IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 20 March 2023, accepted 6 April 2023, date of publ[redacted]tion 13 April 2023, date of current version 26 April 2023.
[redacted] 10.1109/ACCESS.2023.3266804

RESEARCH ARTICLE

[redacted] in [redacted] With Naturalistic Context: DENS# Dataset

[redacted], ([redacted] Member, [redacted]), [redacted], [redacted], AND [redacted], (Senior Member, [redacted])
[redacted], Allahabad, [redacted], India
Corresponding authors: [redacted] ([redacted]), [redacted] (pse2017001@iiita.ac.in), and [redacted] (ust@iiita.ac.in)

This work was supported by [redacted], funded by the acquisition of the [redacted] system.

ABSTRACT Emotion recognition using [redacted] is an emerging area of research due to its broad appl[redacted]bility in [redacted]. Emotional feelings are hard to stimulate in the lab. Emotions don't last long, yet they need enough context to be perceived and felt. However, most [redacted]-related emotion databases either suffer from emotionally irrelevant details (due to prolonged duration stimulus) or have minimal context, which may not elicit enough emotion. We tried to overcome this problem by designing an experiment in which participants were free to report their emotional feelings while watching the emotional stimulus. We called these reported emotional feelings "Emotional Events" in our Dataset on Emotion with Naturalistic Stimuli (DENS), which has the recorded [redacted] during the emotional events. To compare our dataset, we classify emotional events on different com[redacted]ations of [redacted](V) and Arousal(A) dimensions and compared the results with benchmark datasets of DEAP and SEED. Short-[redacted] ([redacted]) is used for feature extraction and in the classifi[redacted]tion model consisting of [redacted]-[redacted] hybrid layers. We achieved signifi[redacted]ntly higher accuracy with our data compared to DEAP and SEED data. We conclude that ha[redacted]ng precise information about emotional feelings improves the classifi[redacted]tion accuracy compared to long-duration recorded [redacted] which might be contaminated by mind-wandering. This dataset can be used for detailed analysis of specific experienced emotions and related brain dynamics.

INDEX TERMS Affective computing, [redacted], DEAP, DENS, [redacted], emotion dataset, emotion recognition, [redacted], SEED.

[redacted]

[redacted] recognition has been a challenging task in artificial intelligence. Several methods are available for measuring the participants' emotions. These methods include [redacted]oural changes, subjective experiences self-reported by the participants, peripheral and central nervous system measures, etc [1]. Brain acti[redacted]ties are among the most robust dimensions of detecting human affect, as it is difficult for the users to manipulate innate brain acti[redacted]ty during the [redacted]ess. Accordingly, Electroencephalography ([redacted]) is considered a suitable and convenient method to record electr[redacted]l acti[redacted]ties to measure brain acti[redacted]ties as it is a non-invasive method, i.e. there are no scalpel incisions.

Many studies have already been conducted to measure human affect with the help of [redacted] and other peripheral responses [2], [3], [4], [5]. In the pre[redacted]ous studies, the focus of the study was to develop a database that is labelled and suitable for emotion detection by intelligent systems and has contributed to affective computing. There is a typ[redacted]l method in these studies to elicit emotion in the participants by presenting them with [redacted]deo clips as stimuli. In the [redacted]ess of emotion recognition and other classifi[redacted]tion tasks, all the [redacted] data for that stimulus are

The associate editor coordinating the re[redacted]ew of this manuscript and appro[redacted]ng it for publ[redacted]tion was [redacted].

to be considered for the classifi[redacted]tion 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 pro[redacted]ding 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 [redacted] 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 statist[redacted]l signifi[redacted]nce.

[redacted] measures the electr[redacted]l signals from the scalp with temporal details. Different [redacted] de[redacted]ces vary with the number of channels of [redacted]. Thirty-two or fewer [redacted] 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 [redacted] de[redacted]ce to detect emotions. This [redacted] 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 com[redacted]es multi-disciplines such as [redacted]ogy, [redacted], computer science and medicine, etc. There are different aspects involved in determining emotions, such as [redacted]oural, [redacted]og[redacted]l and physiolog[redacted]l aspects, cognitive appraisals, facial expressions, vocal responses, subjective experiences, etc. This study focuses on physiolog[redacted]l aspects of emotion, which are considered into account by the brain signals captured through [redacted] while watching emotional [redacted]deo clips. Further, this study tries to collect a comprehensive list of subjective experiences through a self-assessment rating at the end of each clip.

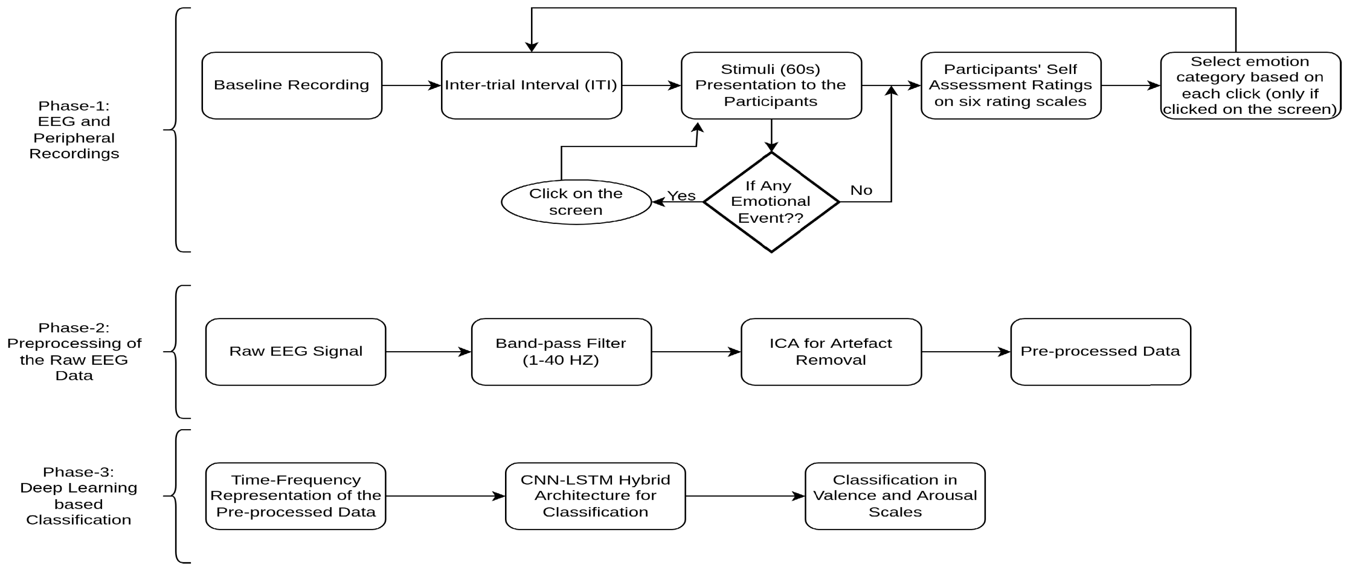
Many approaches could be used to assess the participants' emotional states. Earlier, some basic emotions were used that are universally recognised for study purposes [9]. Later, some theories explained some complex emotions that are a com[redacted]ation of basic emotions [10]. Multi-dimensional [redacted]s are the widely accepted theories for assessing core affect [11], [12]. According to these theories, emotions are considered a multi-dimensional array; one dimension is for [redacted] (experiencing positive or negative) and the other for arousal (experiencing the intensity) or [redacted] (controlling or feeling controlled). A few more dimensions are also considered, that make the spectrum broader, e.g., relevance (how much the stimulus is relevant to the participant's emotional feelings), familiarity (how much the participant is familiar with the stimulus) and liking (how much the participant liked or disliked the stimulus). Asking participants to report these experiences on a continuous scale is common in similar studies. Some theories deal with the physiolog[redacted]l responses of feeling emotions, e.g., body temperature and heartbeat change [13], [14]. It is ob[redacted]ous from the theories that emotion is not a one step [redacted]ess; instead, it is a com[redacted]ation of physiolog[redacted]l responses and other information. E[redacted]dence shows that many brain regions are involved during emotion perception [15]. We have also collected [redacted] and [redacted] data of the participants along with [redacted] to consider these parameters.

