


Affective Video Tagging Framework using Human Attention Modelling through EEG Signals

Shanu Sharma, Amity School of Engineering and Technology, Amity University, Noida, India*

 <https://orcid.org/0000-0003-0384-7832>

Ashwani Kumar Dubey, Amity School of Engineering and Technology, Amity University, Noida, India

Priya Ranjan, Bhubaneswar Institute of Technology, India

ABSTRACT

The explosion of multimedia content over the past years is not surprising; thus, their efficient management and analysis methods are always in demand. The effectiveness of any multimedia content deals with analyzing human perception and cognition while watching it. Human attention is also one of the important parameters, as it describes the engagement and interestingness of the user while watching that content. Considering this aspect, a video tagging framework is proposed in which the EEG signals of participants are used to analyze human perception while watching videos. A rigorous analysis has been performed on different scalp locations and frequency rhythms of brain signals to formulate significant features corresponding to affective and interesting video content. The analysis presented in this paper shows that the extracted human attention-based features are generating promising results with the accuracy of 93.2% using SVM-based classification model, which supports the applicability of the model for various BCI-based applications for automatic classification of multimedia content.

KEYWORDS

Affect, Attention, Classification, EEG, Emotion, Tagging, Video

INTRODUCTION

Over the past years, one can see incredible growth in the direction of multimedia content creation along with the huge transformation in technology and devices. Today, the use and transmission of high-quality digital videos can be seen almost everywhere whether it is in online learning videos, surveillance videos, entertainment videos, or self-made videos on smartphones (Caviedes, 2012). With this explosive increase in the use and transmission of these images and videos, their efficient management techniques are also in high demand. With the advancement in technology, today new

DOI: 10.4018/IJIT.306968

*Corresponding Author

This article published as an Open Access Article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

opportunities are also created for accessing multimedia content archives, but for efficient search methods, proper annotation of these content with significant features is always required (Dimitrova et al., 2002; Smith & Chen, 2005).

Any kind of multimedia content like images, videos, or music usually involves some emotions that the creator wants to bring in its viewers and somehow generates different impacts on viewers' or listeners' emotional states (Isola et al., 2014). This kind of effect is highly subjective and is usually not incorporated while indexing and analyzing multimedia content. Thus, in this paper, the most important factor "Affect" is considered for the analysis of multimedia content. Affectiveness is directly related to the viewer's emotion and can work as a significant subjective feature for the classification and tagging of multimedia content (Siddharth et al., 2019). Affective analysis of multimedia content can be done by analyzing human's perception and cognition while watching that multimedia content (Isola et al., 2014; Siddharth et al., 2019).

In today's era, the advancement in neuroscience and brain-computer interface technologies have provided a deep understanding of complex information processing for automatic detection of a range of cognitive states (Müller, 2008; Hassanien & Azar, 2015). The brain imaging research field can play an effective role in the prediction of non-conscious user reactions to various types of multimedia content like movies, cricket videos, online video advertisements, etc. To date, many technologies have been developed and successfully used for capturing and analyzing brain reactions such as fMRI, PET, CT, EEG, etc. (Ghaemmaghami, 2017). For the past few decades, owing to the development of affordable and portable EEG devices, researchers are trying to explore their use in different possible fields (Hassanien & Azar, 2015). The use of EEG signals for multimedia content assessment is also an active research area.

Motivated by the applicability of EEG signals and the role of affectiveness in multimedia content assessment, in this paper, the work is presented towards proposing a model for automatic tagging of videos by modeling the human's EEG responses corresponding to specific video content. The work presented here is an extension of the previously published work done by Sharma et al. (2021), where the authors explored the relationship between affective content and corresponding EEG responses of viewers. Different brain sections like Frontal, Temporal, Parietal, and Occipital parts have been explored using extracted frequency bands of EEG signals i.e., Alpha, Beta, Gamma, Delta, and Theta. In continuation of this, here an automatic video tagging model is developed using the analysis presented in Sharma et al. (2021). The significant contributions of the presented research are mentioned as follows:

- First, to associate EEG signals and affective video content, EEG signals of 40 participants have been explored corresponding to 4-4 short video clips from two categories of videos.
- Second, as the short video clips are used for the creation of the proposed model, only frequency domain features have been targeted to characterize the EEG signals. Also, brain signals usually have significant information which lies in different frequency ranges, thus different frequency ranges are extracted to further generate meaningful features. These ranges are generally used to portray information in EEG signals.
- Further, to characterize the factors that drive the viewer's attention and affective state, the combined effect at different brain locations is used to model human attention. The analysis presented in Sharma et al. (2021) is used to formulate attention and affect-based features.
- In the last, the extracted features were used to model the automatic tagging of two categories of videos using SVM-based classifiers. The classification results are presented using different combinations of features along with their significance.

