An overlapping sliding window and combined features based emotion recognition system for EEG signals

Shruti Garg, Rahul Kumar Patro, Soumyajit Behera and
Neha Prerna Tigga
Birla Institute of Technology, Ranchi, India, and
Ranjita Pandey
University of Delhi, New Delhi, India

Abstract

Purpose – The purpose of this study is to propose an alternative efficient 3D emotion recognition model for variable-length electroencephalogram (EEG) data.

Design/methodology/approach – Classical AMIGOS data set which comprises of multimodal records of varying lengths on mood, personality and other physiological aspects on emotional response is used for empirical assessment of the proposed overlapping sliding window (OSW) modelling framework. Two features are extracted using Fourier and Wavelet transforms: normalised band power (NBP) and normalised wavelet energy (NWE), respectively. The arousal, valence and dominance (AVD) emotions are predicted using one-dimension (1D) and two-dimensional (2D) convolution neural network (CNN) for both single and combined features.

Findings – The two-dimensional convolution neural network (2D CNN) outcomes on EEG signals of AMIGOS data set are observed to yield the highest accuracy, that is 96.63%, 95.87% and 96.30% for AVD, respectively, which is evidenced to be at least 6% higher as compared to the other available competitive approaches.

Originality/value – The present work is focussed on the less explored, complex AMIGOS (2018) data set which is imbalanced and of variable length. EEG emotion recognition-based work is widely available on simpler data sets. The following are the challenges of the AMIGOS data set addressed in the present work: handling of tensor form data; proposing an efficient method for generating sufficient equal-length samples corresponding to imbalanced and variable-length data.; selecting a suitable machine learning/deep learning model; improving the accuracy of the applied model.

Keywords Electroencephalography (EEG), Emotion recognition (ER),
1D and 2D convolution neural network (CNN)

Paper type Research paper

1. Introduction

Emotions are a manifestation of intuitive states of the mind. They are known to be generated by events occurring in a person's environment or internally generated by thoughts [1]. Identification and classification of these emotions using computers have been widely studied under affective computing and human–computer interface [2].

Emotions are recognised using physiological or non-physiological signals [3]. Electroencephalogram (EEG), electrocardiogram (ECG) [4], galvanic skin response (GSR), blood volume pulse (BVP) [5] and respiratory suspended particulate (RSP) [6] are popular

tools used in literature to obtain physiological signals, while facial expressions [7], speech [8], body gestures and videos [9] give non-physiological signals. The advantage of using the physiological signals for ER is that they are directly captured from human body which gives true response of human intuitions [10] unlike non-physiological signals that can be synthetically elicited. Thus, the EEG signals are suitable tool for current research. However, since EEG signals involve studying human behaviour directly, there is a limitation to the number of samples that can be collected while deep learning (DL) methods required large number of samples to work efficiently. Therefore, there is a need for innovative resampling method to be able to apply DL methods.

The EEG signals are generated by electrical waves corresponding to brain activity presented by external stimuli [11]. The raw signals need to be pre-processed, and then appropriate features need to be extracted to get emotions from the signals. Lastly, an efficient classifier is applied to obtain an appropriate recognition of emotions.

The features of EEG signals were frequently extracted in the time, frequency and time–frequency domains. The features extracted in the time domain are the Hjorth feature [12], fractal dimension feature [13] and higher-order crossing feature [14]. The features used in the frequency domain are power spectral density (PSD) [15], spectral entropy (SE) [16] and differential entropy [17]. Wavelets and a short-time Fourier transform (STFT) [18] have been used to extract the time–frequency domain features.

After feature extraction, the machine learning (ML) and DL methods are primarily applied in literature for classification [19]. The ML methods applied are k-nearest neighbour (KNN), random forest (RF), decision tree (DT), neural network (NN) and support vector machine (SVM) for ER. The DL methods used for ER are convolution neural network (CNN), long short-term memory (LSTM), recurrent neural network (RNN) and several other variants. The DL methods are found to work with greater accuracy [20]. Table 1 shows a summary of DL methods applied in recent years.

Apart from these, nature-inspired algorithms have also been applied on ER tasks for feature selection, such as on the DEAP data set along with particle swarm optimisation (PSO) [21] and firefly optimisation (FO) [30]. At the same time, LSTM and SVM were used as classifiers. Feature selection through FO has been known to achieve an accuracy of 86.90%, while PSO feature selection recorded an accuracy of 84.16%.

Emotions in ER can be classified in two ways: discrete emotions, such as anger, happiness, sadness, disgust, fear and neutral, and emotion models. There are two types of emotion models: two-dimensional (2D) [31] and three-dimensional (3D) [32]. The 2D emotion model consists of valence and arousal; valence represents the measure of pleasant and unpleasant, and arousal represents excitement and calmness. The 3D emotion model comprises AVD. The arousal and valence emotions are the same as in the 2D emotion model. Dominance is the third emotional aspect, representing dependence and independence.

1.1 Contribution

The objective of the present work is to develop an efficient ER model for the AMIGOS [33] data set in 3D emotional space (i.e. AVD) using DL models. The AMIGOS is a new data set among other popular EEG data sets for ER. The following are the challenges of the AMIGOS data set addressed in the present work:

(1) Handling of tensor form data.

(2) Proposing an efficient method for generating sufficient equal-length samples corresponding to imbalanced and variable-length data.

(3) Selecting a suitable ML/DL model.

(4) Improving the accuracy of the applied model.

Ref., year	Emotions recognised	Feature extraction method	Classifier	Data sets	Accuracy %
[21], 2020	2D-emotion model	High-order statistics	LSTM	SEED	90.81
[22], 2020	Negative, positive and neutral	Electrode frequency distribution map + STFT	CNN	SEED, DEAP	90.59
[23], 2020	3D emotion model	Multi-level feature capsule network (end to end network)	Multi-level feature capsule network	DEAP, DREAMER	98.32
[24], 2020	Negative, positive and neutral	Local and global inter channel relation	Regularised graph neural network	SEED, SEED-IV	85.30
[25], 2021	2D emotion model	Differential entropy	Graph convolution Network + LSTM	DEAP	90.60
[26], 2020	Sad, happy, relax, fear	Time frequency representation by smoothed Pseudo-Wigner–Ville distribution	Configurable CNN, Alexnet, VGG-16, Resnet-50	Recorded EEG of students of Indian Institute of information Technology design and Manufacturing, Jabalpur	93.01
[27], 2020	2D emotion model	End-to-end region asymmetric convolution neural network	Region asymmetric convolution neural network	DEAP, DREAMER	95
[28], 2020	2D emotion model	Spectrogram representation	Bidirection LSTM	AMIGOS	83.30
[29], 2021	2D emotion model	Features extracted from topographic and holographic feature map	CNN + SVM	AMIGOS	90.54

Table 1.
Studies on EEG-based emotion recognition using deep learning

The equal-length data samples are generated here by the OSW method. Although the data can be oversampled using the built-in Python function Synthetic Minority Oversampling Technique (SMOTE) [34], SMOTE generates the data by replicating the examples without adding any new information to them. Thus, the OSW method is proposed in the present work, which induces variability in the sample records by avoiding the repetition of the signals. Feature extraction is undertaken in two modes using normalised transformation of band power and a wavelet energy.