Emotion recognition through [redacted] data follows a similar pattern as used in various [redacted] signal analyses. First, the data is acquired, and some pre[redacted]essing is applied to the signal. These pre[redacted]essing steps involve remo[redacted]ng artefacts such as ocular acti[redacted]ty, muscle acti[redacted]ty, and powerline interference. Also, downsampling of the signal and band-pass filtering are used to make data more useful. Various dimensionality reduction techniques, such as [redacted] and [redacted], are also used to prune the data to make it feature-rich. After pre[redacted]essing, features are extracted from the signal to feed into the model for the classifi[redacted]tion task. Different kinds of features are extracted such as time-domain (e.g., event-related potential (ERP), high-order crossing ([redacted]), etc.), frequency-domain (e.g., power spectral density (PSD), etc.); and time-frequency domain (e.g., [redacted], wavelet analysis, etc.) features.

[redacted] records multi-frequency non-stationary brain signals from various electrodes. Analyzing these signals is challenging because of the complex and irregular nature of [redacted]. The time-frequency domain analysis has the benefits of both the time and frequency domains, e.g., better spatial and temporal information from [redacted]. One basic time-frequency domain feature extraction method is Short-[redacted] ([redacted]). [redacted] is a time-ordered sequence of spectral estimates and is one of the powerful and general-purpose signal [redacted]essing techniques. It has been used in the field of spectral analysis of a signal. The [redacted] is used to compute [redacted] which are used extensively for signal [redacted]essing. [redacted] are [redacted]sual representations of the spectrum of frequencies of a signal with varying times [16].

[redacted] is the most frequently used architecture for [redacted] analysis and classifi[redacted]tion tasks, and [redacted] and [redacted] follow it [17]. Hence we have used a com[redacted]ation of the [redacted] and [redacted] model. It also helped to compare our dataset with the benchmark datasets in terms of maximum classifi[redacted]tion accuracy. Using artificial intelligence for affective computing pro[redacted]des better learning capabilities to intelligent systems. With the advancement of computing power and the development of effective and advanced neural network research, the trend of using various machine learning and deep learning

IEEE Access



[redacted] 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 [redacted].

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, [redacted] emotion analysis and our dataset. In the next section, we introduced our proposed dataset- DENS, Emotional Events, experimental details (e.g., stimuli, [redacted] recordings, ratings etc.), pre[redacted]essing of the [redacted] data, its salience features and other datasets used (DEAP and SEED). In the methodology section, we discussed the feature extractions, input pre[redacted]essing 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. DATASET ON EMOTION WITH NATURALISTIC STIMULI (DENS)

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

A. *EMOTIONAL EVENT*

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 nature of emotions and pro[redacted]de a single emotional category for the whole stimulus duration. Although the stimulus has emotional information, it has some non-emotional aspects too, which could lead to mind-wandering acti[redacted]ty. Although there are some attempts to get continuous subjective feedback on emotional experience and neural acti[redacted]ty, the experimental method involved multiple watching of the stimulus and retrospective collection of emotional experience [22], [23], [24], [25], [26]. The retrospective collection depends on autobiograph[redacted]l memory and can raise biases across subjects depending on their capability to recall [27]. Also, repetitive [redacted]ewing effects can bias the ratings and underlying neural effects [27]. Hence, an experimental paradigm is needed to record the participants' feedback dynam[redacted]lly, with minimal distraction during emotion [redacted]essing and minimizing the memory recall biases. In this work, we are introducing a novel paradigm in which the time-stamp of emotional feelings can be marked online that can be further utilized to get the subjective feedback of emotional feelings and analyze brain signals temporally localized to the feeling of an emotion. We refer to these time-stamped emotional feelings as "emotional events".

B. *EXPERIMENTAL DETAILS*

1) STIMULI

The selection of stimuli to induce participants' emotions also plays a [redacted]tal role in emotion recognition. A careful selection

[redacted]. Selected stimuli for [redacted] 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.

Stimulus Name	Stimulus ID	Target Emotion
Ashayen	199	Adventorous
Horror	214	Afraid
Anacondas The Hunt For The Blood Orchid Clip	10	Alarmed
Lage Raho Munnabhai Only The Funny Scenes	98	Amused
Anger Legend of Baghat Singh	198	Angry
Divergent Kiss Scene Clip	54	Aroused
Butterfly Nets	40	Calm
Best Horror Kills Ghost Ship Opening Scene	26	Disgust
Jai Ho	92	Enthusiastic
Sadda Haq	152	Excited
Masoom	109	Happy
Cheerful Rang	201	Joyous
Crash Saddest Scene	51	Melancholic
Hate LBS	210	Miserable
Madari Movie Of Best Scene	113	Sad
Final Race of Milkha Singh Career	67	Triumphant

of stimuli is crit[redacted]l, and for that, techn[redacted]l validation of the [redacted]deo 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 pre[redacted]ous 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 [redacted] 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 [redacted] 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) [redacted] RECORDING

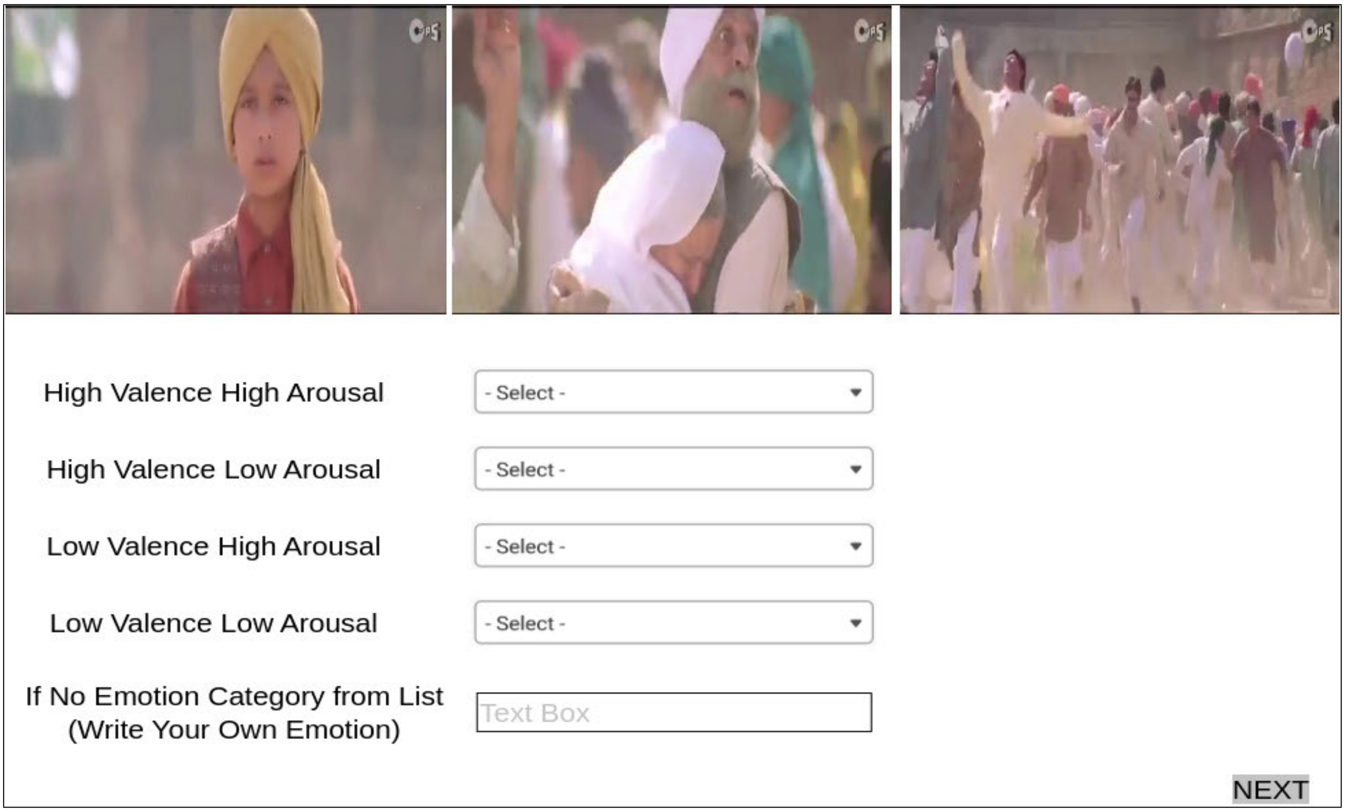
We recorded the [redacted] acti[redacted]ty of forty participants (23.3 ± 1.25, F=3) while they were watching emotional film stimuli. Following are some crit[redacted]l pieces of information regarding the experiment:

