



Audio-visual stimulation based emotion classification by correlated EEG channels

Mitul Kumar Ahirwal¹ · Mangesh Ramaji Kose²

Received: 16 April 2019 / Accepted: 10 November 2019 / Published online: 19 December 2019
© IUPESM and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

In this paper, a new channel selection technique is presented for emotion classification by electroencephalography (EEG) signals. Audio-visual stimulation is used to generate emotions at the time of experiment. After recording of EEG signals, feature extraction and classification has been applied to classify the emotions (happy, angry, sad and relaxing). The main highlights of the study include: 1) identification/characterization of audio-visual stimulation which generate harmful emotions and 2) proposed approach to reduce the number of EEG channels for emotion classification. Intention behind identification of audio-visual stimulation (video) responsible for harmful emotions like sad and anger is to control their access over social media and another public platform. EEG channels are selected on the basis of their activation probability, calculated from the correlation matrix of EEG channels. Three types of features are extracted from EEG signals, time domain, frequency domain and entropy based. After feature extraction three different algorithms, support vector machine (SVM), artificial neural network (ANN) and naïve bayes (NB) are used to classify the emotions. This study is conducted over the DEAP (Database for emotion analysis using Physiological signals) database of EEG signals recorded at different emotional states of several subjects. To compare performance after channel selection, parameters like accuracy, average precision and average recall are calculated. After result analysis, ANN is found as best classifier with 97.74% average accuracy. Among listed features, entropy-based features are found as best features with 90.53% average accuracy.

Keywords EEG signals · Emotion classification · Channel selection · Channel correlation · Feature extraction

1 Introduction

Emotional condition plays key role in our lives. Emotion affects the various areas of our daily life, like decision making, learning and communicating with others [1]. Emotion recognition has large application in medical, science, multimedia and many other fields. For example, the emotional response of the patient for his bad health can affect the improvement of health condition, medicines and treatment given to the patient [2]. Also, for developing video games and robots, emotion recognition technique is used to make them more intelligent. Hence, many researchers are focused towards improving the capability of the machine to

recognize the emotion to make it more natural and more intelligent [3]. Emotion can be expressed using various modalities which include facial expression, voice, motion and gesture etc. [4, 5].

The modalities of expressing different emotions have certain limitations. One of the limitations of these modalities is that, they can be faked. For example, a person can fake the facial expression, gesture, voice and motion which makes these modalities less efficient for recognizing the actual emotions [6]. Hence, for the situations like, crime investigation by lie detector, where actual emotions need to be known, these modalities are not effective. To overcome the limitations of these modalities, the researchers have focused on the modalities which cannot be faked [7]. These modalities are the physiological signals that reflect certain changes as a result of change in the emotional condition of subjects. The physiological signals used for emotion recognition includes electroencephalogram (EEG) signal, electrocardiogram (ECG) signal and electromyography (EMG) signal etc. This paper focuses on EEG signal for Emotion Recognition [8].

EEG signal is the result of electrical activity occurs in the brain. To Record the EEG signal, electrodes are placed at certain location over the scalp of the subject. Various electrode

✉ Mitul Kumar Ahirwal
ahirwalmitul@gmail.com; mkahirwal@manit.ac.in

Mangesh Ramaji Kose
mangeshkose@gmail.com

¹ Maulana Azad National Institute of Technology,
Bhopal, MP 462003, India

² National Institute of Technology, Raipur 492010, India

placement systems are available to record the signals [9]. The EEG signals used in this study are recorded using 32 electrode placement system. Position of each electrode is fixed while recording the EEG signal [10]. The change in emotional condition of subject causes activation of a particular electrode or group of electrodes placed at different region of the brain. The pattern of EEG Electrodes activation must be identified correctly to build the connection between emotional condition and electrodes activation. The features are extracted and classification algorithm is applied to categorize the emotions [11]. Literature survey shows that there are many researchers who implemented and applied various classifiers on different types of features extracted for emotion recognition.

Different types of features can be extracted from EEG signals by transforming them in different domains. Each domain has its own type of features, for example, time domain features includes features like line length (L), activity (1stHjorth parameter), root mean square (RMS) amplitude, mobility (2ndHjorth parameter), complexity (3rdHjorth parameter), zero crossings, minima, maxima and nonlinear energy (N). Frequency domain features includes features like peak power, bandwidth (BW), spectral edge frequency (SEF), total spectral power (TP), peak frequency, intensity weighted bandwidth (IWBW) and intensity weighted mean frequency (IWMF) [12]. Entropy based features includes spectral entropy (HS), shannon entropy (HSH) and approximate entropy (HAP) [13]. Correlation between signals obtained from different electrodes is also an important feature that indicates the synchronization between different brain lodes. Studies based on correlation analysis of EEG signals are reported in [5, 14].

Classification is performed after completion of feature extraction. Variety of classifiers have been applied on emotion recognition problem which includes support vector machine (SVM) which works on the concept of finding an optimal hyper plane, K-nearest neighbor (KNN) works the concept of distance between dataset samples, artificial neural network (ANN) uses the concept of interconnected neurons to learn and classify the dataset, naive bayes classifier which is a probabilistic classification algorithm and many more [15–18].

The dimension of the EEG signal used in this work is high because of 32 electrodes. To reduce the dimension, limited numbers of electrodes have to be selected for feature extraction. Electrodes are selected based on common property possessed by them. There are various techniques available in literature for finding common electrodes such as calculating coherence, pearson correlation and cross-correlation between pair of electrodes [19–21]. In this paper, four types of emotions (sad, angry, happy and relaxing) are classified and an approach to select highly correlated and activated EEG channels is developed by identifying the common electrodes based on the new concept of activation probability. Rest of the paper is arranged as; section 2 is dedicated to methodology and materials used. Section 3 and 4 contains the feature extraction and classification, respectively. In section 5, results and observations are presented. Section 6

contains discussion about the work. Finally, section 7 concludes the overall work proposed in this paper.

2 Methodology

The purpose of the methodology followed in this paper is to identify the activation pattern of EEG recording electrodes. It will be easy to classify the emotion using the activation pattern of electrodes. The EEG signals database used here is prepared by recording EEG data with 32 electrodes. The electrodes are numbered from 1 to 32, starting from Fp1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4 and O2 respectively. To record EEG signal the place of each electrode is fixed. Figure 1 shows the placement of 32 electrodes over scalp of a participant to record the EEG signal. The database used here is downloaded from “DEAP: A Database for Emotion Analysis using Physiological Signals” site which contains multichannel dataset for emotion recognition [7]. This dataset contains 32 records of 40 videos, 40 channels EEG recordings recorded from 32 participants while watching 40 one minute long audio-visual stimuli each. Each participant rated each music video according to level of arousal, valence, familiarity, dominance and like/dislike. The overall size of this dataset is $32 \times 40 \times 40 \times 8064$ (participants, No. of videos, No. of channels and data samples). The DEAP dataset is down sampled at 128 Hz. The original dataset contains data samples of 63 s where first 03 s samples are pre-trial baseline that is to be removed [15]. To identify the class label for each video the arousal, valence rating given by each subject to that video is used. By calculating the average of these ratings for each video and by using the arousal-valence 4 quadrant plot the class label for each video is identified, as shown in Fig. 2. The numbers shown in following figure represents the video or trial number.