The rest of this paper comprises three additional sections. Section 2 provides details of the emotion recognition system proposed in the research, Section 3 details the results and discussions and Section 4 provides the conclusions.

2. Emotion recognition system

The proposed emotion recognition system (ERS) is modelled in three stages:

(1) Data preprocessing,

(2) Feature extraction and

(3) Classification implemented for AVD.

Figure 1 shows the framework adopted for OSW-based ERS.

The important concepts used in present research are described as follows:

2.1 Decomposition of signal using OSW
The emotion samples are amplified in current research using OSW, as a large amount of data is recommended for efficient model building in DL methods [35]. EEG signals produced in different experiments were decomposed into 512 size windows by a shift of 32, as shown in Figure 2.

The portion of signals not covered by the 512 windows was trimmed or not used for computation purposes. The window and shift were decided experimentally.

2.2 Feature extraction
Once signals were decomposed into equal-length samples using overlapping windows, NBP and NWE features were extracted using discrete Fourier transform (DFT) [36] and discrete wavelet transform (DWT) [37].

2.2.1 Normalised band power (NBP). To calculate NBP feature, first Fourier transform X_k was calculated for windowed signal by using Eqn (1):

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi nk/N} \tag{1}$$

where N is length of vector x and $0 \leq k \leq N - 1$.

Once signal is converted to frequency domain, the five frequency bands (4–8 Hz, 8–13 Hz, 13–16 Hz, 16–30 Hz and 30–45 Hz) were extracted. The beta band was decomposed into two (beta1 and beta2) to equalise the dimensions of the wavelet transform. Further, band power and normalised band power were then calculated for each band by Eqns (2) and (3) given below:

$$P_B = \sum_0^k |X_k|^2 \tag{2}$$

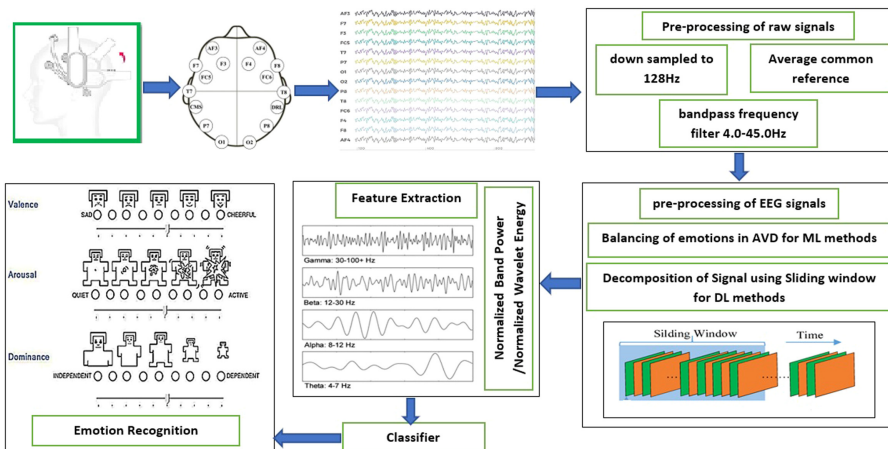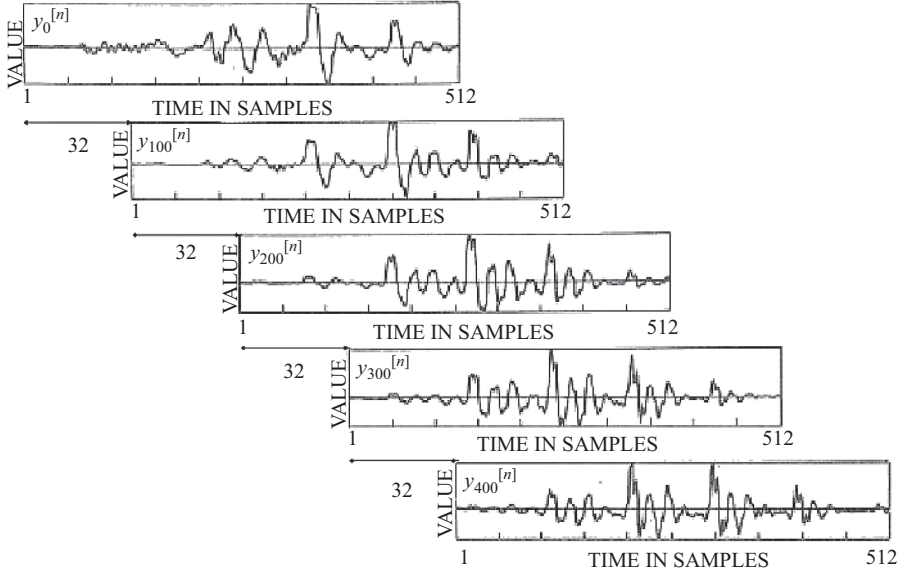where P_B represents power of band B, k is length of each band.

Figure 2.
Overlapping window
signal decomposition

$$\widehat{P}_B = \frac{P_B}{\sum_B P_B} \tag{3}$$

where, $\widehat{P}_B$ is called NBP.

2.2.2 Normalised wavelet energy (NWE). In DWT, the different frequencies of signal are cut at different levels, and the process is called multi-level wavelet transform, defined in Eqn (4)

$$D(\tau, s) = \frac{1}{\sqrt{s}} \sum_{n=0}^{p-1} x_n . \psi\left(\frac{t - \tau}{s}\right) \tag{4}$$

where $\tau = k.2^{-j}$ and $s = 2^{-j}$ represents translation and scale respectively. ψ is called mother wavelet which was taken here Daubechies4(db4) wavelet. The signal is further decomposed into cA_n and cD_n which are called approximation coefficient at nth level (provides low frequencies) and detailed coefficient at nth level (provides high frequencies), respectively. Because the EEG signal provided in the pre-processed data set is in the range of 4 Hz–45 Hz, five-level decomposition is sufficient for required four-band information, as shown in Figure 3.