The presentation of the proposed research is structured in this paper as: the background of EEG signals, the significance of different brain regions and frequency bands followed by the applicability of EEG signals-based approaches in different sectors as well as for multimedia content assessment

is presented in next section i.e., Background and Related Work. Further, various steps followed for the development of the automatic video tagging model along with the dataset description are discussed in the Methodology section. Various results obtained during the experimental design and their explanation and significance are presented in the Results and Discussion Section. The major contributions of the proposed work are provided in the Key Contributions and Observations section followed by the Conclusion of the presented work.

BACKGROUND AND RELATED WORK

Video tagging and categorization is the process of automatic characterization of videos based on their content. The accuracy and effectiveness of various video content assessment-related applications is usually depending upon the effectiveness of their attributes, so feature extraction is the key step in any video analysis procedure (Caviedes, 2012). However, the proper mapping of extracted features from videos like color, texture, motion, or structure along with the semantic and viewer's subjective aspects is not an easy task (Dimitrova et al., 2002). Manual processing of multimedia content is considered to be a highly accurate method to incorporate subjective aspects, but at the same time, it is a very slow process and requires thorough attention (Smith & Chen, 2005). Usually, during the manual processing of videos, the user's attention and perception play a major role (Isola et al., 2014). It is somewhere directly related to the interestingness and affectiveness of the content present in the videos. Affectiveness and interestingness are directly related to the viewer's emotion and attention and can work as significant subjective features for the classification and tagging of multimedia content (Siddharth et al., 2019).

Over the decade, the multimedia community is continuously trying to simulate human cognitive abilities into machines, to make them work more effectively (Müller, 2008). Affective analysis of multimedia content can be done by analyzing human's perception and cognition while watching that multimedia content. The neuroscience research community has done a lot of advancements in understanding information processing in the brain for different cognitive activities. Recently, due to decreasing cost and easy availability of EEG (Electroencephalography) devices, their applicability in the development of BCI devices can be highly seen (Hassanien & Azar, 2015; Ghaemmaghami, 2017). The EEG signals captured through the brain contain valuable information about neural activities inside the brain mostly in the frequency range of 2Hz–64Hz, which can be used for analyzing the physiological state of the various regions of the brain (Siddharth et al., 2019; Ghaemmaghami, 2017). Further, various regions of the human brain respond differently with respect to the task they are performing. The involvement of different brain regions is described below (Müller, 2008; Ghaemmaghami, 2017; Sharma et al., 2021).

- Occipital Cortex (OC) - When a user sees any object through the eyes, this area is active in the processing of visually experienced data and thus used in experiments involving images and videos.
- Parietal Cortex (PC) - When the brain tries to understand the objects, situation from itself motor functions are active and this region is responsible for that.
- Temporal Cortex (TC) - Involvement of this region is more toward the auditory signals, i.e., language processing and speech production.
- Frontal Cortex (FC) - It is more involved in executive function, like maintaining physical control, thinking, planning, and behavior observation.

Whenever the brain is in a certain state it generates a specific pattern of an electrical signal, EEG recording gives variable frequency patterns change, and analysis of those patterns also gives insight into different cognitive processes, (Sharma et al., 2021; Gawali et al., 2012; Alarcao & Fonseca, 2019). The significance of specific frequency ranges is presented in Table 1.

Table 1. EEG band frequency range and commonly measured brain functions

Band	Frequency Range	Significance
Delta	0.5-3 Hz	Sleep quality measurement, Increased concentration when using internal working memory in tasks
Theta	3-8 Hz	Memory encoding and information retrieval from memory Cognitive workload indication, Fatigue level detection.
Alpha	8-12 Hz	Represents a relaxation state, sometimes monitoring attention level also
Beta	12-40 Hz	Brain activity involved in executive function “mirror neuron system” activation indication
Gamma	40-70 Hz	It reflects attentive focus. Can also record rapid eye movement

To date, a lot of efforts have been made toward the development of BCI-based devices using EEG signals. Researchers have explored the use of EEG signals in a vast range of fields such as neuro-rehabilitation (Duan et al., 2017; Kumar et al., 2013; Padfield et al., 2019) emotion recognition (Alarcao & Fonseca, 2019; Hiyoshi-Taniguchi et al., 2013; Li et al., 2017), Brain Robot Interaction (BRI) (Hassanien & Azar, 2015; Ghaemmaghami, 2017) and cognitive state analysis of humans, etc. The use of EEG signals can be seen in discovering the action-oriented effects on the human cognitive state (Kumar et al., 2013; Sharma et al., 2018). Further, EEG-based devices have been explored in the medical field also, for diagnosing and analysis of certain types of psychological disorders (Pham et al., 2020).

