machine	79.28%
Y. Yang [19]	Differential Entropy as Features	Hierarchical Network with Subnetwork Nodes	86.42%
Jin et al. [22]	Differential Entropy	Domain Adaptation Network	79.19%
Proposed method	PCA + <i>t</i> -statistics based feature extraction	SVM, ANN, LDA, kNN	85.85 ± 5.72; 86.57 ± 4.08 ; 82.50 ± 3.45; 73.42 ± 3.67

Bold value indicates the highest classification accuracy among the utilized classifiers.

each emotion were used to train and the rest feature vectors (70 for each emotion) were used to test the performance of the predictive models. The performances of individual classifiers according to the subject independent approach are given in Fig. 11. The resulting performances are presented as confusion matrix where individual emotion recognition performances are also presented. The confusion matrix of the classification performance regarding the classifiers- SVM, ANN, LDA, and kNN are given by Fig. 11(a), (b), (c), and (d), respectively. The outcomes suggest that the performance of SVM and ANN is superior to the other two methods- LDA and kNN. So far the outcomes we have found by LDA and kNN are 72.9% and 66.7% which are also in convincing region. On the other hand, the performance of the ANN (84.3%) and SVM (77.1%) is too good to recognize the emotional EEG data at subject independent approach. From the performance, one thing is necessarily important to notice that the performance of the classifiers depends on the classification of positive and neutral data. The classifiers are mostly confused to recognize the negative emotional EEG data. The improvement of the classification accuracies of one classifier to another was achieved by classifying the neutral and positive emotional EEG data whereas, the recognition rate of the negative emotional data are quite similar for every classifier.

4. Conclusion

This paper presents an efficient feature extraction method from the multichannel EEG signal utilizing the combined

approach of the PCA and *t*-statistics. The method has been applied to the three-class emotional EEG data for classification. In average, the classification performances of the proposed method by ANN, SVM, LDA, and kNN are found 86.57, 85.85, 82.50, and 73.42, respectively. It is also found that the proposed method outperforms several existing states of the art. Although some methods that utilized deep neural network showed slightly higher classification accuracies, could be considered as unfeasible because to classify only three class problems utilizing the deep neural network is questionable. In addition, the proposed method was also utilized for emotional EEG data classification as person independent approach and we found 84.3% and 77.1% for ANN and SVM, respectively. Therefore, the proposed method is suitable for the subject independent approach, as well.

Some more features could be considered for the *t*-test. This issue can be regarded as the limitations of this proposed work. Since there are thousands of statistical features in time, frequency, and time–frequency domain, only the frequently used features in EEG analysis were considered. In addition, feature dimensions also affect the classification accuracy. Due to this problem, we considered limited but significant features so that we could achieve the maximum classification accuracy. Although the proposed method has been used solely on the SEED dataset for emotional state classification so far, we hope that this unique idea of feature extraction method can be applicable for EEG signal classification in broad area.