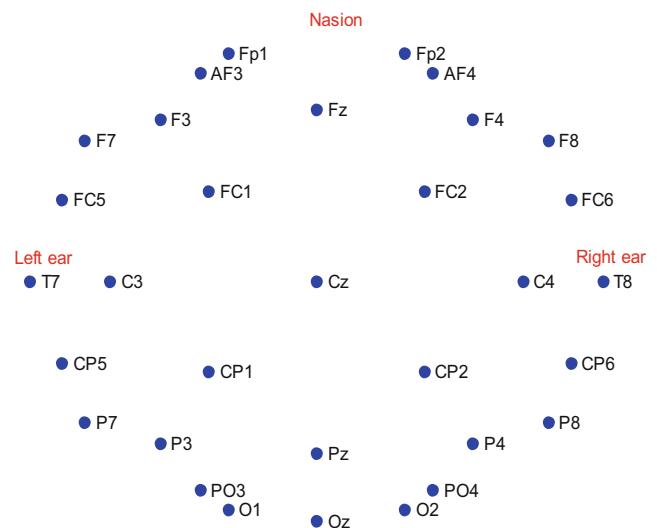
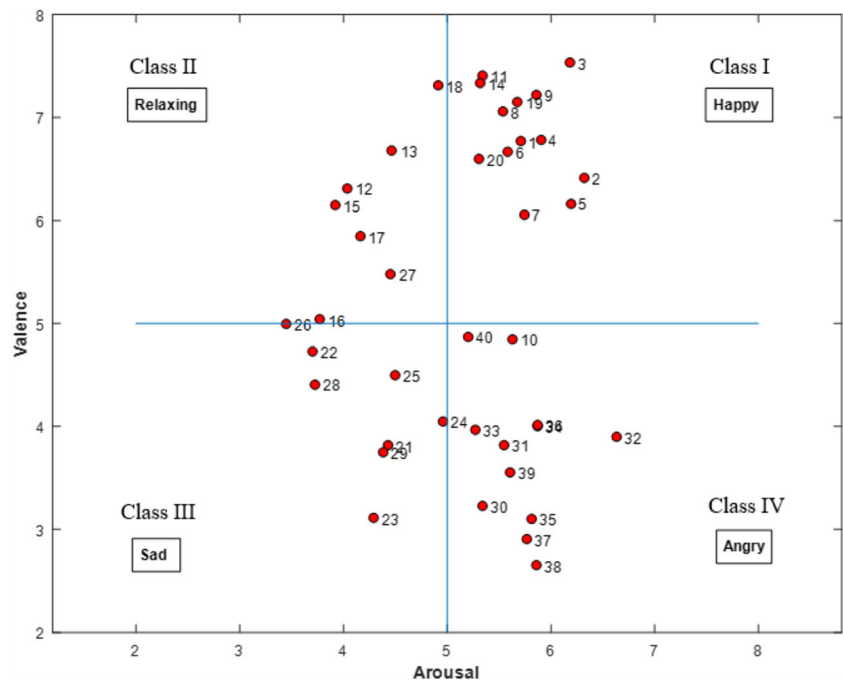


Fig. 1 Electrode placement of 32 channel EEG

Fig. 2 Arousal-valence 4-Quadrant plot



Before using the dataset for this work, the dataset is re-organized to make it music video oriented which was originally subject oriented. The re-organized data consist of 40 records each record consist of one video shown to 32 participants. The original dataset signals were recorded using 40 channels from which first 32 channels are used for EEG signal recording and remaining 8 channels are used for horizontal electrooculogram (hEOG), vertical EOG (vEOG), Zygomaticus EMG (zEMG), trapezius EMG (tEMG), galvanic skin response (GSR), respiration belt, plethysmograph and temperature recording, respectively. In this work electrodes of EEG signal recording are considered (first 32 channels only). The 03 s pre-trial baseline is removed from the 63 s long original signal and recording of only 60 seconds has been used. The overall length of our processed dataset is consisting of $40 \times 32 \times 32 \times 7680$ (No. of videos, No. of participants, No. of channels and No. of Sample). There are 13 videos in happy class, 07 videos in angry class, 08 videos in sad class and 12 videos in relaxing class.

The functional interconnection can be identified by observing the similarities between electrodes recordings. The functional interconnection between each pair of electrodes is identified by calculating pair-wise distance or similarity using some similarity measure. In this paper, correlation as a similarity measure for functional interconnection is used. The correlation among the electrodes which are highly correlated (electrodes correlated at 80 to 100%) are analyzed in this study. Figures 3, 4, 5 and 6 shows the functional interconnection for different classes of emotions. Cases of video no. 01, 12, 21 and 33 are illustrated from happy, relax, sad and angry class, respectively. Connectivity graphs of subject no. 01, 02

and 03 are shown. It is observed that patterns are different for each class of emotions. From these connectivity graphs we can easily identify the pattern of functional interconnection of electrodes. The degree of electrode is also calculated for each electrode, it is the number of other electrodes with which the current electrode is connected. The degree matrix for each video is calculated at the correlation level of 80 to 100%. The degree matrix calculated is further used for calculation of statistical features for selection of electrodes. The mode values are calculated from the degree matrix for 32 subjects who watched that video. Mode value represents the degree of the particular electrode repeated maximum time for that video. The 32 mode values of 32 electrodes for each video belonging to the same class are combined together and then probability of activation of that electrode is calculated for that class of video by eq. (1),

Probability(j^{th} electrode)

$$= \frac{\text{count (No. of videos for which } j^{th} \text{ electrode mode value } > 0)}{\text{Number of videos}} \tag{1}$$

After calculating the probability value of each electrode, for each class of videos, obtained graphs are shown in Fig. 7. From Fig. 7, it is observed that out of 32 electrodes, only four electrodes have positive probability to get activated. Here, these electrodes are CP1, O1, Pz and Po4, numbered as 10, 14, 16 and 31, respectively.

These electrodes have positive probability of activation in all the types of emotions. Hence, the features are extracted from these electrodes to perform the classification. Feature

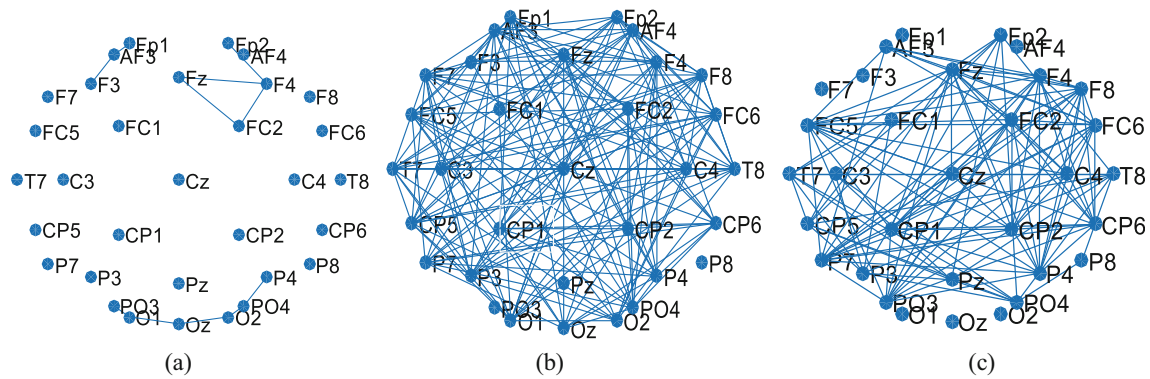


Fig. 3 Functional interconnection among 32 electrodes at 80 to 100% correlation for happy emotion (a) Video 01 Subject 01, (b) Video 01 Subject 02 and (c) Video 01 Subject 03

vector obtained after the completion of aforementioned process is divided into two parts such that 70% instances are used to train the model and remaining 30% instances are used for testing. Figure 8 shows the overall methodology applied in this paper.

3 Feature extraction and Classification

Feature extraction is the key process in any pattern recognition problems, as it is difficult to apply classification on raw dataset. Feature extraction is process of reducing the complexity of original dataset by extracting the attribute values, which will be easily classified by the classifier. Feature extraction techniques vary according to type of dataset on which it is being performed. In this paper, 3 types of features are extracted from EEG signals, time domain features, frequency domain features and entropy based features [12].

3.1 Time domain features

The time domain features of EEG signal are nothing but analysis of EEG signal with respect to time. Here, following time domain features are extracted from EEG signal.

3.1.1 Power (P)

Power of the signal is nothing but its strength. Signal power is calculated as the ratio of sum of square of the signal samples to the number of samples of EEG signal [13]. Power of a signal is calculated as eq.(2),

$$P = \frac{1}{N} \sum_{-\infty}^{\infty} x(t)^2 \quad (2)$$

3.1.2 Line length (L)

Line length is the measure of changes in waveform dimensionality. This measure gets affected by variation of the signal frequency and amplitude. The line length of the signal is calculated as sum of absolute difference between pair of samples [22]. Line Length of a signal is calculated as eq. (3).

$$L = \sum_{t=1}^N \text{abs}[x(t)-x(t-1)] \quad (3)$$

3.1.3 Root mean square (RMS)

As the name suggests RMS is calculated as square root of average of squared samples of an EEG signal [12]. RMS value of a signal is calculated as eq. (4).