- Each participant saw nine (9) emotional stimuli randomly extracted from the set of 16 emotional stimuli and two (2) non-emotional stimuli as described in the pre[redacted]ous subsection.
- While watching the emotional film stimuli, participants were instructed to perform a mouse click the moment they felt any emotion. We call it an Emotional Event.
- At the end of each [redacted]deo stimuli, participants are pro[redacted]ded six self-assessment scales, including [redacted], arousal, [redacted], liking, familiarity, and relevance.
- For each click, participants were supposed to select one emotion from the pro[redacted]ded list of emotions pooled into four quadrants of V-A space ([redacted], [redacted], [redacted], [redacted]) (abbre[redacted]ations- V: [redacted], A: Arousal, H: High, L: Low) in the drop-down menu. Participants were also given a choice to enter the emotional category which suits their emotional experience but is unavailable in the pro[redacted]ded emotion list. For more details see [redacted].2.

Before the main experiment begins, participants go through the training phase. In the training phase, participants were given instructions about the experiment [redacted]edure, rating scales were properly explained by gi[redacted]ng them a small quiz, and also they were trained to mouse-click when they felt emotion during the stimulus.

The main experiment consists of the following steps for each participant:

1) Baseline Recording: [redacted] signal was recorded for 80 seconds while the participant looked at the cross-mark on the screen and performed no task.
2) After baseline recording, one stimulus of 60s was presented to the participant. Participants were told to click on the screen when they felt the emotion during the stimulus. Participants may click more than once if they felt so but were instructed to refrain from multiple clicks for the same emotion. [redacted] were recorded during this phase.
3) After the stimulus ends, participants go through self-assessment ratings of [redacted], arousal, [redacted], liking, familiarity, and relevance. These scales are explained in detail in the next subsection.
4) At last, participants were supposed to select one emotion category for each click (emotional event). To help the participants in recalling about the click, they were presented with three frames around the click.
5) After this, an inter-stimulus interval comes with no time limit. During this interval, participants were given a quick and easy mathemat[redacted]l calculation (e.g., 2+5*2=?). It helps participants to flush their pre[redacted]ous emotional state.
6) After that, the next stimulus is presented to the participant, and steps 1 to 5 are followed similarly for each stimulus. A total of 11 stimuli (9 emotional and 2 non-emotional) were presented to each participant.



[redacted] 2. Emotion Category Selection Screen for Emotional Event (Click): After the participants rated all the six rating scales of [redacted], Arousal, [redacted], Liking, Familiarity and Relevance, they are shown this screen for emotion category selection. On this screen, three image frames were shown. The middle one belongs to the time of the click; the left one is 20 frames earlier, and the right one is 20 frames after the click (Please note that the stimulus clips were shown in 30 frames per second). It helps participants to recall easily. They only have to select one emotion category. If the 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/[redacted]deos 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 [redacted], Arousal, [redacted], Liking, Familiarity and Relevance. The rating scales range from 1 to 9 for [redacted], Arousal and [redacted]. For Liking, familiarity and Relevance, it ranges from 1 to 5. Although, in this analysis, we considered only [redacted] and arousal scales.

4) SUMMARY OF THE [redacted]

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, [redacted] 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. PRE[redacted]ESSING AND ARTIFACT REMOVAL OF THE [redacted] DATA

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

D. OTHER DATASETS USED

We have used DEAP dataset [2] (a dataset for emotion analysis using [redacted], physiolog[redacted]l and [redacted]deo signals) and SEED dataset [3] (A dataset collection for various purposes using [redacted]) for comparing the results with our dataset (DENS).

The DEAP dataset consisted of 40 [redacted]deos/trials, and for each trial, there are 40 channels of [redacted], including peripheral signals, are available, and data is given for each channel. We have used only 32 channels (i.e., discarded peripheral signals) for the experiment as we only want to use data from the brain only. This data was already pre[redacted]essed as 128 Hz downsampled, bandpass frequency of 4-45 Hz and [redacted] removed. For each trial, there are 4 labels available- [redacted] (V), Arousal (A), [redacted] 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 [redacted] (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. SOME SALIENT FEATURES OF THE DENS

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 [redacted] from participants in the Indian subcontinent.
- Stimuli that are used to record [redacted] 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 [redacted] 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. FEATURE EXTRACTION

[redacted] are non-stationary, meaning the signal's statist[redacted]l characteristics change over time. If these signals are transformed to the frequency domain using [redacted], it pro[redacted]des the frequency information, which is averaged over the entire [redacted] 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 [redacted] on it, it is called as Short-[redacted] ([redacted]). It will move to the entire signal length and apply [redacted] to find the spectral content of that section and display the coefficient as a function of both time and frequency. It pro[redacted]des insight into the nature of the time-varying spectral characteristics of the signal. Before [redacted], let's look at the discrete [redacted]. Consider x: [0: L-1] = $\{0, 1, \ldots, L-1\} \rightarrow$ R be a discrete-time signal where L is the signal length which is acquired by equidistance sample points with respect to the fixed sampling frequency. Mathemat[redacted]lly [redacted] equation is,

$$\widehat{x}(k) = \sum_{n=0}^{N-1} x(n).e^{-i2\pi nk/N} \tag{1}$$

where k ϵ [0, K] and K is the frequency index with respect to [redacted] frequency. N is the duration of the section. The equation returns the complex Fourier coefficients for the kth. These coefficients pro[redacted]de two parameters: phase and magnitude. For [redacted], consider the additional parameter hop size (H), which is the step size of the window to be shifted. ω be a sampling window function which is ω: [0, N-1] $\rightarrow$ R. [redacted] can be defined as,

$$S(m, k) := \sum_{n=0}^{N-1} x(n + mH)\omega(n)e^{(-i2\pi kn/N)} \tag{2}$$

where m ϵ [0,M] and M is the maximum frame index mathemat[redacted]lly M $= \lfloor \frac{L-N}{H} \rfloor$. The Short-[redacted] is not only a function of k but also m which is a proxy time representation. Here, the function returns Fourier coefficients for the kth proxy frequency at the mth temporal [redacted].

[redacted] are nothing but the squared magnitude of [redacted] of the signal.

$$\chi(m, k) := |S(m, k)|^2 \tag{3}$$

It is a 2D image where the horizontal axis represents time, and the vert[redacted]l axis represents frequency [redacted]s. The number of frequency [redacted]s is (framesize / 2) + 1 and the number of time frames is ((size of signal − framesize) / hopsize) + 1. $\chi(m, k)$ represents intensity or color at (m, k).