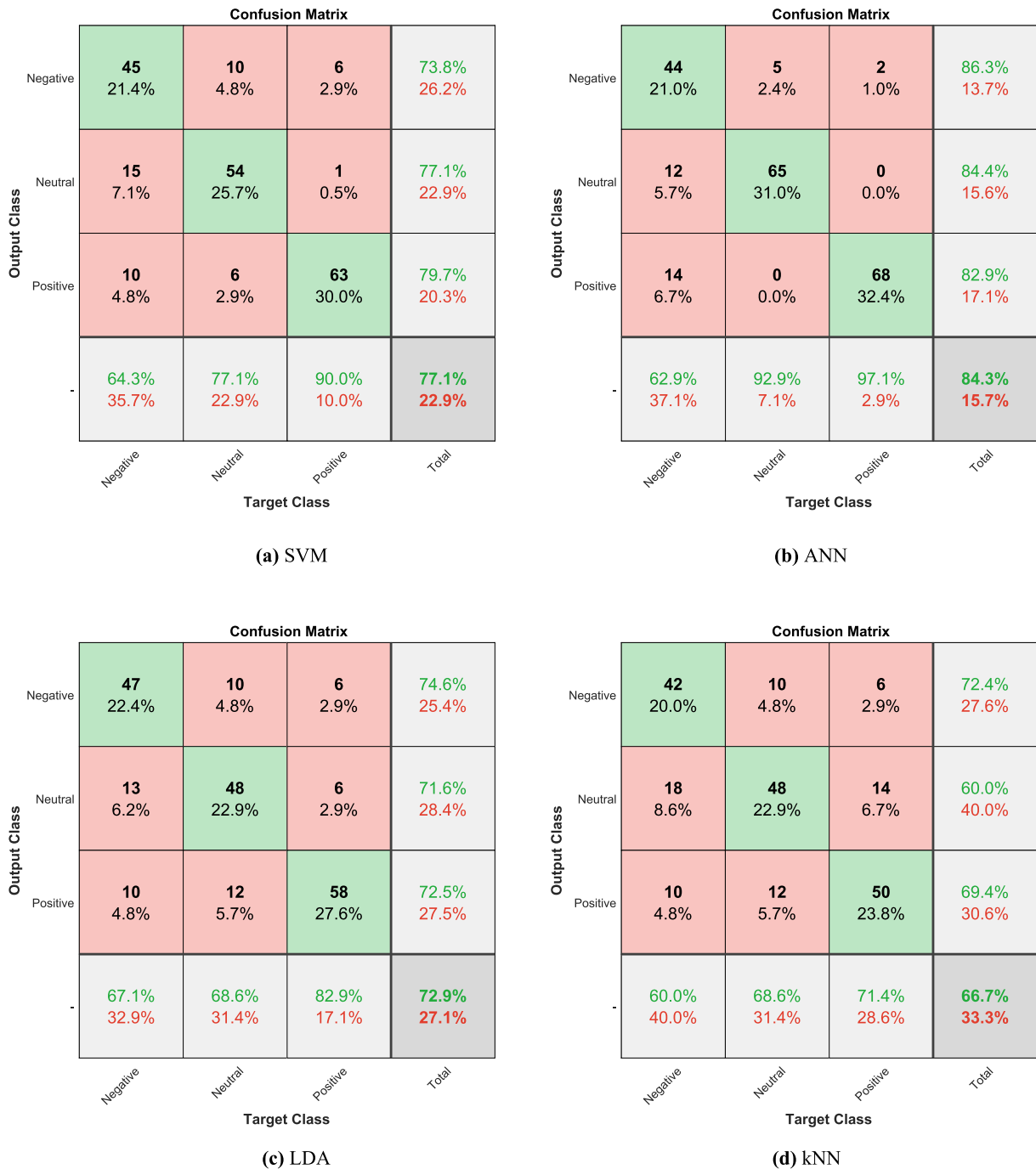


Fig. 11. The performances of the classifiers in emotional EEG signal classification are given by their confusion matrix.

Funding Statement

This research work was not accomplished under any financial support from anybody so far.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] Ajami S, Mahnama A, Abootalebi V. Development of a practical high frequency brain-computer interface based on steady-state visual evoked potentials using a single channel of EEG. *Biocybern Biomed Eng* 2018;38(1):106–14. doi: <https://doi.org/10.1016/j.bbe.2017.10.004>.
- [2] Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: a survey. *IEEE Trans Affective Comput* June 2017. doi: <https://doi.org/10.1109/TAFEC.2017.2714671>.
- [3] Ramirez R, Planas J, Escude N, Mercade J, Farriols C. EEG-based analysis of the emotional effect of music therapy on palliative care cancer patients. *Front Psychol* 2018;9(254).