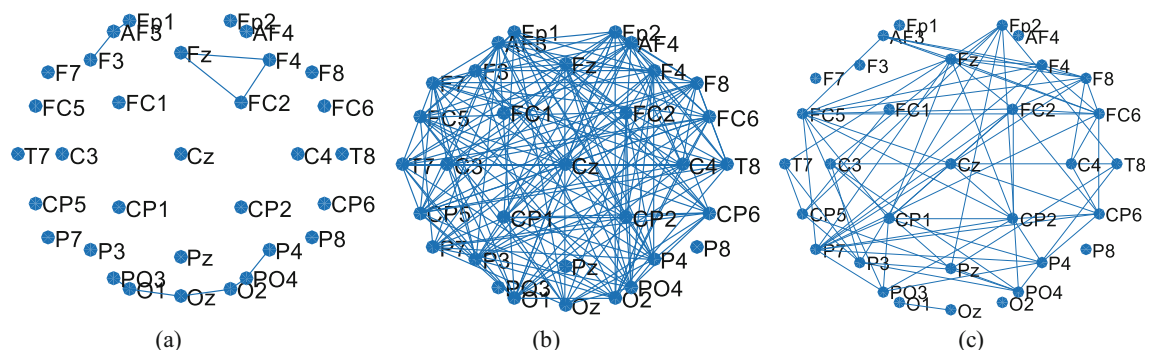


Fig. 4 Functional interconnection among 32 electrodes at 80 to 100% correlation for relax emotion (a) Video 12 Subject 01, (b) Video 12 Subject 02 and (c) Video 12 Subject 03

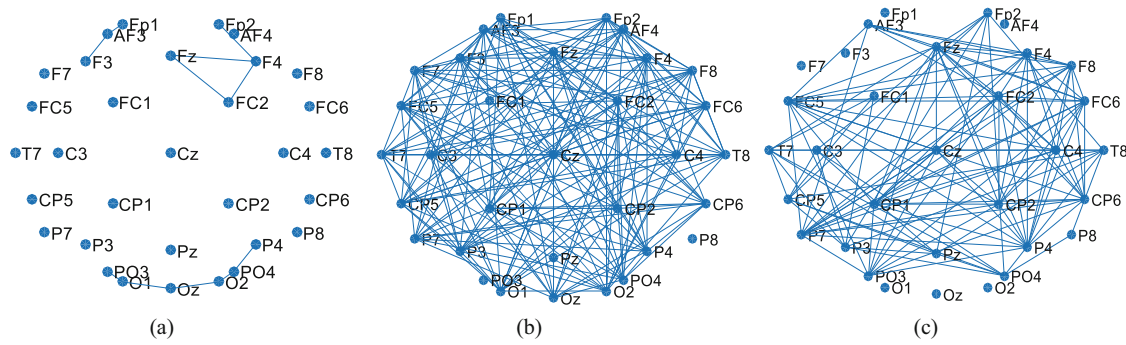


Fig. 5 Functional interconnection among 32 electrodes at 80 to 100% correlation for sad emotion (a) Video 21 Subject 01, (b) Video 21 Subject 02 and (c) Video 21 Subject 03

$$RMS = \sqrt{\frac{1}{N} \sum_{t=1}^N x^2(t)} \tag{4}$$

3.1.4 First difference (D1)

First difference of the signal is calculated as sum of difference between pair of $N-1$ samples upon number of samples-1 [12]. First difference is calculated as eq.(5).

$$D1 = \frac{1}{N-1} \sum_t^{N-1} |x(t+1)-x(t)| \tag{5}$$

3.1.5 Second difference (D2)

Second difference is calculated exactly same as that of first difference but it works on $N-2$ samples rather than $N-1$ sample in first difference [13]. Second difference of a signal is calculated as eq. (6),

$$D2 = \frac{1}{N-2} \sum_{t=1}^{N-2} |x(t+2)-x(t)| \tag{6}$$

Where, N = Total number of EEG samples. $x(t)$ is t^{th} sample of EEG signal.

3.2 Frequency domain features

The frequency domain features of an EEG signals can also be seen as the analysis of EEG signal with respect to time. The signals used in this work are available in time domain. It is necessary to convert time domain signal to frequency domain to extract frequency domain features. Here, fast fourier transform (FFT) is used for time domain to frequency domain conversion of EEG signal. In this paper we have extracted following frequency domain features from the EEG signal.

3.2.1 Dominant /peak power

The peak with largest average power among all the peaks in the spectrum is considered as dominant /peak power [13]. Peak is calculated as eq. (7),

$$Dp = \max\left(\text{abs}(FFT(x))^2\right) \tag{7}$$

3.2.2 Dominant frequency

The frequency corresponding to the dominant/Peak power is defined as dominant frequency [23]. Dominant frequency is calculated as eq. (8).

$$DF = \text{Frequency}\left(\text{Position}\left(\max\left(\text{abs}(FFT(x))^2\right)\right)\right) \tag{8}$$

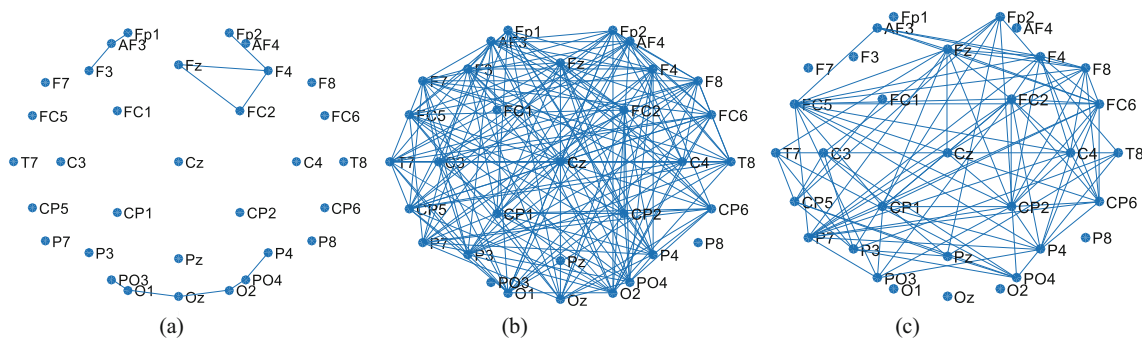
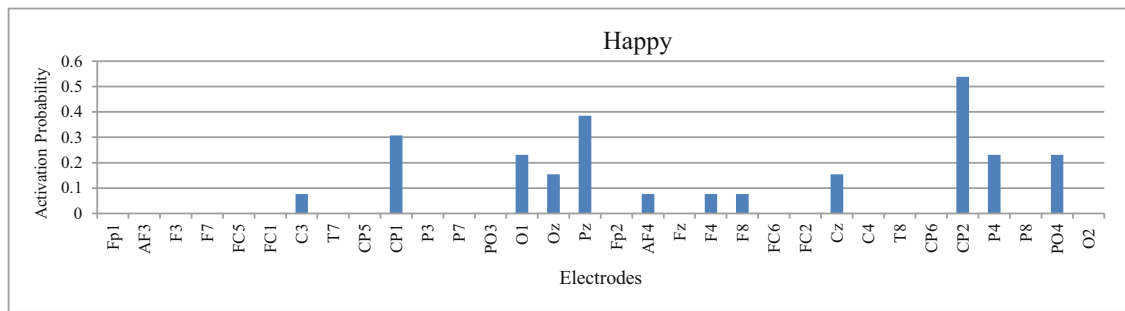
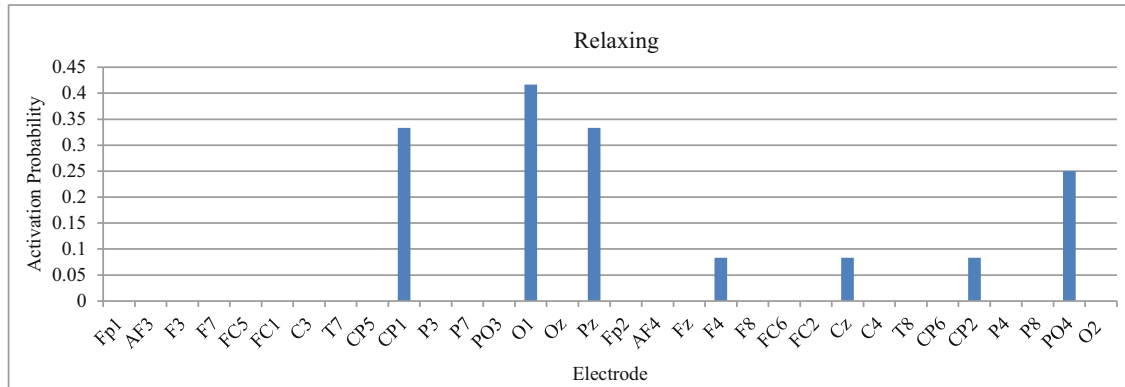


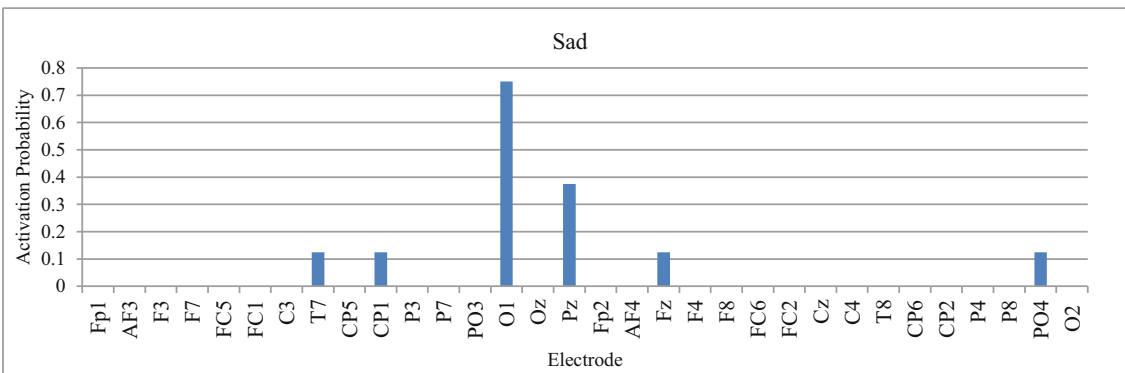
Fig. 6 Functional interconnection among 32 electrodes at 80 to 100% correlation for angry emotion (a) Video 33 Subject 01, (b) Video 33 Subject 02 and (c) Video 33 Subject 03