Over the past few years, due to the progress in low-cost and portable EEG device development, the multimedia research community is also putting efforts into the development of BCI-based mouse-free approaches for multimedia content analysis. The initial work in the field of multimedia using EEG was seen in Bigdely-Shamlo et al. (2008), Wang et al. (2009), and Huang et al. (2011), where authors analyzed the human attentive reaction through EEG responses using a rapid serial visual presentation (RSVP) task. RSVP is a method, where a sequence of images is presented to the viewer along with the target image to capture their attentive response or to observe their visual locations (Lees et al., 2018). Here EEG signals have been successfully used to classify the target image from the rest of the images. After this initial work, EEG signals using RSVP methods were explored to solve a variety of vision-related tasks such as object categorization, object segmentation (Mohedano et al., 2014), object detection (Mohedano et al., 2015), image searching (Huang et al., 2011), etc. Some of the work also explored the combination of different modalities such as EEG, eye tracking, and user ratings for examining artifacts in multimedia content (Tauscher et al., 2017). One of the recent works using EEG-based visual data classification is presented by (Cudlenco et al., 2020). Here authors provided a method of boosting the classification accuracy of objects under six different categories by combining EEG-based extracted features.

In the past few years, the work toward affective video content analysis can also be seen (Baveye et al., 2018). Affective computing field deals with the analysis of viewers’ emotions while watching affective content. To date, various efforts have been made in the field of human emotion recognition for the classification of a particular type of emotion in the valence-arousal emotion space using EEG signals (Suhaimi et al., 2020). Various models using machine learning and deep learning techniques have been developed for the classification of human emotional states using EEG signals (Hiyoshi-Taniguchi et al., 2013; Li et al., 2017). Various emotion-related publicly available datasets exist, where participants’ physiological responses have been recorded while watching affective video clips (Miranda-Correa et al., 2021; Kossaifi et al., 2017; (Abadi et al., 2015; Soleymani et al., 2012). These affective clips are created on the scale of Valence/Arousal space to evoke certain emotions among viewers and are largely based on subjective ratings. Although a lot of affective datasets are

present, very little effort has been reported on the analysis of the affectiveness of these video clips using EEG signals. Classification and tagging of visual content are very simple and effortless tasks for human beings, thus we believe that unconscious signals while watching that video content can be used effectively to provide more meaningful tags to them.

In literature, it has been observed that most of the methods for EEG signals classification involve the extraction of various types of features such as time-domain features, frequency domain features, and time-frequency domain features (Ghaemmaghami, 2017). The combinations of these features have also been used to test different classification models for better accuracy. Some authors have used different deep learning (DL) models, as well as a combination of DL models also for automatic EEG signals classification (Mishra et al., 2020). As discussed earlier, different brain regions and frequency ranges have a specific purpose in the analysis of the human cognitive state, thus proper modeling of EEG signals according to their significance is required for better model preparation. In literature also, it has been found that proper modeling of EEG signals can provide better accuracy in combination with a simple classification model in comparison to powerful deep learning classification models.

To explore the applicability of EEG signals for affective tagging of videos without conscious effort, various related work has been discussed, and accordingly, the work presented in this paper is prepared to focus on the following key points:

- The human brain is a complex architecture, where different brain regions are involved in the processing of information differently. Thus, modeling of specific brain regions effectively can provide a better understanding of the human cognitive state.
- In literature it has been found that EEG signals can effectively capture the brain-related signals corresponding to different activities. These EEG signals have important brain-related information in them which can be represented in different domains such as the time domain, frequency domain, and time-frequency domain. As in this paper, short video clips are used to record the viewer's EEG response, thus only frequency domain representation is used to extract the meaningful features from EEG signals.
- It is evident in Table 1, that in the frequency domain, specific frequency rhythms have a significant role in representing the specific cognitive state. To model, the human affective and attentive state power in these rhythms should be modeled properly. Thus, in this paper, various features according to the most affective brain regions and frequency rhythms have been extracted for further classification.
- Further, to demonstrate the applicability of the model for various BCI-based applications for automatic classification of multimedia content, the proposed extracted features are fed to aid SVM-based classification model.