- [4] Sourina O, Liu Y, Nguyen MK. Real-time EEG-based emotion recognition for music therapy. *J Multimodal User Interface* 2012;5(1–2):27–35. doi: <https://doi.org/10.3389/fpsyg.2018.00254>.
- [5] Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans Intell Transp Syst* 2005;6:156–66. doi: <https://doi.org/10.1109/TITS.2005.848368>.
- [6] Katsis CD, Katertsidis N, Ganiatsas G, Fotiadis DI. Toward emotion recognition in car-racing drivers: a biosignal processing approach. *IEEE Trans Syst Man Cybern* 2008;38:502–12. doi: <https://doi.org/10.1109/TSMCA.2008.918624>.
- [7] Ko BC. A brief review of facial emotion recognition based on visual information. *Sensors* 2018;18(401):pp. doi: <https://doi.org/10.3390/s18020401>.
- [8] Li X, Qi XY, Sun XQ, Xie JL, Fan MD, Kang JN. An improved multi-scale entropy algorithm in emotion EEG features extraction. *J Med Imaging Health Informatics* 2017;7:436–9. doi: <https://doi.org/10.3772/j.issn.1002-0470.2015.10-11.001>.
- [9] Lahane P, Sangaiah AK. An approach to EEG based emotion recognition and classification using kernel density estimation. *Procedia Comput Sci* 2015;48:574–81. doi: <https://doi.org/10.1016/j.procs.2015.04.138>.
- [10] Garcia-Molina G, Tsoneva T, Nijholt A. Emotional brain-computer interfaces. *Int J Autonomous Adaptive Commun Syst* 2013;6(1):9–25. doi: <https://doi.org/10.1504/IJAACS.2013.050687>.
- [11] Murugappan M, Rizon M, Nagarajan R, Yaacob S. Inferring of human emotional states using multichannel EEG. *Eur J Scientific Res* 2010;48:281–99. doi: <https://doi.org/10.4236/jbise.2010.34054>.
- [12] Zheng W. Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. *IEEE Trans Cogn Dev Syst Sept* 2017;9(3):281–90. doi: <https://doi.org/10.1109/TCDS.2016.2587290>.
- [13] Zheng WL, Lu BL. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans Auton Ment Dev* 2015;7(3):162–75. doi: <https://doi.org/10.1109/TAMD.2015.2431497>.
- [14] Jirayucharensak S, Pan-Num S, Israsena P. EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *Sci World J* 2014;2014(627892):1–10. doi: <https://doi.org/10.1155/2014/627892>.
- [15] Ansari K, Chandel G, Pun T. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In: *Proceedings of 15th European Signal Processing Conference*. p. 1241–5. doi: <https://doi.org/10.1109/TAFEC.2007.2712153>.
- [16] DEAP Dataset. Available in : <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/>
- [17] SEED Dataset. Available in: <http://bcmi.sjtu.edu.cn/~seed/download.html>
- [18] MAHNOB Dataset. Available in: <http://www.ibug.doc.ic.ac.uk/resources/mahnob-hci-tagging-database/>
- [19] Yang Y, Wu QMJ, Zheng W, Lu B. EEG-based emotion recognition using hierarchical network with subnetwork nodes. *IEEE Trans Cogn Dev Syst June* 2018;10(2):408–19. doi: <https://doi.org/10.1109/TCDS.2017.2685338>.
- [20] Zheng WL, Zhu JY, Lue BL. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans Affective Comput* 2017. doi: <https://doi.org/10.1109/TAFEC.2017.2712143>.
- [21] Duan R, Zhu J, Lu B. 2013. Differential entropy feature for EEG-based emotion classification. In: *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Diego, CA, 2013, pp. 81–84. doi: 10.1109/NER.2013.6695876
- [22] Jin YM, Luo YD, Zheng WL, Lu BL. EEG-based emotion recognition using adaptation network. *International Conference on Orange Technologies (ICOT)* 2017. doi: <https://doi.org/10.1109/ICOT.2017.8336126>.
- [23] Ahmed R, Rahman MA, Hossain MF. An advanced algorithm combining SVM and ANN classifiers to categorize tumors with position from brain MRI images. *Adv Sci Technol Eng Syst J March* 2018;3(2):40–8. doi: <https://doi.org/10.25046/aj030205>.
- [24] Nazneen T, Rahman MA, Mollah MN. 2019. Towards the effective intrinsic mode functions for motor imagery EEG signal classification. In: *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 7–9 February, 2019, Cox's Bazar, Bangladesh. doi: 10.1109/ECCE.2019.8679120
- [25] Rahman MA, Khanam F, Ahmad M. 2019. Detection of effective temporal window for classification of motor imagery events from prefrontal hemodynamics. In: *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 7–9 February, 2019, Cox's Bazar, Bangladesh. doi: 10.1109/ECCE.2019.8679317
- [26] Rahman MA, Or Rashid MM, Khanam F, Ahmad M. EEG based brain alertness monitoring by statistical and artificial neural network approach. *Int J Adv Comput Sci Appl*. 2019;10(1). doi: <https://doi.org/10.14569/IJACS.2019.0100157>.
- [27] Rahman MA, Ahmad M. 2016. Movement related events classification from functional near infrared spectroscopic signal. In: *Int. Conf. on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 18–20 December, 2016, Dhaka, Bangladesh. doi: 10.1109/ICCITECHN.2016.7860196
