Received 20 March 2023, accepted 6 April 2023, date of publication 13 April 2023, date of curre
Digital Object Identifier 10.1109/ACCESS.2023.3266804
Emotion Recognition Using Temporally Localized Emotional Events in EEG With Naturalistic Context: DENS# Dataset
[SANSÜRLÜ]
, (Graduate Student Member, IEEE), [SANSÜRLÜ]
,
[SANSÜRLÜ], [SANSÜRLÜ]
, (Senior Member, IEEE)
Indian Institute of Information Technology Allahabad, Allahabad, Uttar Pradesh 211012, India
Corresponding authors: Sudhakar Mishra (rs163@iiita.ac.in), Mohammad Asif (pse2017001@ii
(ust@iiita.ac.in)
This work was supported by the Ministry of Education, Government of India, funded by the acqu
ABSTRACT Emotion recognition using EEG signals is an emerging area of research due to its l
applicability in Brain-Computer Interfaces. Emotional feelings are hard to stimulate in the lab. Em
don·t last long, yet they need enough context to be perceived and felt. However, most EEG-rela
emotion databases either suffer from emotionally irrelevant details (due to prolonged duration st
or have minimal context, which may not elicit enough emotion. We tried to overcome this proble
designing an experiment in which participants were free to report their emotional feelings while
the emotional stimulus. We called these reported emotional feelings ··Emotional Events·· in our
on Emotion with Naturalistic Stimuli (DENS), which has the recorded EEG signals during the em
events. To compare our dataset, we classify emotional events on different combinations of Vale
Arousal(A) dimensions and compared the results with benchmark datasets of DEAP and SEED.
Time Fourier Transform (STFT) is used for feature extraction and in the classification model con
CNN-LSTM hybrid layers. We achieved significantly higher accuracy with our data compared to
and SEED data. We conclude that having precise information about emotional feelings improve
classification accuracy compared to long-duration recorded EEG signals which might be contam
mind-wandering. This dataset can be used for detailed analysis of specific experienced emotion
brain dynamics.
[SANSÜRLÜ] Affective computing, CNN, DEAP, DENS, EEG, emotion dataset, emotion recogn
LSTM, SEED.
I. INTRODUCTION
Emotion recognition has been a challenging task in artificial intelligence. Several methods are available for measuring the participants· emotions. These methods include behavioural changes, subjective experiences self-reported by the participants, peripheral and central nervous system measures, etc [1]. Brain activities are among the most robust dimensions of detecting human affect, as it is difficult for the users to manipulate innate brain activity during the process. Accordingly, Electroencephalography (EEG) is considered a
The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li

to be considered for the classification model, as there is no information about the precise temporal location at which a participant may experience the emotion. Models must consider all the data presented for that label, which is unnecessarily computationally expensive and decreases the system·s efficiency by feeding not-so-essential data in the input.

In our approach, we have presented a novel method to overcome this issue by providing precise information about the emotion elicitation, self-reported by the participants. We call it an ·Emotional Event·. In this method, an additional task is given to the participants to mention precise temporal information by clicking on their computer screens while watching the emotional clips if they feel some emotion. Also, to the best of our knowledge, there are no EEG affective datasets available for the Indian subcontinent population. Hence we tried to reduce this research gap in our work. We have considered DEAP dataset [2] and SEED dataset [3] for comparison. We tried to follow a format similar to the benchmark datasets and compared our dataset·s results with these datasets based on statistical significance.

EEG measures the electrical signals from the scalp with temporal details. Different EEG devices vary with the number of channels of EEG. Thirty-two or fewer EEG channels are especially notable in affective computing research [6]. A few studies are also available with up to 64 electrodes. In this work, we used a 128-channel EEG device to detect emotions. This EEG cap follows the International 10-10 system·s standards [7].

Emotions are complex and challenging to understand as many theories exist about emotions, and there is a lack of a single consensus theory [8]. The study of emotions has been an emerging topic that combines multi-disciplines such as psychology, neuroscience, computer science and medicine, etc. There are different aspects involved in determining emotions, such as behavioural, psychological and physiological aspects, cognitive appraisals, facial expressions, vocal responses, subjective experiences, etc. This study focuses on physiological aspects of emotion, which are considered into account by the brain signals captured through EEG while watching emotional video clips. Further, this study tries to collect a comprehensive list of subjective experiences through a self-assessment rating at the end of each clip.

FIGURE 1. Complete Flowgram of the Experiment.
techniques has grown within the last few years [18]. This work employs the widely used state-of-the-art deep learning methods to detect emotions from EEG signals.