After decomposition of signal into multilevel wavelet coefficients, the wavelet energy is calculated using detailed coefficients cD_n of above five levels because the emotion information is mostly available in higher frequencies. The formula for calculating wavelet energy is given in Eqn (5):

$$WE_n = \sum_n |cD_n|^2 \tag{5}$$

NWE is calculated using Eqn (6)

$$\widehat{WE}_n = \frac{WE}{\sum_n WE_n} \tag{6}$$

EEG signal
4-45 Hz

cA1
4-24.5Hz

cD1
24.5-45Hz

cD1 Upper
Beta and
Gamma

cA2
4-14Hz

cD2
14-24Hz

cD2 Beta

cA3
4-9 Hz

cD3
9-14Hz

cD3 Upper
Alpha and
Lower Beta

cA4
4-6.5 Hz

cD4
6.5-9 Hz

cD4 Lower
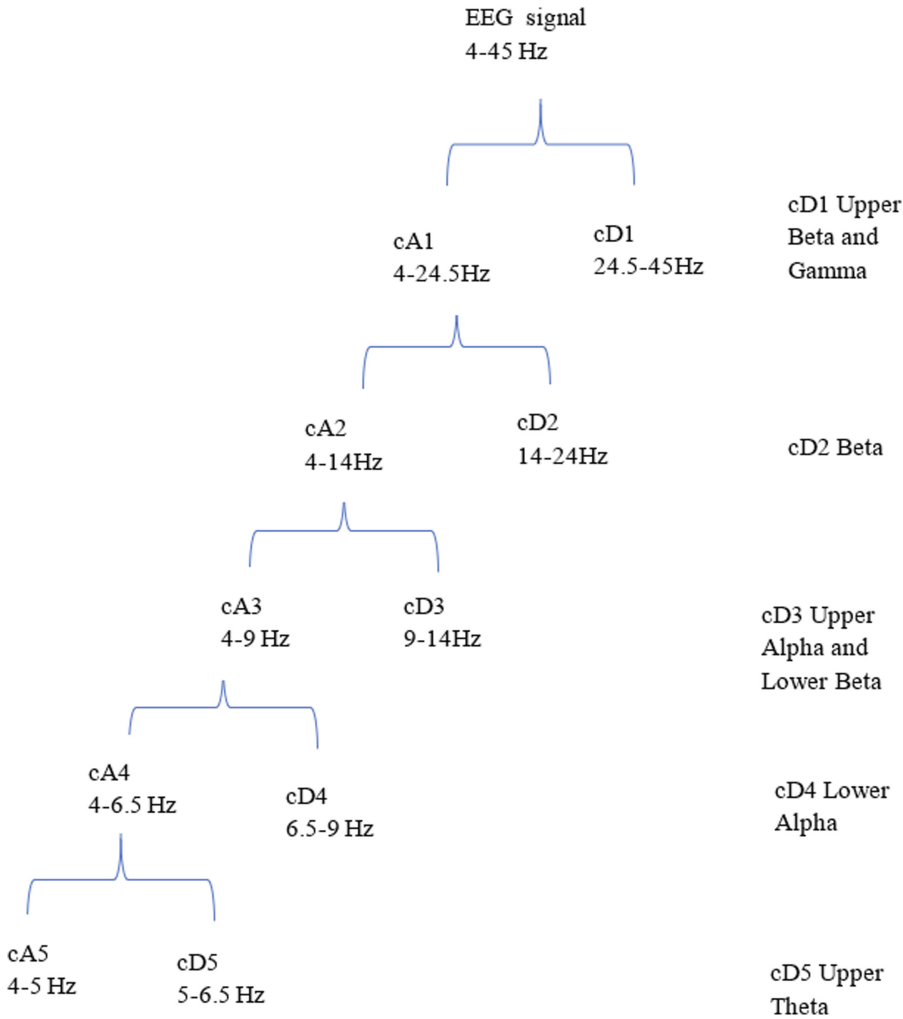Alpha

cA5
4-5 Hz

cD5
5-6.5 Hz

cD5 Upper
Theta

Figure 3.
Wavelet
decomposition of
different bands

2.3 Convolution neural network
A CNN is multilayer structure that consists of different types of layers, including input, convolution, pooling, fully connected, softmax/logistic and output [38]. The extracted features are fed into two different types of CNN: 1D and 2D. Both 1D and 2D CNN followed same architectures, convolutions, including Conv1D and Conv2D, preceded by batch normalisation. A max pooling layer with ReLU activation function is applied to every convolution layer. Lastly, the max pooling layer is connected to an adaptive average pooling layer, which is then passed through a flattening layer and followed by four output dense layers. The first three are dense linear layers, and the last is a sigmoid layer for binary classification among four output dense layers. Architecture of 1D and 2D CNN is shown in Figures 4 and 5, respectively.
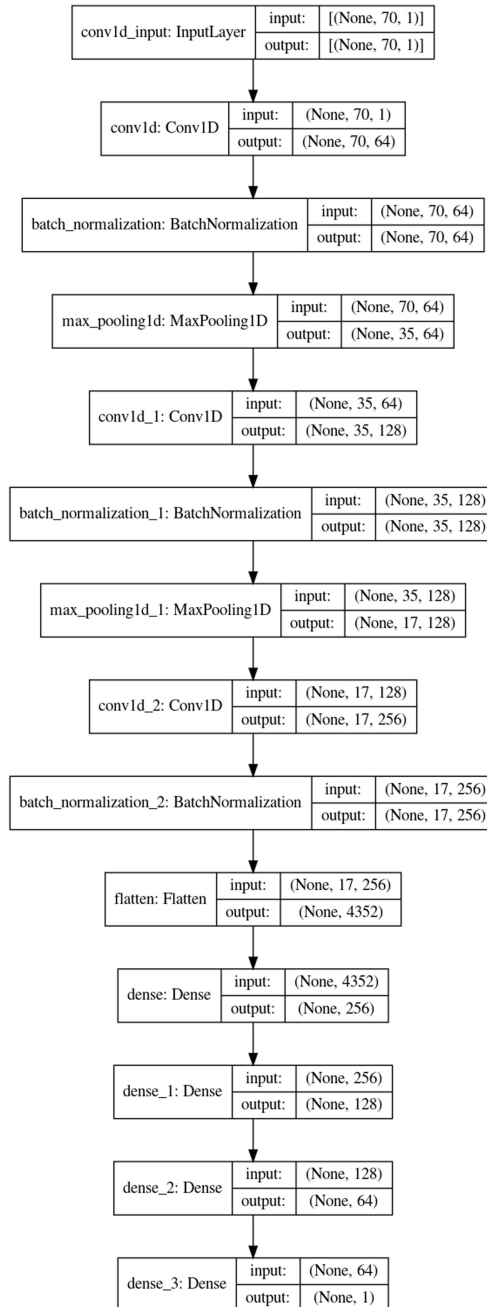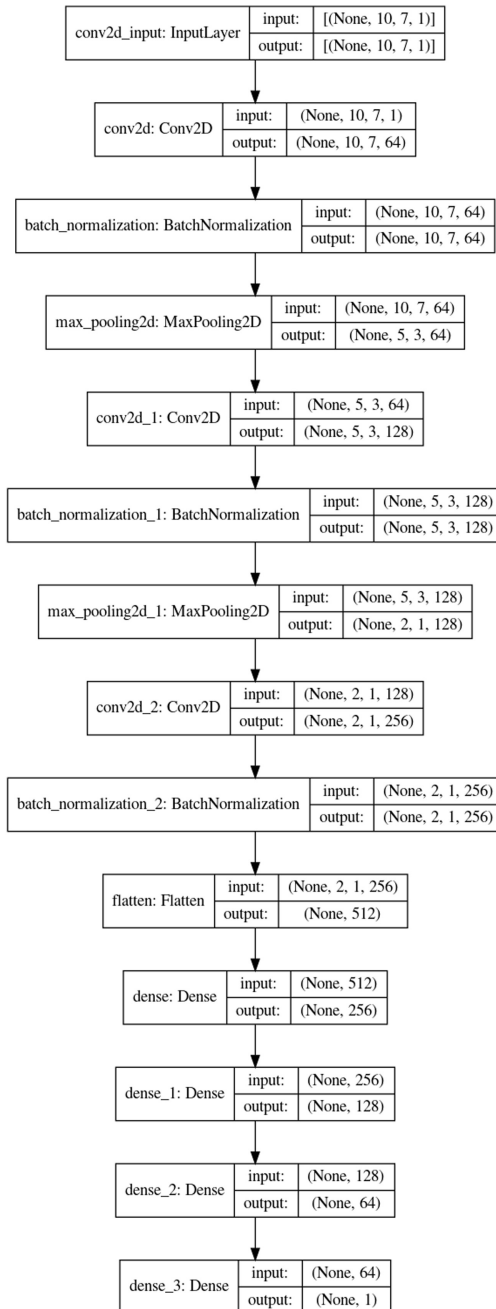
conv1d_input: InputLayer