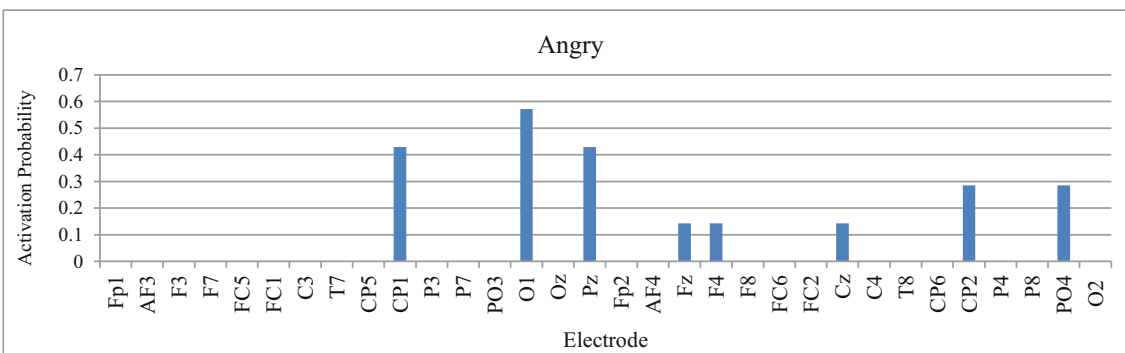
(a)



(b)



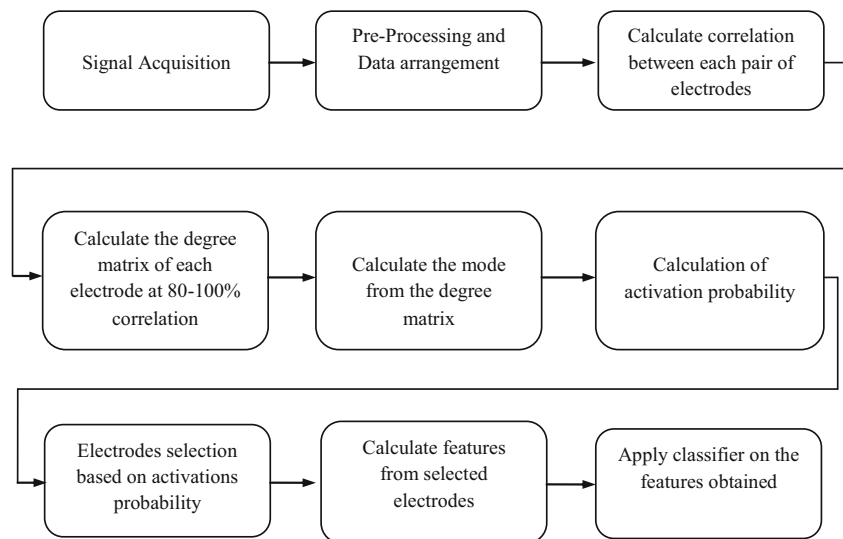
(c)



(d)

Fig. 7 ElectrodeactivationProbability at 80 to 100% correlation, (a) Happy emotion class, (b) Relaxing emotion class, (c) Sad emotion class, (d) Angry emotion class

Fig. 8 Workflow of proposed methodology



3.2.3 Alpha power

The alpha power is calculated from alpha band of EEG signal. The alpha band of EEG signal has frequency range from 8 Hz to 16 Hz. To calculate it we decomposed the original EEG signal using db8 wavelet into 8 levels. Here, we will get 8 detail signals and one approximated signal. The detail coefficient at level 7 i.e. is considered as alpha band signal and then calculated the power of it [24]. Alpha power is calculated as eq. (9).

$$Ap = \sum D7^2 / \text{length}(D7) \tag{9}$$

3.2.4 Delta power

The delta band of the EEG signal is used to calculate the delta power. To calculate delta power the original EEG signal is decomposed into 8 levels using db8 wavelet. Here, we will get 8 detail signals and one approximated signal. The approximated signal is considered as delta band signal and then calculated the power of it [25]. Delta power is calculated as eq. (10).

$$Dp = \sum A^2 / \text{length}(A) \tag{10}$$

3.2.5 Total wavelet energy (TE)

Total wavelet energy is obtained by first decomposing the signal into N levels and then calculating mean energy at each decomposition level (mean energy of detail and approximation coefficients) [26] as given in eq. (11),

$$TE = \left(\frac{\sum_{i=1}^{Nd} DE}{N} \right) + (AE) \tag{11}$$

Where, DE = Detail coefficients energy, AE = Approximation coefficients energy and Nd = Number of detail coefficients.

3.3 Entropy based features

Entropy is the measure of complexity of an EEG signal. Entropy based features have been used by various authors to analyze the behavior of EEG signal during epileptic seizure. In this study we have used following entropy based features of EEG signals.

3.3.1 Spectral entropy

Spectral entropy quantifies the behaviour of the EEG signal, it is calculated using following Equation [26]. Spectral entropy of a signal is calculated as eq. (12).

$$H_S(x) = -\frac{1}{\log N_f} \sum_f P_f(x) \log_e P_f(x) \tag{12}$$

Where, $P_f(x)$ is an estimate of the probability density function. N_f is the number of frequency components in power spectral density (PSD) estimate.

3.3.2 Shannon entropy

An estimate of Shannon entropy (H_{SH}) is obtained by applying histogram estimate of the probability density function, $P_h(X)$ to Shannon's channel entropy formula [26]. Shannon entropy is calculated as eq. (13).

$$H_{SH}(x) = -\sum_f P_h(x) \log_e P_h(x) \tag{13}$$

3.3.3 Sample entropy

It is the modification of approximation entropy used to assess the complexity of timeseries physiological signal [27]. Sample entropy is calculated as eq. (14).

$$S = -\log \frac{A}{B} \tag{14}$$

Where, A is the number of template vector pairs having $d[X_{m+1}(i), X_{m+1}(j)] < r$ of length $m + 1$. B is number of template vector pairs having $d[X_m(i), X_m(j)] < r$ of length m . Generally, the value of m is taken as 2 and value of r is $0.2 \times \text{Standard Deviation}$.

After completion of feature extraction, the next is to classify that feature vector using some classification algorithms.

3.4 Classification

The classification is the process of identifying the class label for the input test feature vector. There are various classifiers available in literature for predicting the target value given the feature vector like decision tree, support vector machine, naive bayes, artificial neural network, K-nearest neighbor etc. From literature, it is found that SVM, ANN and NB are most popular and widely used classification algorithms for performing multiclass classification. Machine learning approaches like SVM outperforms the traditional classifiers in terms of accuracy of classification. The ANN classification algorithm is able achieve maximum accuracy in the presence of noise. Naive Bayes is a probabilistic approach for both binary as well as multiclass classification which uses two different density estimation methods. Considering these factors, many EEG based multiclass prediction systems are designed using SVM, ANN and NB classifiers [9–11, 19]. In this study SVM, ANN and NB classifiers are used.

3.4.1 Support vector machine

Support vector machine is one of the widely used classification technique. This classifier was mainly developed for binary classification problem. The SVM classifier works on the principle of finding an optimal hyper plane that increases the margin between different types of classes. Iterative training algorithm that minimizes error function is used to find an optimal hyper plane [28].

For linear SVM training process includes minimization of error function.

$$\frac{1}{2} W^T W + C \sum_{i=1}^N \xi_i \quad (15)$$

Eq. (15) Subjects to a constraint $y_i(W^T \phi(x_i) + b) \geq 1 - \xi_i$ and $\xi_i \geq 0, i = 1 \dots N$.