In this work, we contribute to the affective computing research by emphasising the importance of considering the duration of the signal encoding information about emotional experience. Emotion duration is the essential component of emotion dynamics [19], which is ignored in other datasets. We take account of emotion duration, which, to the best of our knowledge, had never been considered before. By comparing with other datasets using the same stimulus modality, we show that better emotion recognition accuracy can be achieved if the temporal information is incorporated.

This paper is organized into six sections. In the introduction section, we introduced the ongoing trends in affective computing, EEG emotion analysis and our dataset. In the next section, we introduced our proposed dataset- DENS, Emotional Events, experimental details (e.g., stimuli, EEG recordings, ratings etc.), preprocessing of the EEG data, its salience features and other datasets used (DEAP and SEED). In the methodology section, we discussed the feature extractions, input preprocessing of the extracted features for the classifier and deep learning model architecture for the same. Next, we have the results section, discussing the comparison results of the DENS-DEAP and DENS-SEED data based on several parameters and also comparing our results with recent studies. After that, we have a discussion section discussing the results and future aspects. At last, we concluded our analysis in the conclusion section.

II. [SANSÜRLÜ]
STIMULI (DENS)

The complete flow diagram of our experiment is given in Fig. 1. We call our dataset ·Dataset on Emotion with Naturalistic Stimuli· (abbreviated as DENS) [20].

A. [SANSÜRLÜ]

Emotion is a complex phenomenon which is embedded within a context [21]. Moreover, emotion is transient in nature and is not available throughout the stimulus duration. In fact, more than one aspect could be embedded within the stimulus context, and different participants can feel emotion at different points of time considering various aspects. However, most of the datasets recorded to date [2], [3] ignore the transient

TABLE 1. Selected stimuli for EEG study from the stimuli dataset we created [28]). The time duration of each stimulus is 60s. Stimulus Ids are given for references available in the open science framework repository.

of stimuli is critical, and for that, technical validation of the video clips is crucial to assess if the intended emotional experience is elicited by the stimuli. We have used naturalistic stimuli to elicit emotions in the participants. Naturalistic stimuli are dynamic emotional scenes in which multi-sensory perception is applied. It resembles more to the real-life scenario as compared to static and simple stimuli. In our previous work, we have validated a set of multimedia stimuli and created an affective stimuli database [28]. We selected 16 emotional stimuli from this database to perform our EEG experiment. The selection criteria for these 16 emotional stimuli are based on three factors:

1) A high probability of eliciting target emotions (calculated on the basis of ratings available).
2) Few stimuli must be available for each emotion category.
3) Since this experiment was done on the Indian population, more emphasis was given to Indian clips.

Besides these 16 emotional stimuli, we have validated 2 non-emotional stimuli separately. These clips were rated around 5 mean valence and arousal values (on a scale of 1 to 9). These non-emotional clips included the world·s longest road routes or animated history of the Babylonian era, which may not contribute to eliciting emotions. The inclusion of non-emotional stimuli was to validate the participants· responses and avoid the long accumulation of the affects during the experiment.

For each participant, nine (9) emotional stimuli were selected randomly from the 16 selected emotional stimuli and two (2) non-emotional stimuli. Each stimulus was of 60 seconds.

Table 1 shows the list of 16 emotional stimuli with the target emotions assigned during the stimuli validation.

2) [SANSÜRLÜ]

We recorded the EEG activity of

FIGURE 2. Emotion Category Selection Screen for Emotional Event (Click): After the participan
Dominance, Liking, Familiarity and Relevance, they are shown this screen for emotion category
The middle one belongs to the time of the click; the left one is 20 frames earlier, and the right on
stimulus clips were shown in 30 frames per second). It helps participants to recall easily. They c
experienced emotion is not present in the list, they were free to write their own.

3) RATINGS

Subjective ratings are one of the well-known methods to evaluate the personal emotional experience of the participants. Emotional pictures/videos or audio clips are presented to the participants, and they are asked to rate these clips on different scales based on their personal experiences. These scales include Valence, Arousal, Dominance, Liking, Familiarity and Relevance. The rating scales range from 1 to 9 for Valence, Arousal and Dominance. For Liking, familiarity and Relevance, it ranges from 1 to 5. Although, in this analysis, we considered only valence and arousal scales.

4) [SANSÜRLÜ]

As explained above, 465 emotional events were extracted from the forty participants in this experiment. All the participants clicked at least one time (average 1.29 times) during the stimulus.

Although for each participant and each stimulus, EEG recording is available for the whole stimulus (i.e., for the 60s), we have considered the signal for 7 seconds duration (1 second before the click and 6 seconds after the click) for each emotional event. We have tested for other time durations (e.g., 8s, 9s, up to 10s) but found better results with 7s duration. The recording has a sampling rate of 250 Hz.