input:	[(None, 70, 1)]
output:	[(None, 70, 1)]

conv1d: Conv1D

input:	(None, 70, 1)
output:	(None, 70, 64)

batch_normalization: BatchNormalization

input:	(None, 70, 64)
output:	(None, 70, 64)

max_pooling1d: MaxPooling1D

input:	(None, 70, 64)
output:	(None, 35, 64)

conv1d_1: Conv1D

input:	(None, 35, 64)
output:	(None, 35, 128)

batch_normalization_1: BatchNormalization

input:	(None, 35, 128)
output:	(None, 35, 128)

max_pooling1d_1: MaxPooling1D

input:	(None, 35, 128)
output:	(None, 17, 128)

conv1d_2: Conv1D

input:	(None, 17, 128)
output:	(None, 17, 256)

batch_normalization_2: BatchNormalization

input:	(None, 17, 256)
output:	(None, 17, 256)

flatten: Flatten

input:	(None, 17, 256)
output:	(None, 4352)

dense: Dense

input:	(None, 4352)
output:	(None, 256)

dense_1: Dense

input:	(None, 256)
output:	(None, 128)

dense_2: Dense

input:	(None, 128)
output:	(None, 64)

dense_3: Dense

input:	(None, 64)
output:	(None, 1)

Figure 4.
1D-CNN architecture

conv2d_input: InputLayer	input:	[(None, 10, 7, 1)]
	output:	[(None, 10, 7, 1)]

conv2d: Conv2D	input:	(None, 10, 7, 1)
	output:	(None, 10, 7, 64)

batch_normalization: BatchNormalization	input:	(None, 10, 7, 64)
	output:	(None, 10, 7, 64)

max_pooling2d: MaxPooling2D	input:	(None, 10, 7, 64)
	output:	(None, 5, 3, 64)

conv2d_1: Conv2D	input:	(None, 5, 3, 64)
	output:	(None, 5, 3, 128)

batch_normalization_1: BatchNormalization	input:	(None, 5, 3, 128)
	output:	(None, 5, 3, 128)

max_pooling2d_1: MaxPooling2D	input:	(None, 5, 3, 128)
	output:	(None, 2, 1, 128)

conv2d_2: Conv2D	input:	(None, 2, 1, 128)
	output:	(None, 2, 1, 256)

batch_normalization_2: BatchNormalization	input:	(None, 2, 1, 256)
	output:	(None, 2, 1, 256)

flatten: Flatten	input:	(None, 2, 1, 256)
	output:	(None, 512)

dense: Dense	input:	(None, 512)
	output:	(None, 256)

dense_1: Dense	input:	(None, 256)
	output:	(None, 128)

dense_2: Dense	input:	(None, 128)
	output:	(None, 64)

dense_3: Dense	input:	(None, 64)
	output:	(None, 1)

Figure 5.
2D-CNN architecture

3. Results and discussions

All experiments conducted in the present work are performed on Intel i5 8 GB RAM AMD processor using Python 3.7 programming language. PyTorch version 1.7.0 is used to implement CNN, and the execution of CNN is achieved on Kaggle GPU.

The present work is executed in the following steps:

3.1 Preparation of data

The data set used to pursue research was originally prepared by Correa *et al.* (2018) to identify affect, mood and personality in an intricate format. This data set comprises of 40 folders wherein each folder corresponds to one participant. Further each folder consists of a MATLAB file with the following list, as shown in Table 2.

In the present study, the data for 16 short videos were taken for 14 columns of EEG and their respective labels in AVD, from self-assessment labels list. Emotion indices responses under AVD were coded as 1 and 0 according to Table 3.

3.1.1 Balancing for emotions. After preparing the data set, the number of samples in each category AVD are plotted in Figure 6(a). It is evident that the number of samples recorded as low emotions for each category is significantly fewer than those recorded as high emotional level. Thus, the low and high emotions of each category were balanced by the Python function, SMOTE. The result of upsampling is shown in Figure 6(b).

The resultant number of samples is insufficient for applying DL methods. Moreover, replication in data causes reduction in the accuracy of models as shown in Table 6. To overcome these limitations, the data is being generated by non-overlapping sliding window (NOSW) and OSW in the present work. The resultant number of samples is shown in Table 4.