Where, C is the capacity constant, W is coefficients vector; b is a constant and ξ_i is slack variable that represents the distance between margin plane $\langle w, x \rangle + b = y_i$ and x_i . The index i label the N training cases. The group category is represented as $y \in \pm 1$ and x_i represents the independent variables. Variable ϕ represents the kernel used to transform data from the input to the feature space. It should be noted that the larger the C , the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

The SVM described above is used just for binary classification purpose. To perform SVM on multiclass classification problems there are two types of algorithm is available in literature, first is one vs all method and another method is one vs one. In this paper ‘one vs all’ algorithm for multiclass classification by SVM is used which is as follows:-

Step 1: During training process we need to create the i models of binary SVM classifier Z_i , such that for i^{th} model output for the features corresponding to class i will be 1 otherwise 0.

Step 2: For the 4 class problem we need to create 4 SVM models i.e. $Z_1 \dots Z_4$, such that for Z_1 , all the features corresponding to class1 are labeled as 1, otherwise 0. Similarly, for Z_2 all the features corresponding to class2 are labeled as 1 and all the other as 0, and so on.

Step 3: During testing process the models created during training process will be tested against each feature vector.

Step 4: Class label is determined by the binary classifier which gives maximum output.

Pseudo Code 1:SVM

```

⇒ Import and shuffle the data
[Divide Data into Train_set (70%) and Test_set (30%)]
[Separate features and classlabels from each set of data]
Train_set = [TrainFeat, Train_lbl]
Test_set = [TestFeat, Test_lbl]
FOR each type of classes  $k = 1$  to 4
    Modellabel = (Train_lbl ==  $k$ )
    [ Labels for training the Model  $M(k)$ ]
     $M(k)$  = Create_trainmodel (TrainFeat, Modellabel)
    [ $M$  contains optimized weight  $w$  and bias  $b$ ]
END FOR
FOR each sample  $i$  in TestFeat
    FOR each type of class  $k$ 
        IF (PREDICTSVM( $M(k)$ , TestFeat $_i$ ))
            [Prediction using trained SVM model]
            Result( $i$ ) =  $k$ 
            [class label for Sample  $i$  is  $k$ ]
        End IF
    End FOR
END FOR
PLOT CONFUSION MATRIX

```



```

Define Create_trainmodel ( $X, Y$ )
[ $X$  =Training Feature,  $Y$  =Label]
Initialize weight vector  $w$  and bias  $b$ 
REPEAT
  [Linear classification function:  $f(x) = (w^T X + b)$ ]
  [Classification function after solving dual problem:  $f(x) = \sum_{i=1}^m \alpha_i y^{(i)} \langle X^i, X \rangle + b$ ]
  Select  $\alpha_i, \alpha_j$ 
  [Where  $\alpha$  is Lagrange multiplier]
  [Compute bounds  $L$  and  $H$ ]
  IF ( $y^i \neq y^j$ )
     $L = \max(0, \alpha_j - \alpha_i)$ ;
     $H = \min(C, C + \alpha_j - \alpha_i)$ 
    [C is regularization parameter initialized to 1]
  ELSE
     $L = \max(0, \alpha_j + \alpha_i - C)$ ;
     $H = \min(C, \alpha_j + \alpha_i)$ 

    
$$\alpha_j = \alpha_j - \frac{y_j(E_i - E_j)}{\eta}$$

  [Update  $\alpha_j$ ]
  END IF
  [Find  $\alpha_j$ ]
  IF ( $\alpha_j > H$ )
     $\alpha_j = H$ 
  ELSE IF ( $L < \alpha_j < H$ )
     $\alpha_j = \alpha_j$ 
  ELSE
     $\alpha_j = L$ 
  END IF
   $\alpha_i = \alpha_i + y^i y^j (\alpha_j^{old} - \alpha_j)$ 
  [Find value of  $\alpha_i$ ]
  [Select threshold]
  IF ( $0 < \alpha_i < C$ )
     $b = b_1$ 
    [where  $b_1 = b - E_i - y^i (\alpha_i - \alpha_i^{old}) \langle X^i, X^i \rangle - y^j (\alpha_j - \alpha_j^{old}) \langle X^i, X^j \rangle$ ]
  ELSE IF ( $0 < \alpha_j < C$ )
     $b = b_2$ 
    [where  $b_2 = b - E_i - y^i (\alpha_i - \alpha_i^{old}) \langle X^i, X^j \rangle - y^j (\alpha_j - \alpha_j^{old}) \langle X^j, X^j \rangle$ ]
  ELSE
     $b = (b_1 + b_2) / 2$ 
  END IF
  [Compute  $w$ ]

  
$$W = \sum_{i=1}^m y_i \alpha_i X_i$$

UNTIL Iterations
return  $w, b, \alpha$ 

```

```

Define PREDICTSVM( $M, XT$ )

```

```

[Here,  $M$  is trained model and  $XT$  testSet features]

```

```

 $f = w * XT + b$ 

```

```

IF ( $f > 0$ )

```

```

  Class = 1

```

```

ELSE

```

```

  Class = 0

```

```

END FOR

```

```

return Class

```

Table 1 Implementation details with parameters of classification methods

Classifiers	SVM	ANN	NB
Parameter	Liner kernel function is used	Scaled Conjugate gradient back-propagation training algorithm is used	Kernel distribution is used
	Penalty parameter (C) with value 1	10 Hidden layer	Training Data = 70%
	SMO hyper plane method is used	Training Data = 65%	Test Data = 30%
	Iterations = 10,0000	Validation data = 5%	
	Training Data = 70%	Test Data = 30%	
	Test Data = 30%		

3.4.2 Artificial neural networks

This is a type of supervised learning algorithm, characterized by network topology, activation function, input to the network and weight vector. The supervised learning algorithm has two phases first is training and second is testing. Here, feed forward neural network with sigmoid output neurons is used. ANN needs to be trained using training feature vector X_r and corresponding target values C_r . After training the ANN, testing is performed. In the testing phase the test vector X_t is given as input to ANN and it has to predict the corresponding class label C_t . The neural network is collection of neurons which are connected by links to communicate information from one neuron to another. The weight value W is assigned to each link of the neural network. The working principle of ANN classifier is to adjust the weights to minimize the error [29].

The Training algorithm of ANN classifier is as follows [30]:-.
 Step 1: Initialize weight vector by some random values of size $m \times n$.

Step 2: Initialize reference vector α .

Step 3: Calculate weighted some of input vector at each layer using eq. (16).

$$D(j) = \sum_{i=1}^n \sum_{j=1}^m (W_{i,j} X_i) \tag{16}$$

Step 4: Apply activation function $f(x)$ to the weighted sum value.

Step 5: Calculate error, $E(i) = Y - f(x)$ at output layer.

Step 6: Change the weight using eq. (17) and eq. (18).

$$\Delta w = f'(x) * E \tag{17}$$

		Predicted Class				
		Class 0	Class 1	Class 2	Class 3	
Actual Class	Class 0	TP for class 0	FN for class 0 FP for class 1	FN for class 0 FP for class 2	FN for class 0 FP for class 3	Total number of samples actually belonging to Class 0. (TP + FN)
	Class 1	FN for class 1 FP for class 0	TP for class 1	FN for class 1 FP for class 2	FN for class 1 FP for class 3	Total number of samples actually belonging to Class 1.(TP + FN)
	Class 2	FN for class 2 FP for class 0	FN for class 2 FP for class 1	TPfor class 2	FN for class 2 FP for class 3	Total number of samples actually belonging to Class 2.(TP + FN)
	Class 3	FN for class 3 FP for class 0	FN for class 3 FP for class 1	FN for class 3 FP for class 2	TPfor class 3	Total number of samples actually belonging to Class 3.(TP + FN)
		Total number of samples Predicted as Class 0. (TP + FP)	Total number of samples Predicted as Class 1. (TP + FP)	Total number of samples Predicted as Class 2. (TP + FP)	Total number of samples Predicted as Class 3. (TP + FP)	

Fig. 9 Structure of Confusion matrix for multi-class classificatio

$$W_{ij}(\text{new}) = W_{ij} + \alpha \times f(x) \times \Delta w \quad (18)$$

Step 7: Repeat step 2 to 6 until stopping condition satisfies.

Here, X =training vector $(x_1, \dots, x_i, \dots, x_n)$, T = class for training vector X , w_j = weight vector for j^{th} output unit. C_j = class associated with the j^{th} output unit. m = number of layers.