C. [SANSÜRLÜ]

[SANSÜRLÜ]

The procedure followed to perform the preprocessing is described elsewhere [29]. The critical step which should be described here includes filtering and artifact removal. We had 128-channel EEG raw data with a sampling rate of 250 Hz. The raw signal is filtered using a Butterworth fifth-order bandpass filter with the passband 1-40 Hz. Independent component analysis (ICA) is used to remove artifacts, including heart rate, muscle movement, and eye blink-related artifacts.

D. [SANSÜRLÜ]

We have used DEAP dataset [2] (a dataset for emotion analysis using EEG, physiological and video signals) and SEED dataset [3] (A dataset collection for various purposes

downsampled, bandpass frequency of 4-45 Hz and EOG removed. For each trial, there are 4 labels available- Valence (V), Arousal (A), Dominance and Linking. We have used only V-A space for the experiment purpose.

The SEED dataset was recorded for 15 participants, and emotions were presented to the participants into three categories- positive, negative and neutral emotions (i.e., only valence (V) values were used). We have used only V-space in the DENS dataset to match the number of classes for both the datasets. The data was recorded using 62 channels.

E. [SANSÜRLÜ]

To sum up, we are highlighting some key points of our dataset-

· To the best of our knowledge, the first time, we created a dataset on Emotion with Naturalistic Stimuli (DENS) and recorded EEG signals from participants in the Indian subcontinent.

· Stimuli that are used to record EEG data of the participants are pre-validated on a different set of participants for the selected emotion categories.

· Participants were free to select any emotion category, whatever they felt for the stimuli from the given list.

· We used 128-channel high-density EEG recording for higher spatial resolution.

· Emotional Event: Temporal markers are available for each emotion category when participants feel the emotion, resulting in higher temporal resolution.

III. METHODOLOGY

A. [SANSÜRLÜ]

EEG Signals are non-stationary, meaning the signal·s statistical characteristics change over time. If these signals are transformed to the frequency domain using Fourier Transform, it provides the frequency information, which is averaged over the entire EEG signal. So, information on different frequency events is not analyzed properly. If a signal is cut into minor segments such that it could be considered as stationary and focus on signal properties at a particular section which is called a windowing section and apply Fourier transform on it, it is called as Short-Time Fourier Transform (STFT). It will move to the entire signal length and apply Fourier transform to find the spectral content of that section and display the coefficient as a function of both time and frequency. It provides insight into the nature of

FIGURE 3. Model Architecture: It is consisted of two 2D-convolution layers with 3 × 3 kernels a
max pooling layer followed by a dropout layer and flattening layer. A repeat vector layer of size
Two LSTM layers are used of sizes 256 units and 128 units respectively, each followed by a dr
sizes 64 (followed by a dropout layer) and 4 or 3 (equals the number of the output classes).

2) [SANSÜRLÜ]

SEED dataset contains 45.mat files for 15 subjects for each subject with 3 trials. The label file contains 3 emotional labels ·1 for negative, 0 for neutral, and 1 for positive on the valence scale. After renaming, the labels become 0 for neutral, 1 for positive, and 2 for negative. For classification, we have considered 15.mat files, one trial per subject. Due to the different sizes of data length in each channel, the first 16000 sample for each data which is the first 80s of data, is considered for further processing. EEG cap includes 62 channels according to the 10-20 international system. So, 15 subjects, 15 trials, 62-channels, and 16000 EEG data are converted into a tensor of $X \cdot R13950\times16000$ (i.e., 15 subjects×15 trials×62 channels, 16000 samples)for feature extraction. As mentioned in the DEAP dataset experiment, using STFT with a window size of 0.5s and overlap of 0.25s, each 16000 EEG data is converted into a spectrogram with the shape of (51,319). Then, a hybrid CNN-LSTM classifier was implemented for multi-class classification with input tensor shape $X \cdot R51\times319\times3$.

3) [SANSÜRLÜ]

For the DENS dataset, we have 465.mat files which contain emotional events. All 465 files are picked for the experiment. Each.mat file is a matrix of $X \cdot R128\times1751$, where 128 is the number of EEG channels and 1751 is the sample data for each channel. Then we have converted the data tensor of $X \cdot R465\times128\times1751$ into the form of $R59520\times1751$ (i.e., 465 emotional events × 128 channels, 1751 samples) for feature extraction with window size 0.5s and overlap is 0.25s. After feature extraction, we have 59520 spectrograms, and each spectrogram is in the shape of (63, 26).