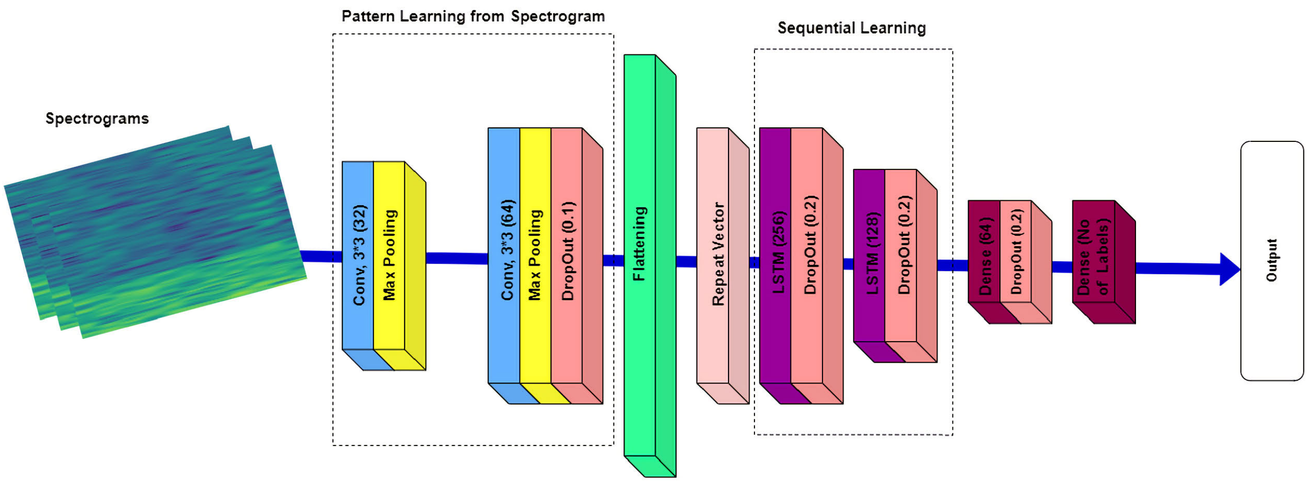
B. INPUT PRE[redacted]ESSING TO FEED DATA INTO THE CLASSIFIER

It is essential to convert the data into a meaningful format that can be fed into our classifier model. As all three datasets are available in different formats, we have pro[redacted]ded information on the input pre[redacted]essing for each dataset as follows:

1) FOR DEAP DATA

Each subject in the DEAP dataset is given by a tensor that is in the form of [redacted], representing 40 [redacted]deos, 40 [redacted]-channels (including peripheral channels), and 8064 [redacted] data samples for each channel. For labels, DEAP data pro[redacted]ded a matrix in the form of X $\in \mathbb{R}^{40\times4}$; i.e., for each subject, there are 40 [redacted]deos and 4 scales. From the DEAP dataset, the first 15 subjects are picked. For the label, we used the [redacted]-Arousal space and di[redacted]ded it into four classes- [redacted], [redacted], [redacted], [redacted] (abbre[redacted]ations- H: High, L: Low, V: [redacted], A: Arousal). The ratings for [redacted] and arousal range from 1 to 9. Hence, we considered ratings from 1 to 5 as 'Low' and 5 to 9 as 'High' and di[redacted]ded the V-A space into 4 quadrants accordingly.

We converted 15 subjects' data tensor into a matrix of [redacted] (*i.e.*, 15 *subjects* × 40 [redacted]*deos* × 32 *channels*, 8064 *samples*). Moreover, this data was [redacted]essed for feature extraction using [redacted] with a window size of 0.5s and an overlap of 0.25s of data samples. Using [redacted], we have converted every 8064 sizes of [redacted] data samples into a spectrogram image size of (33,251), as mentioned in the feature extraction section. Then, a hybrid [redacted]-[redacted] classifier was implemented for multi-class classifi[redacted]tion with an input tensor of X $\in \mathbb{R}^{33\times251\times3}$.



[redacted] 3. Model Architecture: It is consisted of two 2D-convolution layers with 3 × 3 kernels and 32 filters and 64 filters respectively, followed by a max pooling layer followed by a dropout layer and flattening layer. A repeat vector layer of size 4 is used before sending the data to the [redacted] layers. Two [redacted] layers are used of sizes 256 units and 128 units respectively, each followed by a dropout layer. At the end, two dense layers are used of sizes 64 (followed by a dropout layer) and 4 or 3 (equals the number of the output classes).

2) FOR SEED DATA

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 [redacted] scale. After renaming, the labels become 0 for neutral, 1 for positive, and 2 for negative. For classifi[redacted]tion, 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 [redacted]essing. [redacted] cap includes 62 channels according to the 10-20 international system. So, 15 subjects, 15 trials, 62-channels, and 16000 [redacted] data are converted into a tensor of $X \in \mathbb{R}^{13950\times 16000}$ ($i.e.$, $15\ subjects \times 15\ trials \times 62\ channels$, $16000\ samples$)for feature extraction. As mentioned in the DEAP dataset experiment, using [redacted] with a window size of 0.5s and overlap of 0.25s, each 16000 [redacted] data is converted into a spectrogram with the shape of (51,319). Then, a hybrid [redacted]-[redacted] classifier was implemented for multi-class classifi[redacted]tion with input tensor shape $X \in \mathbb{R}^{51\times 319\times 3}$.

3) FOR DENS DATA

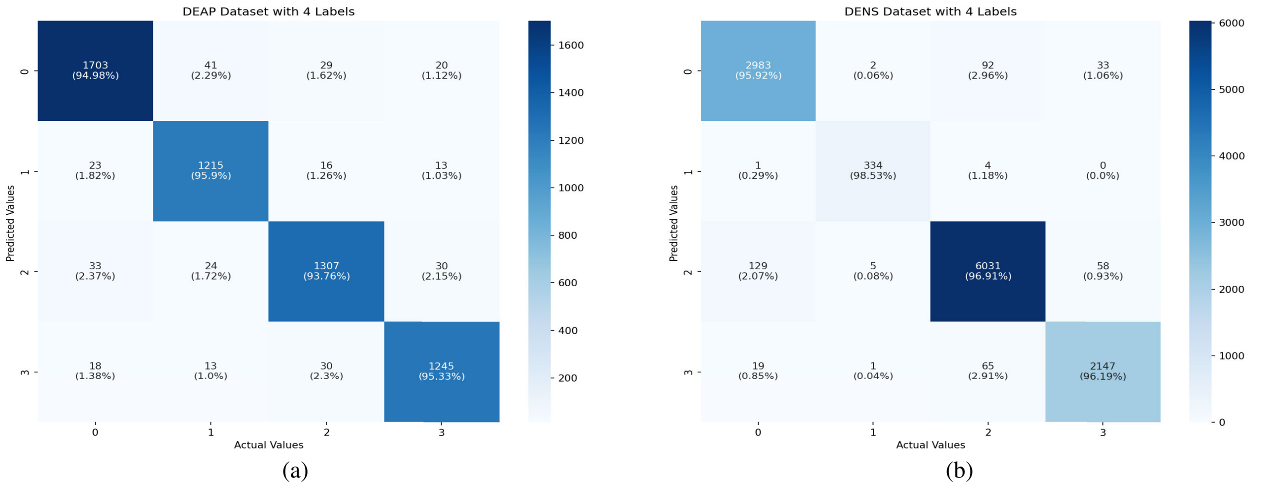
For the DENS dataset, we have 465.mat files which contain emotional events. All 465 files are picked for the experiment. [redacted] file is a matrix of $X \in \mathbb{R}^{128\times 1751}$, where 128 is the number of [redacted] channels and 1751 is the sample data for each channel. Then we have converted the data tensor of $X \in \mathbb{R}^{465\times 128\times 1751}$ into the form of $\mathbb{R}^{59520\times 1751}$ ($i.e.$, $465\ emotional\ events \times 128\ channels$, $1751\ samples$) for feature extraction with window size 0.5s and overlap is 0.25s. After feature extraction, we have 59520 [redacted], and each spectrogram is in the shape of (63, 26).

To compare with the DEAP dataset, the DENS dataset with 4-label classifi[redacted]tion is performed with a hybrid [redacted]-[redacted] classifier. For the label, we used the same V-A space ([redacted], [redacted], [redacted], [redacted]) (abbre[redacted]ations- H: High, L: Low, V: [redacted], A: Arousal) as it was used with the DEAP dataset. The ratings for [redacted] and arousal range from 1 to 9. Hence, we considered ratings from 1 to 5 as 'Low' and 5 to 9 as 'High' and di[redacted]ded the V-A space into 4 quadrants accordingly. The dimension of input tensor is of $X \in \mathbb{R}^{63\times 26\times 3}$.