Pseudo Code 2: ANN

```

=> Import and shuffle the data
Divide Data into Train_set (60%) , Validation_set(10%) , Test_set (30%);
[separate features and classlabels from Train and Test set of data]
Randomly initialize Weights  $W$  , bias and threshold
Initialize number of Hidden_layers,  $\alpha$ 
[TRAINING ]
REPEAT
  FOR each node  $i$  in layer  $j$ 
     $D_j = \sum_j \sum_i (W_{ji} \cdot X_i)$ 
    [Weighted sum of input Train_set X]
    Output  $_j = f(D)$ 
    [Apply activation function to weighted sum value]
  END FOR
   $E = Y - \text{Output } _j$ 
  [Calculate error at output layer]
   $W(\text{new}) = W(\text{old}) + \alpha * f(D) * E$ 
  [Update weight 'w' in backward direction]
UNTIL stopping condition satisfies

[TESTING]
FOR each Sample  $n$  in Test_Set
  FOR each node  $i$  in layer  $j$ 
     $DT_j = \sum_j \sum_i (W_{ji} \cdot XT_i)$ 
    [Weighted sum of input Test_set XT ]
    Output  $_j = f(DT)$ 
    [Applying activation function  $f$  ]
  END FOR
  Result( $n$ )= label of max (node  $_{i,output}$  )
END FOR
PLOT CONFUSION MATRIX

```

3.5 Naive Bayes

Naive bayes classifier is simple algorithm with probabilistic approach based on the bayes theorem with naive independency among attributes. Naive bayes classifier is also used for text classification problem. In text classification category of new document is identified by estimating the product of probability of each word if the class is given (Likelihood), multiplied by probability of particular class (prior). After calculating above for every class, the class having highest probability will be considered as winner and selected [29]. The concept of Naive bayes classifier is as follows: -.

Bayes theorem provides a path to calculate posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$, as given in the eq. (19),

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (19)$$

Here, $P(c|x)$ is the posterior probability of class (c, target) given predictor (x, attributes). $P(c)$ is the prior probability of class. $P(x|c)$ is the likelihood which is the probability of predictor belongs to the given class. $P(x)$ is the prior probability of feature values (predictor).

Then naive bayes classification algorithm is described as below [31].

Step 1: Let X is a feature sample with n characteristic attributes that is $X = \{x_1, x_2, \dots, x_n\}$.

Step 2: Class labels for all the feature samples are represented by C . There are m feature samples belongs to m class labels respectively, that is $C = \{c_1, c_2, \dots, c_m\}$.

Step 3: For an incoming feature sample X which is to be classified. If $P(c_i|X) > P(c_j|X)$, $i \neq j$, then the feature sample X is considered as it belongs to Class label c_i based on naive bayes.

Step 4: Using bayes theorem, $P(c_i|X) = P(X|c_i)P(c_i)/P(X)$.

Step 5: For all categories, $P(X)$ is constant, so that the MAP $P(c_j|X)$ can be translated into the MPP $P(X|c_j)P(c_j)$.

Step 6: On account of the characteristic attribute of various categories is mutual independence,

$P(X|c_i)P(c_i) = P(x_1|c_i)P(x_2|c_i) \dots P(x_n|c_i)P(c_i) = P(c_i) \prod_{j=1}^n P(x_j|c_i)$, $0 < j < n$.

Pseudo Code 3:NB

⇒ Import and shuffle the data

[Divide data into Train_set (70%) and Test_set (30%)]

[separate features and classlabels from Train and Test set of data]

Calculate probability (PC_i) for each class

FOR each class i

FOR each feature k

$x_i = k^{th}$ feature from i^{th} class samples from training set

$u_i = k^{th}$ feature of all the samples in test set

$DS(i, k).fv = ksdensity(x_i, u_i)$

[Compute probability density estimate]

[length(fv) = length (Test_set)]

END FOR

END FOR

FOR each sample S_N in the test set

$FD(i, k) = DS(i, k).Fv(N)$

[Getting probability estimate value corresponding to sample N]

$P(N, :) = PC_i * prod(FD, 2)$

[Probability of sample N belonging to class i] [size(P) = (N, i)]

Result (N) = Index of max(P)

END FOR

PLOT CONFUSION MATRIX

In Table 1, the details of implementation are provided. This table includes the main parameters with their values of each classification methods used.

4 Results

After the feature extraction process, obtained feature matrix is shuffled, so that samples belonging to different category should be mixed. After this, the complete feature matrix is divided into two parts, first is 70% training set (samples of each emotion category) and second is 30% testing set (used as input samples for testing). On completion of classification process next step is to analyze the performance of each classifier. Confusion matrix is an important tool for accessing the performance of classification algorithm. The structure of confusion matrix for multiclass classification is different from that of binary classification as shown in Fig. 9. The confusion matrix shows the numbers of correctly classified instances as well as numbers of misclassified instances. The confusion matrix consists of 4 terminologies i.e. TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative.

If Class 0 is considered as positive class and Class 1 as negative class then, true positive represents the number of actual Class 0 instances predicted as Class 0. False negative represents the number of actual Class 0 instances misclassified as Class 1. Similarly, False positive represents the numbers of actual Class 1 misclassified as Class 0 instances and true negative represents the Class 1 instances correctly classified as Class 1 instances.

The process of calculating performance parameters using multiclass confusion matrix is discussed below. To calculate the *precision* and *recall* three variables i.e. TP , FP and FN are required. From multiclass confusion matrix, the TP for $class_i$ is the number of samples which are correctly classified as $class_i$ samples, as shown in green box. The FP for $class_i$ is calculated as sum of all the samples which are wrongly predicted as $class_i$ samples. FN for $class_i$ is the sum of all the misclassified samples actually belonging to $class_i$. Considering these variables, the performance parameters from multiclass confusion matrix are calculated as follows.

4.1 Accuracy

It is the measure of classification rate of the classifier. Calculated as ratio of correctly classified class to the total number of classes [32], as given in eq. (20),

$$Accuracy = \frac{\sum_{i=1}^N TP \text{ of } class_i}{total \text{ number of samples}} \quad (20)$$

Where, TP is an abbreviation for true positive. It represents the number of *positive samples* classified correctly.

4.2 Precision [33]

Precision is an important measure to access the performance of classification. It is used to find that, out of total predicted classes as '*class i*' how many are actually *class_i* in terms of

confusion matrix terminology it is calculated as the ratio of sum of precision for each class to the number of classes, as given in eq. (21).

Here, N is the number of class labels.

$$\text{Precision} = \frac{\sum_{i=1}^N \frac{TP_i}{TP_i + FP_i}}{N} \tag{21}$$

4.3 Recall [34]

Recall also called as sensitivity use to measure the performance of classification algorithm. It is use to find that, out of total actual classes how many classes are truly classified, i.e. in terms of confusion matrix terminologies it is calculated as ratio of sum of recall for each class to the number of classes, as given in eq. (22).

$$\text{Recall} = \frac{\sum_{i=1}^N \frac{TP_i}{TP_i + FN_i}}{N} \tag{22}$$

4.4 Kappa index

It is an important metric to evaluate the performance of classifiers. Kappa index is computed by comparing the observed accuracy and expected accuracy [35]. It is calculated using eq. (25).

Kappa Index

$$= \frac{(\text{Observed Accuracy} - \text{Expected Accuracy})}{(1 - \text{Expected Accuracy})} \tag{25}$$

Where, *Observed accuracy* is the accuracy obtained from the classifier as mentioned in eq. (20).

Expected accuracy is calculated using eq. (26).

$$\text{Expected accuracy} = \frac{\sum_{i=1}^N \left[\frac{\text{Actual}(i) * \text{Predicted}(i)}{\text{TotalNumberofinstances}} \right]}{\text{TotalNumberofinstances}} \tag{26}$$

First of all, the results are calculated using all 32 channels and compared with the results of channels selected after correlation analysis-based activations probability calculation. Tables 2, 3 and 4 consists of results of selected channels. Table 2 shows the results calculated for time domain features using the SVM, ANN and NB classifiers, in terms of accuracy, average precision, average recall and kappa index. In same manner Tables 3 and 4 contains the accuracy, average precision, average recall and kappa index value for frequency

Table 2 Performance estimation for Time Domain Features

Classifiers	Accuracy	AVG_Precision	AVG_Recall	Kappa Index
SVM	76.5625	0.7599	0.7679	0.6869
ANN	96.6145	0.9583	0.9599	0.9568
NB	87.5000	0.87507	0.8772	0.8295

domain and entropy features calculated using SVM, ANN and NB classifiers. From Tables 2, 3 and 4, it has been observed that the performance of ANN is better as compared to other classifiers for all types of features.

Table 5 provides comparative analysis of the results calculated using 32 electrodes and results calculated using selected electrodes. The comparison is performed by means of average accuracy value calculated for each type of feature and classifier. As it has been observed from the Table 5 that the results obtained from the selected channels are better than the results obtained from all the channels. It is noticed that the average classification accuracy of ANN is increased up to 97.74% from 93.22%. Improvement is also noticed in case of entropy-based features, 90.53% average classification accuracy has been reached with selected channels that were only 81.16% previously with all channels. From the observation of confusion matrix at the time of testing, it is found that, samples belonging to happy and sad are easily classified as compared to relaxing and angry emotions. Difference of horizontal and vertical sum of each class in confusion matrix is used to identify which class is easy to classify. Happy and sad have less value of difference as compared to sad and angry class.