To compare with the DEAP dataset, the DENS dataset with 4-label classification is performed with a hybrid CNN-LSTM classifier. For the label, we used the same V-A space (HVHA, HVLA, LVLA, LVHA) (abbreviations- H: High,

FIGURE 4. Comparison of Confusion matrices for DEAP and DENS datasets over Valence-Aro
and assigned a label to it (0-HVHA, 1-HVLA, 2-LVHA and 3-LVLA). 4a: DEAP Dataset; 4b: DEN
A:Arousal; L:Low; H:High. The color bar represents the number of samples in the class.

the classification of emotions. The hybrid CNN-LSTM model utilizes the ability of convolutional layers for feature extraction from data, and LSTM layers are for long-term and short-term dependencies. The same model is used to compare all three datasets. The model classifier and its details are shown in Fig. 3.

CNN is often placed in the initial layers as it helps in local pattern learning from spectrogram or in general input data. The Pattern learning block consists of two 2D-convolutional blocks, each with a kernel size of (3 × 3). The feature map, which is the output of convolutional layers, keeps track of the location of the features in the input. A max-Pooling layer is added in between two consecutive convolutional layers. A pooling layer is added after the convolutional layer to reduce the feature-map dimension; hence it reduces the computational cost, and the activation function is applied to enhance the capability of the model. Rectified Linear Unit (ReLU) activation function which has been widely used to resilient vanishing gradient problem. In between, the dropout layer is used in some places to avoid the overfitting problem. The flattening layer transforms these feature maps into one-dimensional vectors. The repeat vector gives extra dimension for the LSTM layer. The sequential learning block consists of 2 LSTM layers which capture the long-term temporal dependencies from the feature map extracted by CNN layers. 1st LSTM layer consists of 256 cells with a return sequence set to ·True· while 2nd LSTM consists of 128 cells and as it is the last LSTM layer return sequence is ·False·. Between LSTM layers, dropout layers with rate = 0.2 are added to avoid overfitting issues. Finally, two fully-connected layers where 1st layer with 64 neurons and 2nd layer with the number of classes as neurons are added for further processing. As we have the multi-class classification, the SoftMax activation function is used in the output layer as it outputs a vector representing the probability distributions of a list of potential classes.

TABLE 2. Parameter Settings for the Model.

The parameter setting for the developed deep learning model is mentioned in Table 2.

IV. RESULTS

FIGURE 5. Comparison of Confusion matrices for SEED and DENS datasets over Valence spa
dataset provided data with three classes, while DENS data is divided into three classes based o
participants) and assigned a label to it as follows: For SEED:0 for neutral, 1 for positive and 2 fo
ratings range from 1-4.5), 1 for non-emotional data (valence ratings range from 4.5-5.5, as well
high-valence (valence ratings ranges from 5.5-9). 5a: SEED Dataset; 5b: DENS Dataset. The c
class.
TABLE 3. Comparison Table with Other Recent Studies.
FIGURE 6. F1 scores of DEAP vs DENS for all the 25 trials.
d estimate: ·13.70 (large), 95 percent confidence interval:
[·16.51 ·10.89].
TABLE 4. DEAP vs DENS with mean F1 scores.
TABLE 5. SEED vs DENS with mean F1 scores.
B. [SANSÜRLÜ]
For SEED vs DENS comparison, label classification we have used 3 labels on the valence scale. Comparison between SEED and DENS results is mentioned in Table 5. The loss and accuracy graphs are mentioned in Fig. 8.

FIGURE 7. F1 scores of SEED vs DENS for all the 25 trials.
Fig. 7 shows an F1 score comparison between SEED and DENS datasets per trial. Using t-test statistical testing, the 25 F1 score of SEED dataset (M = 95.65%, SD = 0.37%) compared with the 25 F1 score of DENS dataset (M = 97.68%, SD = 0.13%), DENS dataset shows better results with absolute t (31) = 25.466, p < 0.0001, Cohen·s d estimate: ·11.37 (large), 95 percent confidence interval: [·13.73 ·9.02].

C. [SANSÜRLÜ]

We have included some other recent studies and given a comparative table for their results in Table 3. The studies consist of CNN-RNN Hybrid models, R2G-STNN model that is based on regional to global BiLSTM with Attention layer, Attention-based CNN-RNN Hybrid model (ACRNN), BiDCNN that is Bi-hemisphere Discrepancy CNN model and ECLGCNN that is a fusion model of Graph CNN and LSTM model.