Name	Size	Content
Joined_data	1×20	In this array, there are 20 columns corresponding to 16 short videos and 4 long videos shown to the participants. Each cell consists of a matrix of size $y \times 17$. Here, y is variable the value of which depends on the length of video. In this matrix, there are total 17 columns out of which 14 are corresponding to EEG signals, 2 are corresponding to ECG and last is corresponding to GSR signal
Labels_self-assessment	1×20	In this array, there are 20 columns corresponding to 16 short videos and 4 long videos shown to the participants. Each cell consists of a matrix of size 1×12, wherein 12 columns correspond to 12 assessments (arousal, valence, dominance, liking, familiarity and seven basic emotions) by participant for every video. The first five dimensions of emotions are measured on a scale of 1–9, where 1 is the lowest and 9 is the highest. The seven basic emotions (neutral, disgust, happiness, surprise, anger, fear and sadness) are displayed in binary (i.e. 0 or 1)
Labels_ext_annotation	1×20	In this array, there are 20 columns corresponding to 16 short videos and 4 long videos shown to the participants. Each cell consists of a matrix of size $z \times 3$, where z is number of segments in a video each of length 20 seconds and three columns holds the value for segment number, arousal and valence

Table 2.
Content of
MATLAB files

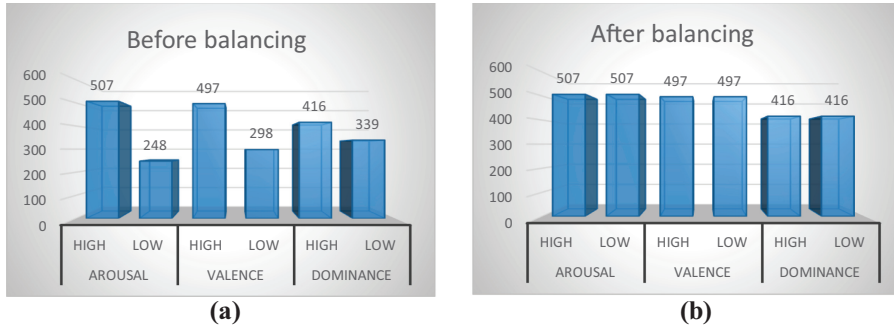| High arousal (HA) = 1	Low arousal (LA) = 0	High valence (HV) = 1	Low valence (LV) = 0	High dominance (HD) = 1	Low dominance (LD) = 0
>4.5	≤4.5	>4.5	≤4.5	>4.5	≤4.5

Table 3.
Coding of AVD from
1–9 to 0–1

Figure 6.
High and low emotion
indices in AVD (a) prior
to balancing (b) after
balancing

S. No	Sampling technique	Arousal	Valence	Dominance
1	Original samples	755	795	755
2	After resampling by SMOTE	1,014	994	832
3	Samples generated after decomposition of signal by NOSW	29,382	29,382	29,382
4	Samples generated after decomposition of signal by OSW	458,664	458,664	458,664

Table 4.
Number of samples
generated after
resampling methods

3.2 Feature extraction and classification
The decomposed signals were cleaned by removing NaN values. Moreover, five NBP and five NWE features corresponding to five EEG band were extracted by Fourier and wavelet transform, respectively. A combined vector of both features {NBP, NWE} is also formed by appending the NWE features to the NBP features. A total of 70 (=14 × 5) features were extracted for 14 EEG channels by each of NBP and NWE separately. Thus, there are 140 features present for combined vector. The features extracted by different resampling methods are shown in Table 5.

CNN classifiers discussed in Section 2.3 were applied to individual and combined features. The train, validation and test samples are divided into a 70:40:30 ratio. The learning rate, batch size and optimiser are taken as 0.001, 32 and "adam", respectively. A binary cross-entropy function was used as loss function.

The training of CNN continues until the accuracy of the network does not become constant/start decreasing. The emotion recognition accuracies of two DL classifiers were compared with baseline ML model, SVM in Table 6; the highest accuracy is shown in italic.

Resampling method	Feature modality	Arousal	Valence	Dominance
SMOTE	Individual feature	1014 × 70	994 × 70	884 × 70
	Combined feature	1014 × 140	994 × 140	884 × 140
NOSW	Individual feature	29382 × 70	29382 × 70	29382 × 70
	Combined feature	29382 × 140	29382 × 140	29382 × 140
OSW	Individual feature	458664 × 70	458664 × 70	458664 × 70
	Combined feature	458664 × 140	458664 × 140	458664 × 140

Table 5.
Feature dimensions
obtained after feature
extraction

Resampling method	Classifier	Feature	Arousal	Valence	Dominance
SMOTE	SVM	NBP	*80.76*	68.81	63.35
		NWE	76.92	*73.11*	67.17
		{NBP,NWE}	77.88	67.74	*74.04*
	1D CNN	NBP	61.7	58.09	61.93
		NWE	63.99	50.49	60.01
		{NBP,NWE}	61.86	58.17	58.87
	2D CNN	NBP	62.35	53.4	58.52
		NWE	67.89	63.56	60.01
		{NBP,NWE}	68.11	61.22	55.53
NOSW	SVM	NBP	71.21	62.85	54.15
		NWE	80.11	73.73	76.27
		{NBP,NWE}	81.14	75.19	78.39
	1D CNN	NBP	71.33	66.56	65.66
		NWE	78.64	69.55	73.13
		{NBP,NWE}	78.3	69.97	70.09
	2D CNN	NBP	75.41	71.81	71.37
		NWE	80.22	75.5	76.67
		{NBP,NWE}	*81.79*	*75.59*	*78.67*
OSW	SVM	NBP	84.56	88.21	87.02
		NWE	87.05	83.22	85.18
		{NBP,NWE}	70.64	63.79	56.25
	1D CNN	NBP	91.45	92.93	93.47
		NWE	89.9	85.95	88.21
		{NBP,NWE}	93.66	93.14	92.62
	2D CNN	NBP	94.22	93.78	94.08
		NWE	92.3	90.35	91.51
		{NBP,NWE}	*96.63*	*95.87*	*96.3*

Table 6.
Accuracy obtained
after applying ML/DL
classifiers

From Table 6, it is evident that the accuracies obtained after resampling by SMOTE are at least 1% lower than the NOSW and at least 15.87% lower than OSW. Comparing the ML and DL methods, SVM performs better under SMOTE resampling since sample size is small. But, in NOSW and OSW resampling, 2D CNN performs best since sample size is large. A feature-wise comparison of all methods is shown in Figure 7.

From Figure 7(a), it can be observed that SVM outperforms other methods, and no specific pattern is observed that can indicate which among individual or combined features perform better. It is evident from 7(b) that the NWE feature is providing higher accuracy in the case of NOSW. In contrast, the NBP feature is providing higher accuracies with the OSW for DL methods shown in Figure 7(c). The combined features for 2D CNN give higher accuracies for both NOSW and OSW shown in Figure 7(b) and 7(c). Thus, by combining the observations made by Table 6, Figure 7(b) and 7(c), a 2D CNN classifier with a combined feature vector found best for all the emotional indices with 96.63%, 95.87% and 96.30% accuracies, respectively.