5 Discussion

This paper presents an approach for emotion classification supported by EEG channel selection methods proposed as activation probability of EEG channels. This will reduce the computation complexity of the large feature vector obtained from all channels. The EEG signal used in this paper is recorded from 32 subjects using 32 electrodes arrangement. Each user/subject has gone through 40 videos of one minute each that are categorized in four classes (Happy, Angry, Sad and Relaxing). In this way the dimension of dataset used for feature extraction is about 32 (subjects) × 32 (channels) × 40

Table 3 Performance estimation for Frequency Domain Features

Classifiers	Accuracy	AVG_Precision	AVG_Recall	Kappa Index
SVM	78.6458	0.7651	0.7628	0.7123
ANN	97.9166	0.9743	0.9747	0.9719
NB	72.3958	0.6881	0.3241	0.6138

Table 4 Performance estimation for Entropy Based Features

Classifiers	Accuracy	AVG_Precision	AVG_Recall	Kappa Index
SVM	83.5937	0.7470	0.7295	0.7750
ANN	98.6979	0.9483	0.9561	0.9820
NB	89.3229	0.7235	0.7429	0.8535

(videos) \times 7680 (EEG data samples). To extract features from this large dataset is tedious job.

The method proposed in this paper aims to reduce the number of electrodes/channels used for feature extraction. For this, each electrode is pair wise analyzed through correlation. Electrodes having correlation of 80 to 100 percentage correlation are used to form functional interconnection graph. The degree matrix of each electrode is calculated that shows how many other electrodes recording is synchronized at the same time. Now, mode value is calculated from degree matrix for each electrode that shows the maximum degree of that electrode among all 32 subjects for same video. Mode values of all videos are combined according to the type of video for which mode is calculated. For example, if video no. 01, 02, 03, 04 and 05 belongs to same class (Happy) then mode values calculated for these videos are combined to form matrix. In this way four mode matrix are obtained for four types of emotions.

Finally, probability of activation of particular electrode for particular type of video is calculated. It is the count of number of videos of that class for which this electrode has positive mode value divided by total number of videos in that particular class. In this way activation probability is calculated for each electrode for each type of emotion classes. The electrodes showing positive probability for all types of emotions are chosen for feature extraction. In this study, four electrodes are found to have activation probability for all types of classes these electrodes are CPI1, O1, Pz and Po4. Hence, features extracted from these electrodes are used for classification purpose.

SVM classifier with linear kernel is used in this study for performing multiclass classification on feature vector

Table 5 Comparative studies of Results calculated using 32 electrodes and results calculated using selected channels

Average Accuracy		
Classifiers	All channels	Selected channels
SVM	69.87	79.60
ANN	93.22	97.74
NB	59.02	83.07
Features		
Time Domain	77.07	86.89
Frequency Domain	63.88	82.98
Entropy Based	81.16	90.53

extracted from EEG signals corresponding to four emotions. The performance of SVM is also dependent to kernel function, for complex classification problem, non-linear kernel SVM are suitable. On the other hand, feed forward neural network having 10 hidden layers with backpropagation algorithm is used for the same task. Complete feature vector obtained after feature extraction has size 1280×14 . The ANN algorithm performs better for such large number of feature vector. The reason behind the poor performance of SVM as compared to ANN in this study is use of liner kernel function. To improve the performance of SVM classifier liner kernel function can be replaced by non-liner kernel function.

Emotion classification over DEAP dataset is also reported in several studies [18, 36–42]. In [18], investigation of time–frequency representation of EEG signals for emotional classification is done. A recent and advanced time-frequency analyzing method called as multivariate synchrosqueezing transform (MSST) is adopted as a feature extraction method. In [36], a multi-model approach for emotion classification is presented. Combination and individual performance of EEG signals, other peripheral physiological signals and eye gaze data is evaluated for emotion classification. Face expressions analysis for emotion classification are also utilized. In [37], a sample entropy based emotion recognition approach was presented. Sample entropy of all channels are calculated, after that channels are selected based on changes observed in sample entropy. In [38], classification of emotions has been attempted with only two channels with the help of higher order spectral analysis (HOSA) and derived features of bispectrum. In [39], feature extraction and selection methods are utilized before the classification process from 32 channels EEG signals. Accuracy of the emotion classification system is improved by utilizing mutual information based feature selection methods known as minimum-Redundancy-Maximum-Relevance (mRMR). In [40], EEG signals (32 channels) and other peripheral physiological signals (13 channels) are used for emotion classification. A multiple-fusion-layer based ensemble classifier of stacked autoencoder (MESAE) is proposed for emotion classification. In [41], classification of emotions is done by applying logistic regression to nonlinear features extracted from EEG signals. Recurrence quantification analysis (RQA) was used to measure the complex dynamics of signals. Comparison between RQA features and spectral features with least absolute shrinkage and selection operator (LASSO), naive bayes, and SVM is presented. As compared to other studies, on an average less number of channels and features are used in present study.

The recent study in [42], suggest the use of selected 04 channels (F3, F4, O1, T4) of EEG based on their experiments performed. In the beginning independent component analysis (ICA) is applied over all four classes of emotions. Various

source separation algorithms are used that are complex in nature. As compared with study in [42], the proposed method is very simply based on just the calculation of probability and also provides more classification accuracy. Simple probability calculation and reduced number of channels with few numbers of extracted is the key difference between present study and above mentioned studies. *Novelty of this study lies in the way of probability calculation for identification of channels that are activated during all emotions.*

Table 6 compares the performance in terms of classification accuracy of existing studies with presented study. Number of channels used and total number of features extracted are also reported in this Table. From Table 6, it can be easily observed that, emotion recognition based on proposed approach achieved maximum classification accuracy among all listed studies.

The social application of this work is visualized as to control the access of video over social media platform that generates harmful emotions like sad and anger. Because this work is capable to rate or tag a movie or advertisements in different categories like family or adult based on the principle of implicit rating. The implicit rating is calculated directly from the EEG signal analysis as EEG signals are true responses of emotions and feelings. The system will put restrictions over vulgarity in movies. The screening of movies can also be done, because they lead to violence, exposure, premature sex and crime, which will ultimately affect today's teenagers and youths.

6 Conclusion

In this paper, the number of channels is reduced by finding the electrodes that plays important role in all the types of

emotion classes. Three types of features are extracted from these selected electrodes for performing classification. Three types of features are time domain feature, frequency domain feature and entropy-based features for classification. SVM, ANN and NB classifiers are used for classification. The performance of the classifiers is calculated for selected electrode features in terms of accuracy, average precision and average recall and also compared with results of all channels. Among all the features, entropy-based features have maximum average accuracy of 90.53%. The average accuracy for SVM, ANN and NB classifier is 79.60%, 97.74% and 83.07% respectively. After analyzing the performance of these classifiers for the features extracted from all the electrodes and the features extracted from selected electrodes, it is observed that the classification accuracy for the features extracted from selected electrodes is better than the features extracted from all the electrodes. Hence, the performance of classification is improved by selected number of channels. Other aspect of this work is helpful in controlling bad media to come in market. The proposed study and development easily identify scenes, which generate unwanted and harmful emotions.

Future scope Further improvement in emotion classification accuracy is achieved by combining different domain of features extracted from EEG signals. Presently, deep learning is also famous for classification of data, feature extraction part of this study can be reduced by applying learning and classification by deep learning techniques. In addition to EEG signal other physiological signals like EOG, EKG, and EMG can be extracted and used for emotion recognition to achieve more accuracy.

Table 6 Comparing performance of proposed approach with existing techniques over DEAP dataset

Authors	Year	Channel used	Total no. of features extracted	Classification Accuracy
M. Soleymani and J. Lichtenauer [36]	2012	32 channels for EEG, 08 channels for other peripheral physiological signals	216 from EEG, 102 from other peripheral physiological signals	52.40% (Arousal) 57.00% (Valence)
X. Jie et al. [37]	2014	08 channels for EEG	08 features	79.11%
N. Kumar et al. [38]	2016	02 channels for EEG	10 features	64.86% (Arousal) 61.17% (Valence)
J. Atkinson and D. Campos [39]	2016	32 channels for EEG	320 features	60.72% (Arousal) 62.39% (Valence)
Zhong Yin et al. [40]	2016	32 channels for EEG, 13 for other peripheral physiological signals	344 from EEG, 81 from other peripheral physiological signals	77.19% (Arousal) 76.17% (Valence)
Ahmet Mert Aydin Akan [18]	2018	32 channels for EEG	500 features	82.11% (Arousal) 82.03% (Valence)
Miaolin Fan and Chun-An Chou [41]	2018	32 channels for EEG	128 features	75.7%
Zangeneh Soroush et al. [42]	2018	04 channels	36 features	90.54%
Proposed approach	–	04 channels	52 features	98.69%

Acknowledgments This research and study is done at Deptt. Of Computer Applications at NIT Raipur CG, under a project entitled “Development of Computational Model for Decision Making based on Emotion Recognition through EEG signal” in file no. ECR/2017/000250, funded by **SCIENCE & ENGINEERING RESEARCH BOARD (SERB)** a statutory body of the Department of Science & Technology, government of India.