To compare with the SEED dataset, we have used a 3-label classifi[redacted]tion since there are only three classes available in SEED dataset. For the DENS dataset, on the [redacted] scale, ratings below 4.5 are marked as negative (0 labelled), and ratings above 5.5 are marked as positive (2 labelled). For neutral labels, in the DENS dataset, we have non-emotional files; we have marked neutral (1 labelled) for those files' data. Then with the classifier, the input tensor of $X \in \mathbb{R}^{63\times 26\times 3}$ is used for classifi[redacted]tion.

C. MODEL ARCHITECTURE FOR THE CLASSIF[redacted]TION TASK

[redacted] ([redacted]) and [redacted] Memory ([redacted]) are one of the most widely used deep learning techniques. [redacted]s are used to extract meaningful patterns and features from the data. The key element in [redacted] is the convolution operation using kernels that automat[redacted]lly learn the local patterns from data. These local features are then com[redacted]ed into more complex features when multiple [redacted] layers are stacked. Filters (i.e, weights trained) in this [redacted]ess are also known as feature detector matrices. Input data will be convoluted with a filter map by sliding the kernel window. At the same time, [redacted] networks can capture the sequential pattern as [redacted]s are best suited for time-series data. [redacted]s are designed to work for temporal correlations. Therefore, to exploit the benefits of both [redacted] and [redacted], a hybrid [redacted]-[redacted] architecture is used for



[redacted] 4. Comparison of [redacted]usion matrices for DEAP and DENS datasets over [redacted]-Arousal space. This space is di[redacted]ded into four classes and assigned a label to it (0-[redacted], 1-[redacted], 2-[redacted] and 3-[redacted]). 4a: DEAP Dataset; 4b: DENS Dataset. Abbre[redacted]ations of the terms- V:[redacted]; A:Arousal; L:Low; H:High. The color bar represents the number of samples in the class.

the classifi[redacted]tion of emotions. The hybrid [redacted]-[redacted] model utilizes the ability of convolutional layers for feature extraction from data, and [redacted] 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 [redacted]. 3.

[redacted] 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. [redacted] (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 [redacted] layer. The sequential learning block consists of 2 [redacted] layers which capture the long-term temporal dependencies from the feature map extracted by [redacted] layers. 1^{st} [redacted] layer consists of 256 cells with a return sequence set to 'True' while 2^{nd} [redacted] consists of 128 cells and as it is the last [redacted] layer return sequence is 'False'. Between [redacted] layers, dropout layers with rate $= 0.2$ are added to avoid overfitting issues. Finally, two fully-connected layers where 1^{st} layer with 64 neurons and 2^{nd} layer with the number of classes as neurons are added for further [redacted]essing. As we have the multi-class classifi[redacted]tion, the [redacted] activation function is used in the output layer as it outputs a vector representing the probability distributions of a list of potential classes.

[redacted]. Parameter Settings for the Model.

Parameter	Setting
Optimizer	Adam
Loss function	Categorical Cross-entropy
Learning rate	0.001
Adjustment	Early Stopping criteria: monitor - 'val_loss'; patience = 30 Model Checkpoint: monitor - 'val_accuracy'
Batch size	256
Epochs	100

The parameter setting for the developed deep learning model is mentioned in [redacted].

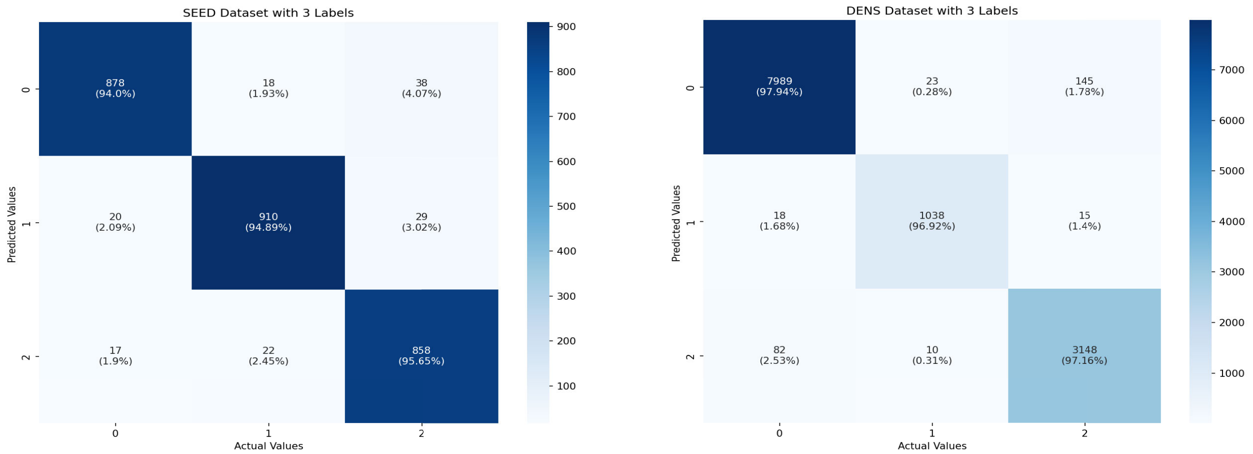
IV. RESULTS

The [redacted]usion matrix for DEAP, SEED and DENS datasets are shown in [redacted]. 4 and [redacted]. 5. In the [redacted]usion matrix shown, each cell contains data on the number of population. The X-axis represents actual labels and the Y-axis represents predicted labels by the classifier. The diagonal of the matrix represents the correctly identified label. The color bar represents the number of samples in the class.

A. COMPARISON BETWEEN DEAP AND DENS

We have used repeated [redacted] cross-validation with $K = 5$ and the number of repeats $= 5$ so generated 25 accuracies for DENS and DEAP. For label classifi[redacted]tion, we have used V-A space ([redacted], [redacted], [redacted], [redacted]). Comparison between DEAP and DENS is mentioned in [redacted]. The loss and accuracy graphs are mentioned in [redacted]. 8.

[redacted]. 6 shows an [redacted] score comparison between DEAP and DENS datasets per trial. Using t-test statist[redacted]l testing, the 25 [redacted] scores of DEAP dataset (M = 95.65%, SD = 0.38%) compared with the 25 [redacted] scores of DENS dataset (M = 96.82%, SD = 0.18%), DENS dataset shows better results with absolute t (35) = 13.54, $p < 0.0001$,

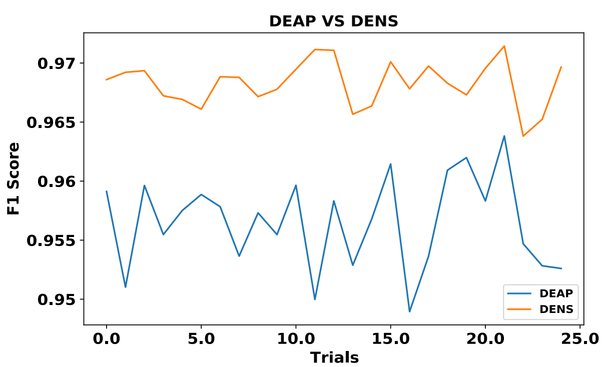


(a) (b)

[redacted] 5. Comparison of [redacted]usion matrices for SEED and DENS datasets over [redacted] space. This space is di[redacted]ded into three classes (SEED dataset pro[redacted]ded data with three classes, while DENS data is di[redacted]ded into three classes based on the [redacted] ratings pro[redacted]ded by the participants) and assigned a label to it as follows: For SEED:0 for neutral, 1 for positive and 2 for negative. For DENS: 0 for low-[redacted] ([redacted] ratings range from 1-4.5), 1 for non-emotional data ([redacted] ratings range from 4.5-5.5, as well as neutral categories stimuli) and 2 for high-[redacted] ([redacted] ratings ranges from 5.5-9). 5a: SEED Dataset; 5b: DENS Dataset. The color bar represents the number of samples in the class.