V. DISCUSSION

In this work, we captured emotional experiences within the ecologically valid naturalistic environment with a precise temporal marker than any study to date. As per recent theories, emotional experience is a constructing phenomenon which involves networks of the brain, including the default mode network, salience network, and fronto-parietal network. These networks are not specific to emotional experiences. In fact, these networks are domain-general networks which are involved in perception (in general). Though, the connectivity among these networks might not be the same in different perceptions which is apparently shown in our previous work [35]. In addition, different from normal perception, emotional experiences involve changes in body physiology [29]. Putting together the above-mentioned ideas from recent results hints that the emotional experiences can be easily confused with other perceptions, which might not be an emotional experience.

One of the major concerns is the mind-wandering activity while using the film stimuli. In the previous research, the whole stimulus is considered to elicit a single emotional experience. And the duration of the stimulus varied from seconds to minutes. Research shows that averaging the participant·s feedback for the whole duration of the stimulus might not be correctly capturing emotional experience (in

FIGURE 8. Loss and Accuracy Graphs for All the datasets Used.

experience in the same stimuli, but it is certainly possible that it can have more than one positive or more than one negative feeling in the movie.

VI. CONCLUSION

The work presented in this article is based on the concept that emotion is a short-lived phenomenon which might last for very few seconds. Hence, using long-duration EEG signals recorded during emotional stimulus watching might not contain emotional information for the whole duration. Therefore, we hypothesized that using only the duration of the signal where an emotional event is reported without compromising the ecological validity of the stimuli will contain more emotional information. To test the hypothesis, we designed an EEG experiment which uniquely marks the duration of the emotional event in the continuous recording of brain waves using EEG. We performed deep learning analysis using a hybrid CNN and LSTM model and found results that significantly favoured our hypothesis. In this work, we saw the problem with a different aspect which has not attracted the attention of the researcher. We suggest that future research on emotion recognition should adapt our approach to collect more such kinds of data so that emotion recognition using EEG can go beyond the emotions only and move towards recognizing and analyzing more complex emotions.

ACKNOWLEDGMENT

# Dataset on Emotion with Naturalistic Stimuli. Availability at (https://openneuro.org/datasets/ds003751). (Mohammad Asif and Sudhakar Mishra contributed equally to this work.)

REFERENCES

[1] I. B. Mauss and M. D. Robinson, ··Measures of emotion: A review,·· Cognition Emotion, vol. 23, no. 2, pp. 209·237, Feb. 2009.

[2] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, ··DEAP: A database for emotion analysis; using physiological signals,·· IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 18·31, Jan./Mar. 2012.

[3] W.-L. Zheng and B.-L. Lu, ··Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks,·· IEEE Trans. Auto. Mental Develop., vol. 7, no. 3, pp. 162·175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[4] S. Katsigiannis and N. Ramzan, ··DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices,·· [SANSÜRLÜ]. Biomed. Health Informat., vol. 22, no. 1, pp. 98·107, Jan. 2018, doi: 10.1109/JBHI.2017.2688239.

[33] D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, and D. Jiang, ··Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for EEG emotion recognition,·· Neurocomputing, vol. 448, pp. 140·151, Aug. 2021.
[34] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, ··EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM,·· Appl. Soft Comput., vol. 100, Mar. 2021, Art. no. 106954.
[35] S. Mishra, N. Srinivasan, and U. S. Tiwary, ··Dynamic functional connectivity of emotion processing in beta band with naturalistic emotion stimuli,·· Brain Sci., vol. 12, no. 8, p. 1106, Aug. 2022.
[36] H. Saarimäki, ··Naturalistic stimuli in affective neuroimaging: A review,·· Frontiers Hum. Neurosci., vol. 15, p. 318, Jun. 2021.
[37] Y.-Y. Lee and S. Hsieh, ··Classifying different emotional states by means of EEG-based functional connectivity patterns,·· PLoS ONE, vol. 9, no. 4, Apr. 2014, Art. no. e95415.
[SANSÜRLÜ] (Graduate Student Member, IEEE) received the bachelor·s degree in computer science and the master·s degree in cognitive science and in information technology (specializing in software engineering). He is currently a Research Scholar with the Indian Institute of Information Technology Allahabad, Allahabad. His research interest includes affective computing. He is also working on emotion recognition using brain signals. He is using EEG for emotion detection using validated stimuli. He is also working on deep learning architectures.
[SANSÜRLÜ] received the master·s degree in human·computer interaction from the Indian Institute of Information Technology Allahabad, Prayagraj, India, where he is currently pursuing the Graduate degree. He is also doing research on spatio-temporal dynamics of emotions. He has conducted two important experiments on Indian samples, which results in the availability of stimuli dataset (validated on an Indian sample) and the availability of EEG dataset with unique information about the time of emotional experience during watching the naturalistic multimedia stimuli. He is a member of the Society