Compliance with ethical standards

Conflict of interest Mitul Kumar Ahirwal and Mangesh Ramaji Kose declare that they have no conflict of interest.

Ethical approval This material has not been published in whole or in part elsewhere. The manuscript is not currently being considered for publication in another journal. All authors have been personally and actively involved in substantive work leading to the manuscript.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Cheng Y, Liu GY, Zhang H. The research of EMG signal in emotion recognition based on TS and SBS algorithm. In: Information Sciences and Interaction Sciences (ICIS), 2010 3rd international conference on, 2010, pp. 363–366.
- Krisnandhika B, Faqih A, Pumamasari PD, Kusumoputro B. Emotion recognition system based on EEG signals using relative wavelet energy features and a modified radial basis function neural networks. In: Consumer Electronics and Devices (ICCED), 2017 international conference on, 2017, pp. 50–54.
- Basu S, Bag A, Aftabuddin M, Mahadevappa M, Mukherjee J, Guha R. Effects of emotion on physiological signals. In: IEEE Annu. India Conf., 2016, pp. 1–6.
- Ahirwal MK, Kumar A, Londhe ND, Bikrol H. Scalp connectivity networks for analysis of EEG signal during emotional stimulation. In: Communication and Signal Processing (ICCSP), 2016 international conference on, 2016, pp. 592–596.
- Ahirwal MK, Kose MR. Emotion recognition system based on EEG signal: a comparative study of different features and classifiers. In: International Conference on Computing Methodologies and Communication (ICCMC 2018), Erode, India, pp. XX–XX.
- Gu Y, Tan SL, Wong KJ, Ho MHR, Qu L. Using GA-based feature selection for emotion recognition from physiological signals. In: Intelligent signal processing and communications systems, 2008. ISPACS 2008. International Symposium on, 2009, pp. 1–4.
- Koelstra S, et al. Deap: a database for emotion analysis; using physiological signals. *IEEE Trans Affect Comput.* 2012;3(1):18–31.
- Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans Affect Comput.* 2012;3(1):42–55.
- Xu Y, Hübener I, Seipp AK, Ohly S, David K. From the lab to the real-world: an investigation on the influence of human movement on emotion recognition using physiological signals. In: Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on, 2017, pp. 345–350.
- Kalaivani M, Kalaivani V, Devi VA. Analysis of EEG signal for the detection of brain abnormalities. *Int J Comput Appl.* 2014;1(2):1–6.
- Chen J, Hu B, Wang Y, Dai Y, Yao Y, Zhao S. A three-stage decision framework for multi-subject emotion recognition using physiological signals. In: Bioinformatics and Biomedicine (BIBM), 2016 IEEE international conference on, 2016, pp. 470–474.
- Greene BR, Faul S, Mamane WP, Lightbody G, Korotchkova I, Boylan GB. A comparison of quantitative EEG features for neonatal seizure detection. *Clin Neurophysiol.* 2008;119(6):1248–61.
- Jenke R, Peer A, Buss M. Feature extraction and selection for emotion recognition from EEG. *IEEE Trans Affect Comput.* 2014;5(3):327–39.
- Ahirwal MK, Kumar A, Singh GK, Londhe ND, Suri JS. Scaled correlation analysis of electroencephalography: a new measure of signal influence. *IET Sci Meas Technol.* 2016;10(6):585–96.
- Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: a survey. *IEEE Trans Affect Comput.* 2017;10(3):374–393.
- Becker H, et al. Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources. 2017;3045(1949):1–14.
- Bhakre SK, Bang A. Emotion recognition on the basis of audio signal using Naive Bayes classifier. In: 2016 Int. Conf. Adv. Comput. Commun. Informatics, pp. 2363–2367, 2016.
- Mert A, Akan A. Emotion recognition based on time – frequency distribution of EEG signals using multivariate synchrosqueezing transform. *Digit Signal Process.* 2018;81:106–15.
- Sargolzaei S, Cabrerizo M, Goryawala M, Eddin AS, Adjouadi M. Functional connectivity network based on graph analysis of scalp EEG for epileptic classification. In: Signal Processing in Medicine and Biology Symposium (SPMB), 2013 IEEE, 2013, pp. 1–4.
- Ma M, Li Y, Xu Z, Tang Y, Wang J. Small-world network organization of functional connectivity of EEG gamma oscillation during emotion-related processing. In: Biomedical Engineering and Informatics (BMEI), 2012 5th international conference on, 2012, pp. 597–600.
- Dimitrakopoulos GN, et al. Task-independent mental workload classification based upon common multiband EEG cortical connectivity. *IEEE Trans Neural Syst Rehabil Eng.* 2017;25(11):1940–9.
- Kumar Y, Dewal ML, Anand RS. Relative wavelet energy and wavelet entropy based epileptic brain signals classification. *Biomed Eng Lett.* 2012;2(3):147–57.
- Gotman J, Flanagan D, Zhang J, Rosenblatt B. Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalogr Clin Neurophysiol.* 1997;103(3):356–62.
- Deshprabhu AA, Shenvi N. Sub-band decomposition of EEG signals and Feature Extraction for Epilepsy Classification. 2015;4(3):108–11.
- Blanco-Velasco M, Cruz-Roldán F, Godino-Llorente JI, Blanco-Velasco J, Armiens-Aparicio C, López-Ferreras F. On the use of PRD and CR parameters for ECG compression. *Med Eng Phys.* 2005;27(9):798–802.
- Nguyen HT, Tran H, Vu TT, Bui TTQ. A combination of independent component analysis, relative wavelet energy, and support vector machine for mental state classification. no. Iccas, pp. 0–5, 2016.
- Zhivolupova Y, Tsvetkov O. The method for increasing of EEG signal sample entropy stability and its application for human state monitoring. In: Open Innovations Association (FRUCT), 2017 20th conference of, 2017, pp. 519–525.
- Kumar RG, Kumaraswamy Y. Performance analysis of soft computing techniques for classifying cardiac arrhythmia. *Ind J Comp Sci Eng.* 2014;4:6.
- Bojanić M, Cmojević V, Delić V. Application of neural networks in emotional speech recognition. In: Neural Network Applications in Electrical Engineering (NEUREL), 2012 11th symposium on, 2012, pp. 223–226.
- De Chazal P, Celler BG, Zee M. Selectin \$ a neural network structure for ECG diagnosis. 1998;20(3):1422–5.

31. Ma Y, Liang S, Chen X, Jia C. The approach to detect abnormal access behavior based on naive bayes algorithm. In: *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*, 2016 10th international conference on, 2016, pp. 313–315.
32. Islam SMR, Sajol A, Huang X, Ou KL. Feature extraction and classification of EEG signal for different brain control machine. In: *Electrical Engineering and Information Communication Technology (ICEEICT)*, 2016 3rd international conference on, 2016, pp. 1–6.
33. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag.* 2009;45(4): 427–37.
34. Junker M, Hoch R, Dengel A. On the evaluation of document analysis components by recall, precision, and accuracy. In: *Document Analysis and Recognition, 1999. ICDAR'99. Proceedings of the fifth international conference on, 1999*, pp. 713–716.
35. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20(1):37–46.
36. Soleymani M, Lichtenauer J. A multimodal database for affect recognition and implicit tagging. 2012;3(1):1–14.
37. Jie X, Cao R, Li L. Emotion recognition based on the sample entropy of EEG. *Biomed Mater Eng.* 2014;24(1):1185–92.
38. Kumar N, Khaund K, Hazarika SM. Bispectral analysis of EEG for emotion recognition. *Procedia Comput Sci.* 2016;84:31–5.
39. Atkinson J, Campos D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst Appl.* 2016;47:35–41.
40. Yin Z, Zhao M, Wang Y, Yang J, Zhang J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Comput Methods Prog Biomed.* 2017;140:93–110.
41. Fan M, Chou CA. Recognizing affective state patterns using regularized learning with nonlinear dynamical features of EEG. In: *2018 IEEE EMBS Int. Conf. Biomed. Heal. Informatics, BHI 2018*, vol. 2018–Janua, no. March, pp. 137–140, 2018.
42. Zangeneh Soroush M, Maghooli K, Setarehdan SK, Nasrabadi AM. A novel approach to emotion recognition using local subset feature selection and modified Dempster-Shafer theory. *Behav Brain Funct.* 2018;14(1):1–15.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.