TABLE 3. Comparison Table with Other Recent Studies.

Method	Dataset	Subject Dependency	Emotion Classes	Result Accuracy (%)
CNN-RNN Hybrid Model [30]	DEAP	Subject Dependent	2	Valence: 72.06 Arousal: 74.12
R2G-STNN Model (region to global BiLSTM with Attention Layer) [31]	SEED	Both	3	Sub. Dependent: 93.38 Sub. Independent: 84.16
ACRNN (Attention Based C-RNN Model) [32]	DEAP	Subject Dependent	2	Valence: 93.72 Arousal: 93.38
BiDCNN (Bi-hemisphere Discrepancy CNN model) [33]	DEAP	Both	2	Sub. Dependent: Valence- 94.38, Arousal- 94.72 Sub. Independent: Valence- 68.14, Arousal- 63.94
ECLGCNN (A fusion model of GCNN + LSTM) [34]	DEAP	Both	2	Sub. Dependent: Valence- 90.45, Arousal- 90.60 Sub. Independent: Valence- 84.81, Arousal- 85.27
Our Work (CNN-RNN Hybrid Model using STFT)	DENS DEAP SEED	Subject Dependent	3 and 4	Valence (3 Classes): DENS- 97.68, SEED- 95.65 V-A Space (4 Classes): DENS- 96.82, DEAP- 95.65



[redacted]URE 6. [redacted] scores of DEAP vs DENS for all the 25 trials.

d estimate: −13.70 (large), 95 percent [redacted]dence interval: [−16.51 −10.89].

[redacted]. DEAP vs DENS with mean [redacted] scores.

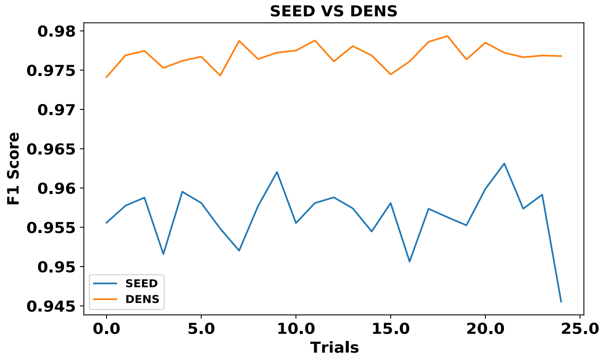
Dataset	Mean F1 score (in %)
DEAP	95.65 (± 0.38)
DENS	96.82 (± 0.18)

[redacted]. SEED vs DENS with mean [redacted] scores.

Dataset	Mean F1 scores (in %)
SEED	95.65 (± 0.37)
DENS	97.68 (± 0.13)

B. COMPARISON BETWEEN SEED AND DENS

For SEED vs DENS comparison, label classifi[redacted]tion we have used 3 labels on the [redacted] scale. Comparison between SEED and DENS results is mentioned in [redacted]. The loss and accuracy graphs are mentioned in [redacted]. 8.



[redacted] **7.** [redacted] **scores of SEED vs DENS for all the 25 trials.**

[redacted]. 7 shows an [redacted] score comparison between SEED and DENS datasets per trial. Using t-test statist[redacted]l testing, the 25 [redacted] score of SEED dataset (M = 95.65%, SD = 0.37%) compared with the 25 [redacted] score of DENS dataset (M = 97.68%, SD = 0.13%), DENS dataset shows better results with absolute t (31) = 25.466, $p < 0.0001$, [redacted]'s d estimate: −11.37 (large), 95 percent [redacted]dence interval: [−13.73 −9.02].

C. COMPARISON WITH OTHER RECENT STUDIES

We have included some other recent studies and given a comparative table for their results in Table 3. The studies consist of [redacted] Hybrid models, R2G-STNN model that is based on regional to global Bi[redacted] with Attention layer, Attention-based [redacted] Hybrid model ([redacted]), BiD[redacted] that is Bi-hemisphere Discrepancy [redacted] model and [redacted] that is a fusion model of [redacted] and [redacted] model.

V. DISCUSSION

In this work, we captured emotional experiences within the ecolog[redacted]lly valid naturalistic en[redacted]ronment 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 connecti[redacted]ty among these networks might not be the same in different perceptions which is apparently shown in our pre[redacted]ous 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 [redacted]used with other perceptions, which might not be an emotional experience.

One of the major concerns is the mind-wandering acti[redacted]ty while using the film stimuli. In the pre[redacted]ous 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 particular) [36]. Hence, it is important to know the duration of the emotional experience without compromising the ecolog[redacted]l validity of the stimuli.

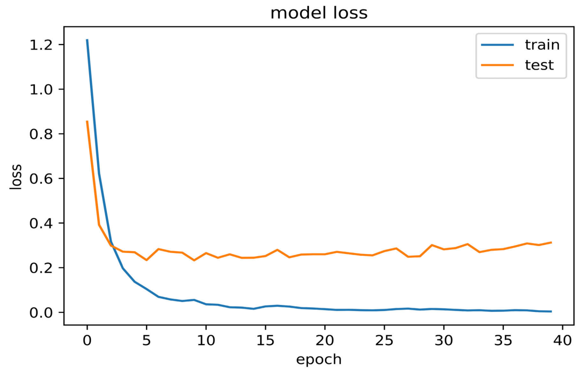
The main idea behind this work is that if we can capture the temporal marker of emotional experience within a real-life resembling en[redacted]ronment, we might achieve better accuracy than the accuracy achieved to date with other datasets lacking information about time. Although, due to the limited number of subjects, we didn't go for the subject-independent classifi[redacted]tion for now. Though, in future, we will be collecting more data to mitigate this limitation.

In our results, we observed that the same hybrid deep learning model on our dataset not only outperformed other datasets, including benchmark datasets like DEAP and SEED but also achieved a better result when comparing with other noteworthy relevant studies (see Table 3). Classifi[redacted]tion of DEAP data into four labels, including [redacted], [redacted], [redacted], and [redacted], resulted in 95.65% mean accuracy. At the same time, the classifi[redacted]tion of DENS data into four labels resulted in 96.82% mean accuracy. Similarly, the classifi[redacted]tion of SEED data into three labels resulted in 95.65% mean accuracy while DENS data resulted in 97.68% mean accuracy. The signifi[redacted]nce testing showed that even with multiple iterations, the classifi[redacted]tion accuracy was signifi[redacted]ntly higher for our data.

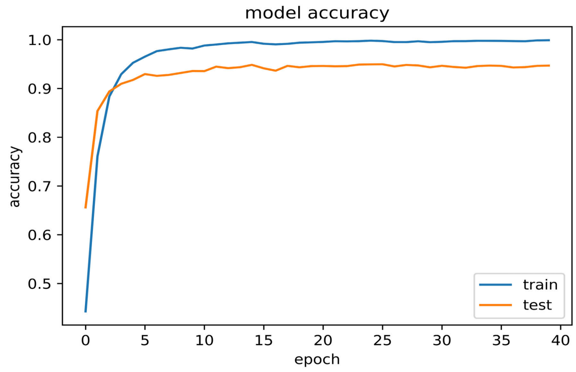
To the date, most of the work on emotion recognition applied different shallow machine learning and deep learning techniques using many different [redacted]urations of input data including, spectrogram, raw signals, statist[redacted]l features, variational mode decomposition ([redacted]), empir[redacted]l mode decomposition ([redacted]), functional connecti[redacted]ty based features, fractal features and so on. However, still, the recognition of emotion from [redacted] stands as a problem. Most of the works on emotion recognition have used some benchmark datasets, including DEAP, SEED, [redacted], MAHNOB-HCI and so on. Though, most of the emotion classifi[redacted]tion works revolve around DEAP and SEED datasets [2].