- [28] Rahman MA, Milu MMH, Anjum A, Khanam F, Mollah MN, Ahmad M. 2018. Classification of motor imagery events from prefrontal hemodynamics for BCI application. In: *International Joint Conference on Computational Intelligence (IJCCI)*, 14–15 December, 2018, Daffodil International University, Dhaka, Bangladesh. pp. 1–06. doi: 10.1007/978-981-13-7564-4_2
- [29] Mammone N, Morabito FC. Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal. *Entropy* 2014;16(12):6553–72. doi: <https://doi.org/10.3390/e16126553>.
- [30] Mammone N, Foresta FL, Morabito FC. Automatic artifact rejection from multichannel scalp EEG by wavelet ICA. *IEEE Sens J* 2012;12(3):533–42. doi: <https://doi.org/10.1109/JSEN.2011.2115236>.
- [31] Abdi H, Williams LJ. Principal component analysis 2010;vol. 2, July/August. doi: <https://doi.org/10.1002/wics.101>.
- [32] Jolliffe IT. *Principal Component Analysis*. Second Edition. Springer; 2002.
- [33] Bostanov V. BCI competition 2003–data sets Ib and IIb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Trans Biomed Eng* 2004;51(6):1057–61. doi: <https://doi.org/10.1109/TBME.2004.826702>.
- [34] Wang T, Deng J, He B. Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clin Neurophysiol* 2004;115(12):2744–53. doi: <https://doi.org/10.1016/j.clinph.2004.06.022>.
- [35] Reality AI Blog. It's all about the features, September 2017. Available in: <https://reality.ai/it-is-all-about-the-features/>.
- [36] Harne BP. Higuchi fractal dimension analysis of EEG signal before and after OM chanting to observe overall effect on brain. *Int J Electrical Comput Eng (IJECE)* August 2014;4(4):585–92. doi: <https://doi.org/10.11591/ijece.v4i4.5800>.
- [37] Gradshteyn IS, Ryzhik IM. 2000. *Tables of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic Press, p. 1115. ISBN-13: 978-0-12-373637-6
- [38] Horn RA, Johnson CR. "Norms for Vectors and Matrices", *Matrix Analysis*. Cambridge, England: Cambridge University Press; 1990.
- [39] Rieke F, Bialek W, Warland D. *Spikes: Exploring the Neural Code (Computational Neuroscience)*. MIT Press; 1999.
- [40] Millers S, Childers D. 2012. *Probability and random processes* Academic Press. pp. 370–375. ISBN: 9780123870131
- [41] Vakkuri A, Yli-Hankala A, Talja P, Mustola S, Tolvanen-Laakso H, Sampson T, et al. Time-frequency balanced spectral entropy as a measure of anesthetic drug effect in central nervous system during sevoflurane, propofol, and thiopental anesthesia. *Acta Anaesthesiol Scand* 2004;48(2):145–53. doi: <https://doi.org/10.1111/j.0001-5172.2004.00323.x>.
- [42] Haykin S. *Neural Network and Learning Machines*. 3rd Edition. New Jersey: Pearson Education Inc.; 2009.
- [43] Zhang Y, Deng Q, Liang W, Zou X. An efficient feature selection strategy based on multiple support vector machine technology with gene expression data. *Biomed Res Int* 2018;2018(7538204):1–11. doi: <https://doi.org/10.1155/2018/7538204>.
- [44] Shin J, Jeong J. 2014. Multiclass classification of hemodynamic responses for performance improvement of functional near infrared spectroscopy-based brain-computer interface. *J Biomed Optics* 19(6) no. 067009. doi: 10.1117/1.JBO.19.6.067009
- [45] Naseer N, Hong MJ, Hong K-S. Online binary decision decoding using functional near-infrared spectroscopy for the development of brain-computer interface. *Exp Brain Res* 2014;232(2):555–64. doi: <https://doi.org/10.1007/s00221-013-3764-1>.
- [46] Huang S, Cai N, Pacheco PP, Narandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics* 2018;15:41–51. doi: <https://doi.org/10.21873/cgp.20063>.
- [47] Li X, Jiang C, Tang J, Chen Y, Yang D, Chen Z. A Fisher's criterion-based linear discriminant analysis for predicting the critical values of coal and gas outbursts using the initial gas flow in a borehole. *Math Problems Eng* 2017;2017(7189803):1–11. doi: <https://doi.org/10.1155/2017/7189803>.
- [48] Luan L, Wang Y, Li X, Hu W, Li K, Li J, et al. Application of multiple classifier fusion in the discriminant analysis of near infrared spectroscopy for agricultural products. *J Near Infrared Spectrosc* 2016;24:363–72. doi: <https://doi.org/10.1255/jnirs.1236>.
- [49] Heijden FV, Duin RPW, Ridder D, Tax DMJ. Classification, parameter estimation and state estimation: an engineering approach using MATLAB. John Wiley & Sons Ltd. 2004. doi: <https://doi.org/10.1002/0470090154>.
- [50] Li Y, Huang J, Zhou H, Zhong N. Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks. *Appl Sci* 2017;7(1060):1–20. doi: <https://doi.org/10.3390/app7101060>.
- [51] Vazquez RA, Salazar-Varas R. Classification of EEG signals using fractal dimension features and artificial neural networks. *IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI 2017;2017:1–6. doi: <https://doi.org/10.1109/SSCI.2017.8285230>.