An execution history of 2D CNN with combined features for overlapping window is shown in Figure 8 in terms of loss and accuracy curve for arousal, valence and dominance, respectively. Loss curves represent the training and validation loss, which is expected to be as close as possible. The accuracy curve shows accuracy obtained for each emotion indices for 20 epochs.
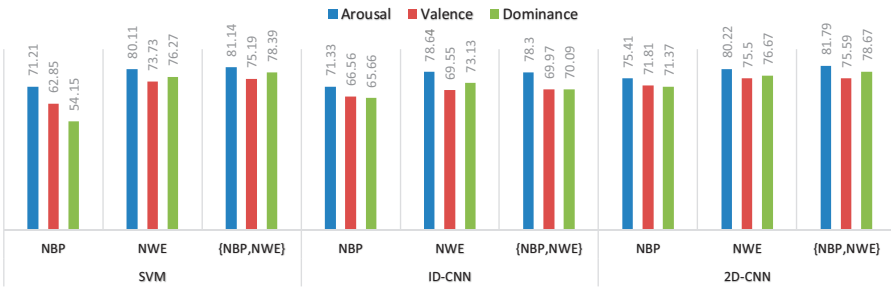
The results are also compared for time of execution of individual feature versus combine features shown in Figure 9.

Figure 9 shows that as the sample size increases from SMOTE → NOSW → OSW, the time for execution increases significantly in case of SVM for both individual and combined
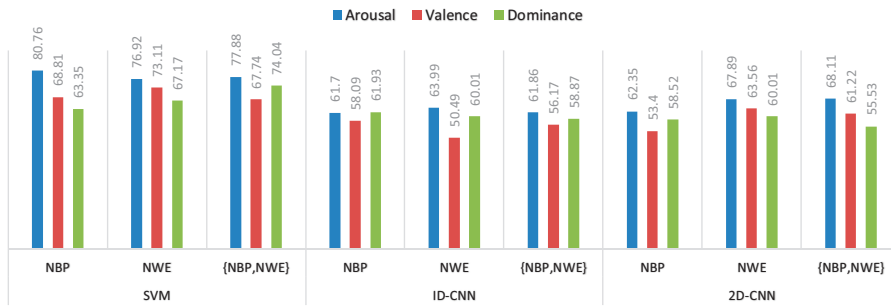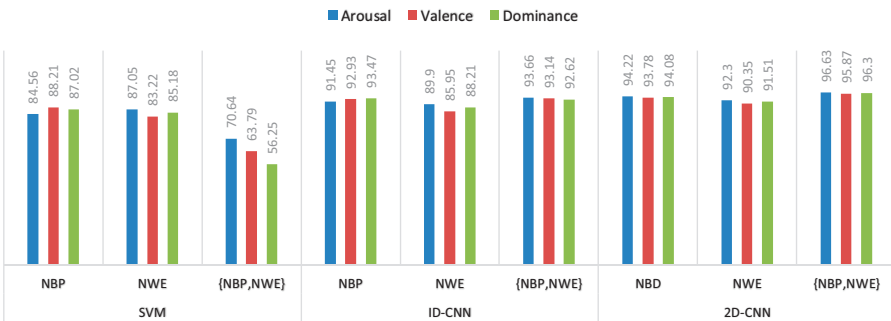
Comparison of feature extraction methods for SMOTE

(a)

Comparison of feature extraction methods for NOSW

(b)

Comparison of feature extraction methods for OSW

(c)

Figure 7.
Comparison of
methods performance
(in terms of accuracies)
for different feature
extraction method

features. The reason for this observation is the fact that SVM cannot be executed on GPUs since it involves complex calculations. As observed in this study that basic SVM performs poorly when the sample size is large (as in case of combined feature with OSW in Table 6), the same is being reported in [39].

Table 7 compares the results obtained in present study with ERS articles published from 2018 onwards on AMIGOS data set.
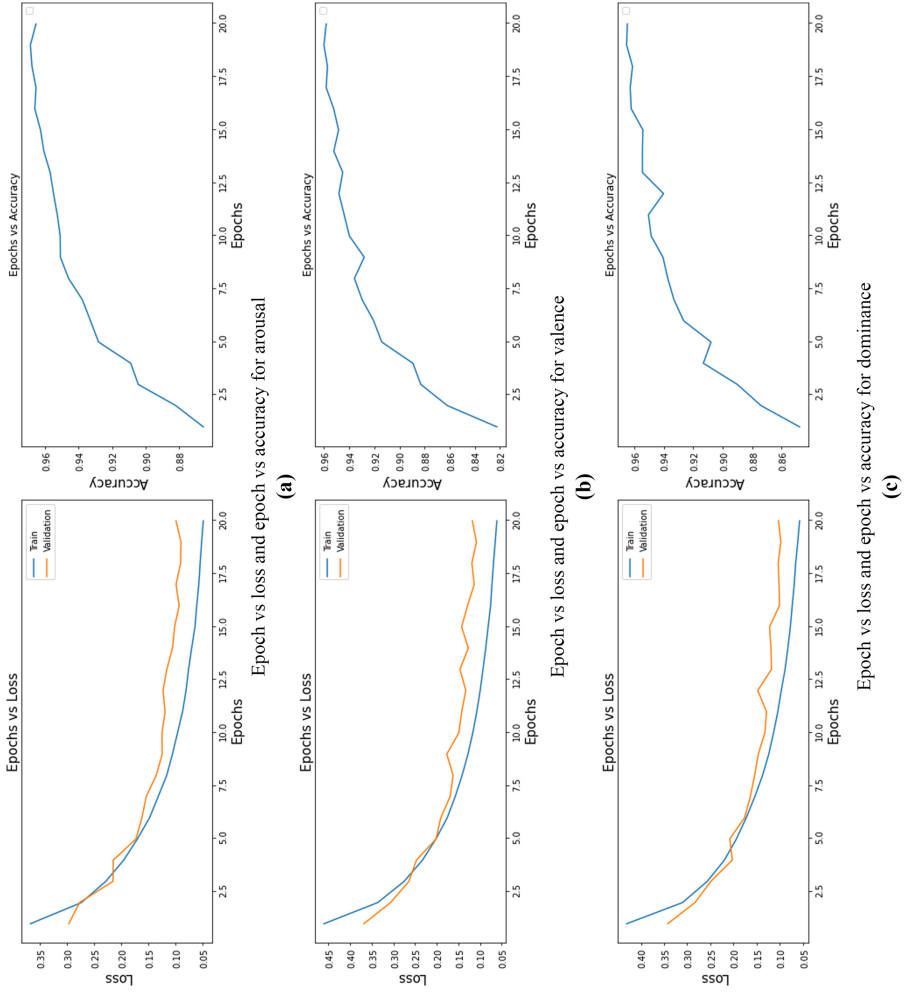
Figure 8.
Loss and accuracy
curve for AVD
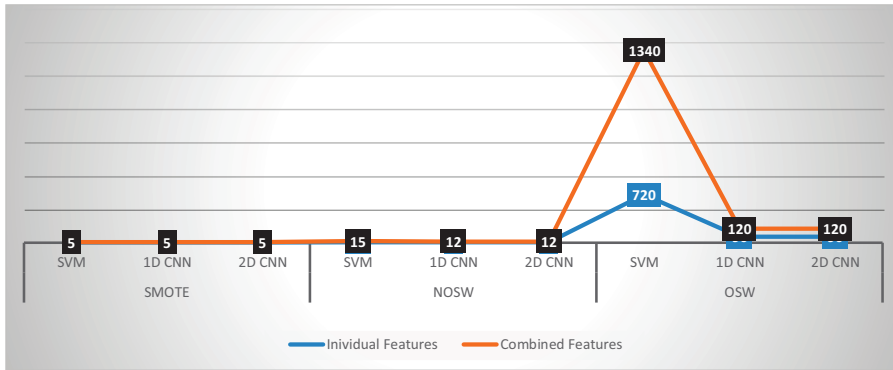
Ref, year	Arousal %	Emotions Valence %	Dominance %	Modality	Features	Classifier
[33], 2018(Original paper)	57.7	56.4	–	EEG	All band, PSD, spectral power asymmetry between 7 pairs of electrodes in the five bands	SVM
[40], 2018	71.54	66.67	72.36	EEG	Conditional entropy (CF) feature, CNN based feature using EEG topography	Extreme learning machine (ELM)
[41], 2018	68.00	84.00	–	EEG + ECG + GSR	Time, frequency and entropy domain features	GaussianNB, XGBoost
[42], 2019	83.02	79.13	–	EEG	PSD, Conditional entropy, PSD image based deep learning features	LSTM
[28], 2020	83.30	79.40	–	EEG + ECG + GSR	Spectrogram representation	Bidirectional LSTM
[29], 2021	90.54	87.39	–	EEG	Features extracted from topographic and holographic feature map	CNN + SVM
Our method	96.63	95.87	96.30	EEG	NBP + NWE	2D CNN