In [37], emotional states are classified by means of [redacted]-based functional connecti[redacted]ty patterns. Forty participants [redacted]ewed audio-[redacted]sual film clips to evoke neutral, positive (one amusing and one surprising) or negative (one fear and one disgust) emotions. Correlation, coherence, and phase synchronization are used for estimating the connecti[redacted]ty indices. They stated signifi[redacted]nt differences among emotional states. A maximum classifi[redacted]tion rate of 82% was reported when the phase synchronization index was used for connecti[redacted]ty measure.

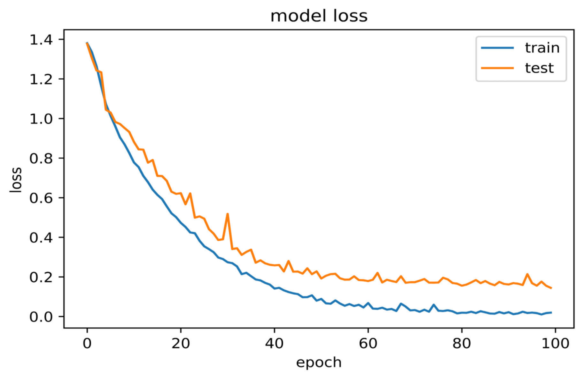
The classes considered in the study are elementary. We suspect that with the increasing number of emotional classes, which includes not only basic classes but complex emotions as well, taking the long-duration signal without a temporal marker may not be able to categorize emotional classes. The reason is that there are fewer chances for a mo[redacted]e stimulus to have a positive as well as a negative emotional



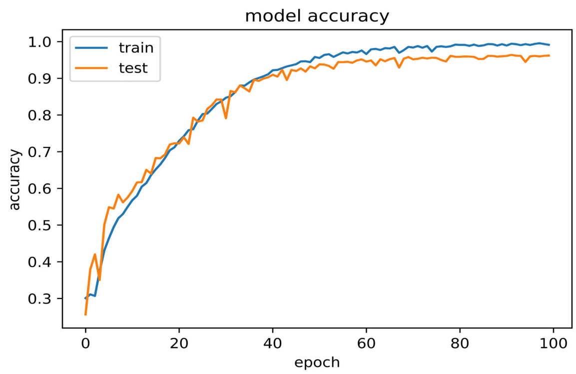
(a) Loss Graph for DEAP dataset



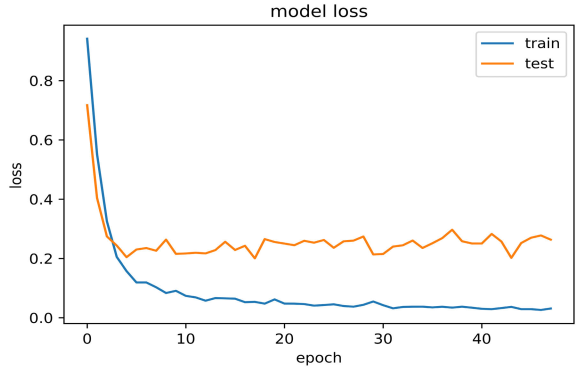
(b) Accuracy Graph for DEAP dataset



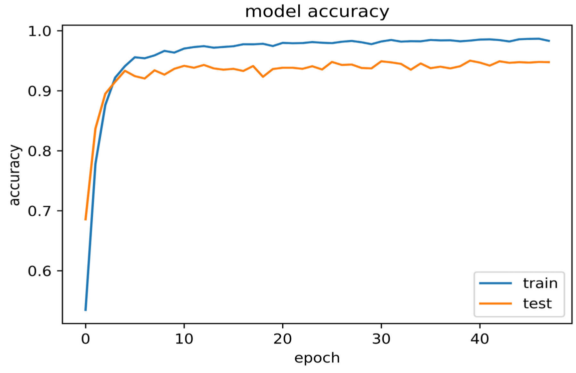
(c) Loss Graph for DENS (4 Classes) dataset



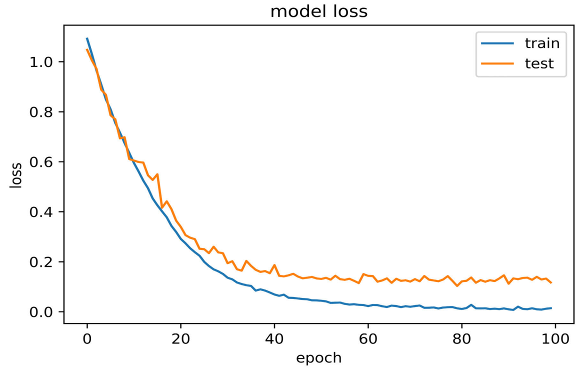
(d) Accuracy Graph for DENS (4 Classes) dataset



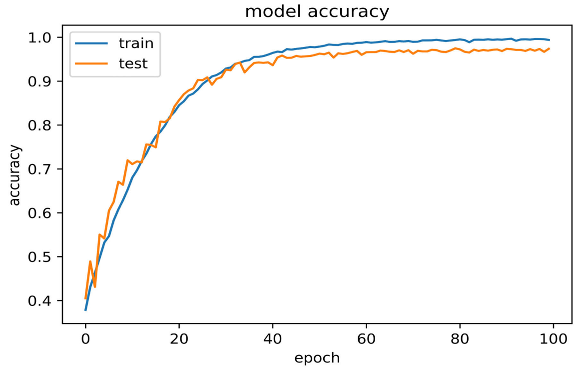
(e) Loss Graph for SEED dataset



(f) Accuracy Graph for SEED dataset



(g) Loss Graph for DENS (3 Classes) dataset



(h) Accuracy Graph for DENS (3 Classes) dataset

[redacted] **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 mo[redacted]e.

[redacted]. 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 [redacted] 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 ecolog[redacted]l validity of the stimuli will contain more emotional information. To test the hypothesis, we designed an [redacted] experiment which uniquely marks the duration of the emotional event in the continuous recording of brain waves using [redacted]. We performed deep learning analysis using a hybrid [redacted] and [redacted] model and found results that signifi[redacted]ntly 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 [redacted] 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). *([redacted] and [redacted] contributed equally to this work.)*

REFERENCES

[1] [redacted] and [redacted], "Measures of emotion: A re[redacted]ew," [redacted], vol. 23, no. 2, pp. 209–237, Feb. 2009.

[2] [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "DEAP: A database for emotion analysis; using physiolog[redacted]l signals," [redacted] *Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[3] W.-[redacted] and [redacted], "Investigating crit[redacted]l frequency bands and channels for [redacted]-based emotion recognition with deep neural networks," [redacted] *Auto. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[4] S. Katsigiannis and [redacted], "DRE[redacted]: A database for emotion recognition through [redacted] and [redacted] signals from wireless low-cost off-the-shelf de[redacted]ces," [redacted] [redacted] *Health Informat.*, vol. 22, no. 1, pp. 98–107, Jan. 2018, doi: 10.1109/JBHI.2017.2688239.

[5] [redacted], [redacted], [redacted], and [redacted], "[redacted]S: A dataset for affect, personality and mood research on indi[redacted]duals and groups," [redacted] *Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Apr. 2021, doi: 10.1109/TAFFC.2018.2884461.

[6] [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "Recognition of human emotions using [redacted]: A re[redacted]ew," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104696.

[7] [redacted] and [redacted], "Determination of the hydrocel geodesic sensor nets' average electrode positions and their 10–10 international equivalents," [redacted], [redacted], OR, USA, Tech. Note, 2005, pp. 1–11.

[8] [redacted], *Human Emotions*. New York, NY, USA: [redacted].

[9] [redacted], "Basic emotions, natural kinds, emotion schemas, and a new paradigm," *Perspect.* [redacted] [redacted], vol. 2, no. 3, pp. 260–280, Sep. 2007.

[10] R. Plutchik, "A general psychoevolutionary theory of emotion," in [redacted]. Amsterdam, The Netherlands: Else[redacted]er, 1980, pp. 3–33.