Note(s): *The "–" in dominance column mean the study is conducted for 2D emotions only

From Table 7, the emotions are recognised using only EEG data in [29, 33, 40, 42]. The other studies were carried out using multimodal data. The first study [33] conducted on the AMIGOS data set provides an initial analysis, produces very low accuracy – 57.7% and therefore poses an open research challenge. The accuracy was improved to 71.54% in [40] in same year, in which the features were extracted using CNN. A multimodal ERS was proposed in [28, 41], producing accuracies of up to 84%. Highest accuracy achieved on AMIGOS data set prior to this work is 90.54% using CNN + SVM model in [29]. Finally, the present proposed model has improved the accuracy up to 96.63% with a single modality (EEG) through a 2D CNN classifier.

Siddharth *et al.* in 2019 [42] worked on four data sets (DEAP, DREAMER, AMIGOS and MAHNOB-HCI) using LSTM. However, it is observed by them that LSTM is difficult to

implement on AMIGOS data set, which has varying lengths of data. This indicates necessity for executing an efficient pre-processing method prior to classification. The present paper offers the most efficient classification strategy for EEG records of varying lengths through decomposition of data using an OSW approach which provides an efficient alternative for handling imbalanced variable-length data prior to the classification.

4. Conclusions

Despite significant development in the field of DL and its suitability to various applications, almost 59% of researchers have used an SVM with RBF kernels for BCIs [19]. This is due to the unavailability of a large-scale data set for BCIs. However, DL models are widely applied in speech and visual modality. A BCI data set provides genuine human responses as they are taken directly from the human body. Thus, ER using brain signals is preferred. There is a need for an "off-the-shelf" method to conduct research on BCIs with a high accuracy. The accuracy found in BCIs is generally low – especially for the AMIGOS data set.

The present contribution focusses on obtaining predictive outcome of the 3D emotion responses to EEG signals in context of imbalanced variable-length records. Novelty of the present paper is that it proposes application of OSW for CNN to the intricate AMIGOS data set aimed at highly accurate prediction of 3D emotions in contrast to the accuracy achieved by the existing approaches available in literature. Most of the earlier analysis of AMIGOS data set has been pivoted on 2D emotion analysis. The current paper views EEG (14 channels) on 3D emotions for predictive inference and presents a comparative assessment of the predictive accuracy with that of Siddharth *et al.* (2018) [40]. Thus, the present approach is found to have the highest accuracy with respect to all the three AVD emotion indices as compared to similar works referenced in literature (Table 7).

The present work can be further extended for multiple modalities in physiological signals as well as with the inclusion of response to video interventions such as in automatic video recommendation system for enhancing the mood of individuals. Another possible extension of this work can be accomplished by representing the signal features in 2D/3D form and subsequently combining them with the respective video/image features.

References

1. Kövecses Z. Emotion concepts. New York: Springer Science and Business Media; 2012 Dec 6.

2. Alarcao SM, Fonseca MJ. Emotion recognition using EEG signals: a survey. IEEE Trans Affect Comp. 2017 Jun 12; 10(3): 374-93.

3. Saxena A, Khanna A, Gupta D. Emotion recognition and detection methods: a comprehensive survey. J Art Int Sys. 2020 Feb 7; 2(1): 53-79.

4. Lin YP, Wang CH, Jung TP, Wu TL, Jeng SK, Duann JR, Chen JH. EEG-based emotion recognition in music listening. IEEE (Inst Electr Electron Eng) Trans Biomed Eng. 2010 May 3; 57(7): 1798-806.

5. Santamaria-Granados L, Munoz-Organero M, Ramirez-Gonzalez G, Abdulhay E, Arunkumar NJ. Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). IEEE Access. 2018 Nov 23; 7: 57-67.

6. Xiefeng C, Wang Y, Dai S, Zhao P, Liu Q. Heart sound signals can be used for emotion recognition. Sci Rep. 2019 Apr 24; 9(1): 1.

7. Recio G, Schacht A, Sommer W. Recognizing dynamic facial expressions of emotion: specificity and intensity effects in event-related brain potentials. Biol Psychol. 2014 Feb 1; 96: 111-25.