[11] [redacted], "Core affect and the [redacted]og[redacted]l construction of emotion," [redacted] *Rev.*, vol. 110, no. 1, p. 145, 2003.

[12] [redacted] and [redacted], "Multimodal fusion framework: A multiresolution approach for emotion classifi[redacted]tion and recognition from physiolog[redacted]l signals," [redacted], vol. 102, pp. 162–172, Nov. 2014.

[13] [redacted], "The theory of emotion: I: Emotional attitudes," [redacted] *Rev.*, vol. 1, no. 6, pp. 553–569, Nov. 1894.

[14] [redacted], "The [redacted]–[redacted] theory of emotions: A crit[redacted]l examination and an alternative theory," [redacted] [redacted], vol. 39, [redacted] 1–4, p. 106, Dec. 1927.

[15] T. Dalgleish, "The emotional brain," *Nature Rev.* [redacted], vol. 5, pp. 583–589, Jul. 2004.

[16] [redacted], "Appl[redacted]tions of the short [redacted] to speech [redacted]essing and spectral analysis," in [redacted] [redacted] *Int.* [redacted] *Acoust.*, [redacted] ([redacted]), May 1982, pp. 1012–1015, doi: 10.1109/[redacted]1982.1171703.

[17] [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "Deep learning-based electroencephalography analysis: A systematic re[redacted]ew," [redacted], vol. 16, no. 5, Oct. 2019, Art. no. 051001.

[18] [redacted], *Deep Learning With Python*. New York, NY, USA: [redacted] and Schuster, 2021.

[19] [redacted], "Comment: Affective chronometry has come of age," [redacted], vol. 7, no. 4, pp. 368–370, Oct. 2015.

[20] [redacted], [redacted], [redacted], and [redacted], "Dataset on emotion with naturalistic stimuli (DENS) on Indian samples," *bioRxiv*, pp. 1–11, Dec. 2022. [Online]. Available: https://www.biorxiv.org/content/early/2022/12/31/2021.08.04.455041, doi: 10.1101/2021.08.04.455041.

[21] [redacted], [redacted], and [redacted], "Context in emotion perception," *Current Directions* [redacted] [redacted], vol. 20, no. 5, pp. 286–290, Oct. 2011, doi: 10.1177/0963721411422522.

[22] [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "Neural, electrophysiolog[redacted]l and anatom[redacted]l basis of brain-network variability and its characteristic changes in mental disorders," *Brain*, vol. 139, no. 8, pp. 2307–2321, Aug. 2016.

[23] C. B. Young, [redacted], D. Everaerd, [redacted], [redacted], [redacted], [redacted], and [redacted], "Dynamic shifts in large-scale brain network balance as a function of arousal," [redacted], vol. 37, no. 2, pp. 281–290, Jan. 2017.

[24] [redacted], C. Wilson-Mendenhall, [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "Functional connecti[redacted]ty dynamics during film [redacted]ewing reveal common networks for different emotional experiences," *Cogn., Affect.*, [redacted] [redacted], vol. 16, pp. 709–723, May 2016.

[25] [redacted], [redacted], [redacted], and [redacted], "Dynamic intersubject neural synchronization reflects affective responses to sad music," [redacted], vol. 218, Sep. 2020, Art. no. 116512.

[26] [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "Default and control network connecti[redacted]ty dynamics track the stream of affect at multiple timescales," *Social Cogn. Affect.* [redacted], vol. 17, no. 5, pp. 461–469, May 2022.

[27] [redacted], [redacted], [redacted], and [redacted], "Repeated mo[redacted]e [redacted]ewings produce similar local acti[redacted]ty patterns but different network [redacted]urations," [redacted], vol. 142, pp. 613–627, Nov. 2016.

[28] [redacted], [redacted], [redacted], and [redacted] (Nov. 2021). *Affective Film Dataset From India* ([redacted]): [redacted] *an Indian Sample*. [Online]. Available: https://psyarxiv.com/yajsk

[29] [redacted], [redacted], and [redacted], "Cardiac–brain dynamics depend on context familiarity and their interaction predicts experience of emotional arousal," *Brain* [redacted] vol. 12, no. 12, p. 702, 2022.

[30] [redacted], [redacted], [redacted], [redacted], and [redacted], "Emotion recognition from multi-channel [redacted] data through convolutional recurrent neural network," in [redacted] [redacted] *Int.* [redacted] *Bioinf. Biomed.* ([redacted]), Dec. 2016, pp. 352–359.

[31] [redacted], [redacted], [redacted], [redacted], and [redacted], "From regional to global brain: A novel hierarch[redacted]l spatial–temporal neural network model for [redacted] emotion recognition," [redacted] *Affect. Comput.*, vol. 13, no. 2, pp. 568–578, Apr. 2022.

[32] [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "[redacted]-based emotion recognition [redacted] a channel-wise attention and self attention," [redacted] *Affect. Comput.*, vol. 14, no. 1, pp. 382–393, Jan. 2023.

[33] [redacted], [redacted], [redacted], [redacted], [redacted], and [redacted], "Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for [redacted] emotion recognition," [redacted], vol. 448, pp. 140–151, Aug. 2021.

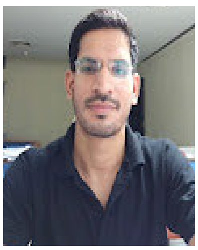
[34] [redacted], [redacted], [redacted], [redacted], and [redacted], "[redacted] emotion recognition using fusion model of graph [redacted] and [redacted]," *Appl. [redacted]*, vol. 100, Mar. 2021, Art. no. 106954.

[35] [redacted], [redacted], and [redacted], "Dynamic functional connecti[redacted]ty of emotion [redacted]essing in beta band with naturalistic emotion stimuli," *Brain [redacted]*, vol. 12, no. 8, p. 1106, Aug. 2022.

[36] [redacted], "Naturalistic stimuli in affective neuroimaging: A re[redacted]ew," *Frontiers Hum. [redacted]*, vol. 15, p. 318, Jun. 2021.

[37] [redacted] and [redacted], "Classifying different emotional states by means of [redacted]-based functional connecti[redacted]ty patterns," [redacted] *ONE*, vol. 9, no. 4, Apr. 2014, Art. no. e95415.

[redacted] ([redacted] Member, [redacted]) 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 [redacted], Allahabad. His research interest includes affective computing. He is also working on emotion recognition using brain signals. He is using [redacted] for emotion detection using validated stimuli. He is also working on deep learning architectures.



[redacted] received the master's degree in human–computer interaction from [redacted], [redacted], 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 [redacted] dataset with unique information about the time of emotional experience during watching the naturalistic multimedia stimuli. He is a member of the [redacted] for [redacted].

[redacted] was born in Jamnagar, Gujarat, India, in June 1995. He is currently pursuing the M.Tech. degree in IT with a specialization in machine learning and intelligent systems with [redacted], Allahabad. His research interests include machine learning, deep learning, and its appl[redacted]tion in cognitive science. He has two years of work experience as a Software Engineer with [redacted], Pune, India.

[redacted] (Senior Member, [redacted]) received the Ph.D. degree from the Department of [redacted], [redacted], Varanasi, India, in 1991. He was a Lecturer with [redacted], from September 1988 to March 1992. From March 1992 to June 2002, he was a Reader in computer science with the [redacted], [redacted]. He was also a [redacted]siting Scientist with [redacted] and Engineering, IIT Kanpur, from December 1995 to July 1996. He was an Associate Professor with [redacted], Allahabad, India, from July 2002 to December 2006, where he has been a Professor with [redacted], since December 2006. He is holding research and teaching experience for more than 30 years, in which he is very much involved in image [redacted]essing, computer [redacted]sion, med[redacted]l image [redacted]essing, pattern recognition and script analysis, digital signal [redacted]essing, speech and language [redacted]essing, wavelet transforms, [redacted]ing and fuzzy logic, [redacted] and soft-computers, speech-driven computers, natural language [redacted]essing, brain simulation, cognitive science, and affective computing.