8. El Ayadi M, Kamel MS, Karray F. Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recogn. 2011 Mar 1; 44(3): 572-87.

9. Gunes H, Piccardi M. Bi-modal emotion recognition from expressive face and body gestures. J Net Comp Appl. 2007 Nov 1; 30(4): 1334-45.

10. Song T, Liu S, Zheng W, Zong Y, Cui Z, Li Y, Zhou X. Variational instance-adaptive graph for EEG emotion recognition. IEEE Trans Affect Comp. 2021 Mar 9, Early access. doi: 10.1109/TAFFC.2021.3064940.

11. Barlow JS. The electroencephalogram: its patterns and origins. Cambridge, MA and London: MIT Press; 1993.

12. Yazıcı M, Ulutaş M. Classification of EEG signals using time domain features. 2015 23nd Signal Processing and Communications Applications Conference (SIU): IEEE; 2015 May 16. 2358-2361.

13. Liu Y, Sourina O. Real-time fractal-based valence level recognition from EEG. Transactions on computational science; Berlin, Heidelberg: Springer: 2013; 18. p. 101-120.

14. Petrantonakis PC, Hadjileontiadis LJ. Emotion recognition from EEG using higher order crossings. IEEE Trans Inf Tech Biomed. 2009 Oct 23; 14(2): 186-97.

15. Kim C, Sun J, Liu D, Wang Q, Paek S. An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI. Med Biol Eng Comput. 2018 Sep; 56(9): 1645-58.

16. Zhang R, Xu P, Chen R, Li F, Guo L, Li P, Zhang T, Yao D. Predicting inter-session performance of SMR-based brain–computer interface using the spectral entropy of resting-state EEG. Brain Topogr. 2015 Sep; 28(5): 680-90.

17. Zhang J, Wei Z, Zou J, Fu H. Automatic epileptic EEG classification based on differential entropy and attention model. Eng Appl Artif Intelligence. 2020 Nov 1; 96: 103975.

18. Al-Fahoum AS, Al-Fraihat AA. Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. Int Scholarly Res Notices. 2014: 1-7.

19. Gu X, Cao Z, Jolfaei A, Xu P, Wu D, Jung TP, Lin CT. EEG-based brain-computer interfaces (BCIS): a survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. IEEE ACM Trans Comput Biol Bioinf. 2021 Jan 19. doi: 10.1109/TCBB.2021.3052811.

20. Tao W, Li C, Song R, Cheng J, Liu Y, Wan F, Chen X. EEG-based emotion recognition via channel-wise attention and self attention. IEEE Trans Affect Com. 2020 Sep 22. doi: 10.1109/TAFFC.2020.3025777.

21. Sharma R, Pachori RB, Sircar P. Automated emotion recognition based on higher order statistics and deep learning algorithm. Bio Sig Pro Cont. 2020 Apr 1; 58: 101867.

22. Wang F, Wu S, Zhang W, Xu Z, Zhang Y, Wu C, Coleman S. Emotion recognition with convolutional neural network and EEG-based EFDMs. Neuropsychologia. 2020 Sep 1; 146: 107506.

23. Liu Y, Ding Y, Li C, Cheng J, Song R, Wan F, Chen X. Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. Comput Biol Med. 2020 Aug 1; 123: 103927.

24. Zhong P, Wang D, Miao C. EEG-based emotion recognition using regularized graph neural networks. IEEE Trans Affect Comp. 2020 May 11. doi: 10.1109/TAFFC.2020.2994159.

25. Yin Y, Zheng X, Hu B, Zhang Y, Cui X. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. Appl Soft Comput. 2021 Mar 1; 100: 106954.

26. Khare SK, Bajaj V. Time-frequency representation and convolutional neural network-based emotion recognition. IEEE Trans Neural Net Lear Sys. 2020 Jul 31; 32(7): 2901-2909.

27. Cui H, Liu A, Zhang X, Chen X, Wang K, Chen X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. Knowl Base Syst. 2020 Oct 12; 205: 106243.

28. Li C, Bao Z, Li L, Zhao Z. Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition. Inf Process Management. 2020 May 1; 57(3): 102185.

29. Topic A, Russo M. Emotion recognition based on EEG feature maps through deep learning network. Eng Sci Tech Int J. 2021 Apr 16. doi: 10.1016/j.jestch.2021.03.012.

30. He H, Tan Y, Ying J, Zhang W. Strengthen EEG-based emotion recognition using firefly integrated optimization algorithm. Appl Soft Comput. 2020 Sep 1; 94: 106426.

31. Russell JA. A circumplex model of affect. J Personal Soc Psychol. 1980 Dec; 39(6): 1161.

32. Verma GK, Tiwary US. Affect representation and recognition in 3D continuous valence–arousal–dominance space. Multimed Tool Appl. 2017 Jan; 76(2): 2159-83.

33. Correa JA, Abadi MK, Sebe N, Patras I. Amigos: a dataset for affect, personality and mood research on individuals and groups. IEEE Trans Affect Com. 2018 Nov 30; 12(2): 479-493.

34. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif intelligence Res. 2002 Jun 1; 16: 321-57.

35. Angelov P, Sperduti A. Challenges in deep learning. In: ESANN 2016 - 24th European symposium on artificial neural networks. ESANN 2016 - 24th European symposium on artificial neural networks. i6doc.com publication, BEL; 2016. 489-496. ISBN 9782875870278.

36. Bracewell RN, Bracewell RN. The Fourier transform and its applications. New York, NY: McGraw-Hill; 1986 Feb.

37. Daubechies I. Ten lectures on wavelets, CBMS conf. Series Appl Math; 1992 Jan 1; 61.

38. Le QV. A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks. Google Brain. 2015 Oct 20: 1-20 [online] Available from: https://cs.stanford.edu/~quocle/tutorial2.pdf.

39. Cervantes J, Li X, Yu W, Li K. Support vector machine classification for large data sets via minimum enclosing ball clustering. Neurocomputing. 2008 Jan 1; 71(4–6): 611-9.

40. Siddharth, Jung TP, Sejnowski TJ. Multi-modal approach for affective computing. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); IEEE; 2018 Jul 18. 291-294.

41. Tung K, Liu PK, Chuang YC, Wang SH, Wu AY. Entropy-assisted multi-modal emotion recognition framework based on physiological signals. 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES); IEEE: 2018 Dec 3. 22-26.

42. Siddharth S, Jung TP, Sejnowski TJ. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. IEEE Trans Affect Com. 2019 May 14. doi: 10.1109/TAFFC.2019.2916015.

Annexure
Annexure is available online for this article.

Corresponding author
Shruti Garg can be contacted at: gshruti@bitmesra.ac.in