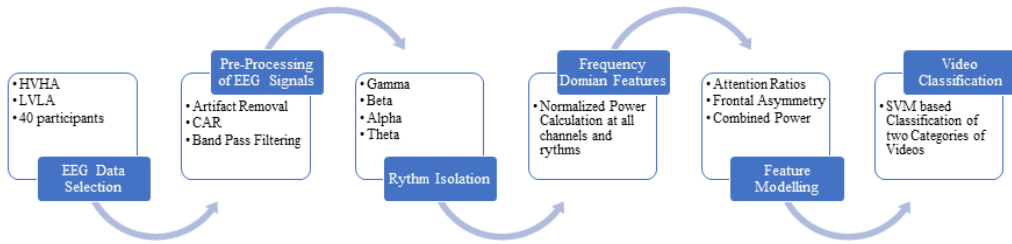
METHODOLOGY AND EXPERIMENTATION

In this paper participant's implicit response and attention level with respect to video stimuli are analyzed using corresponding EEG signals. The computational procedure followed for the development of the proposed model is presented in Figure 1. Various steps followed in the experimental procedure like pre-processing of EEG data, frequency range decomposition using DWT, feature extraction, human attention modeling, and classification results are briefly described.

Dataset Description: AMIGOS (Miranda-Correa et al., 2021)

"AMIGOS: A dataset for affect, personality and mood research on individuals and groups" is a dataset for assessing affect, personality, and mood of participants while watching videos. In AMIGOS forty participants (27male) were involved in the experiment for providing their physiological recordings using different modalities in form of EEG, EMG, and GSR. To stimulate the cognitive effects two different experiments were performed by 40 participants (27 Male) i.e., a short video experiment and

Figure 1. Computational procedure of proposed system



a long video experiment. The videos used in the experiment were selected based on their position in the valence/arousal dimension from two different datasets as presented in Figure 2 (Abadi et al., 2015; Soleymani et al., 2012). In a short video experiment, 16 short videos (4 videos under each quadrant) of < 250-sec duration, and in a long video experiment, 4 long-duration videos (~20 min) were presented to the participants. Participants’ affective responses were recorded in two cases, first when they were alone and second when they were watching videos in the group i.e., as part of the audience.

Short Video Experiment (Miranda-Correa et al., 2021): For the proposed experimental analysis, data corresponding to the short video experiment is selected from the AMIGOS dataset where 16 videos were presented to participants as visual stimuli. The selection of videos was done based on their position in the valence/arousal dimension, such that each quadrant will consist of four videos. In this paper, for the development and testing of the model, two quadrants of valence/arousal were considered i.e., high arousal high valence ($H_V H_A$), and low arousal low valence ($L_V L_A$), and EEG data corresponding to eight video clips are used for further experimentation. Description of these eight videos under HVHA and LVLA categories along with their video ids given in Table 2.

EEG Recordings Structure: During the short video experiment, the different types of video stimuli presented to the participants are termed trials. Before each trial participants’ self-assessment was

Figure 2. Valence / arousal affective space

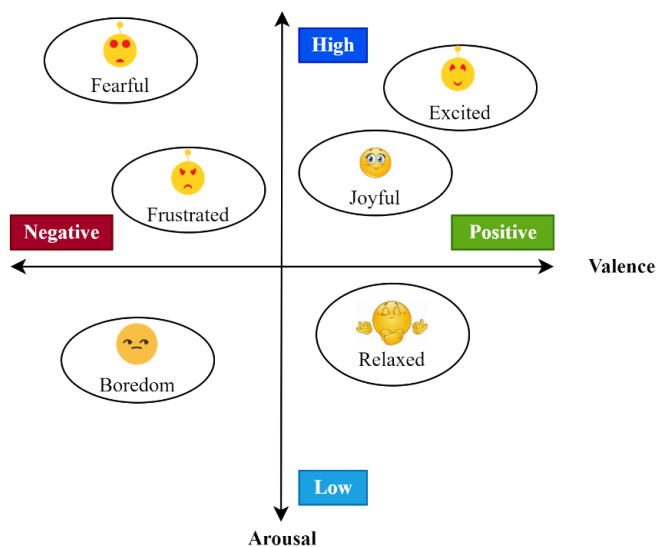


Table 2. Description of videos used in experiment (Miranda-Correa et al., 2021)

Video Category	Video No.	Video Description
H _V H _A	12,13,15,16	Airplane, When Harry met Sally, Hot Shots, Love
L _V L _A	3,5,6,7	Exorcist, My Girl, My Body Guard, The Thin Red Line Fatigue level detection

also recorded to assess their mood, familiarity, level of valence and arousal, etc. Then EEG signals were recorded for five second baseline period followed by the trial (visual stimuli) period. After each trial, self-assessment is again recorded for assessing the participants' explicit response corresponding to those visual stimuli as shown in Figure 3. EEG signals were recorded using 14 channel Emotic EPOC neuroheadset, in which a 10-20 electrode placement system was adopted as shown in Figure 4.

EEG Data Preprocessing and Rhythm Isolation

Due to low voltage variations, EEG signals usually have a very low signal-to-noise ratio. Further, signals are also effected by different artifacts like signal line noise, muscle movement-related noise, noise due to eye blinks, etc. To remove the noise from EEG signals, first, Independent Component

Figure 3. Stimuli trial structure

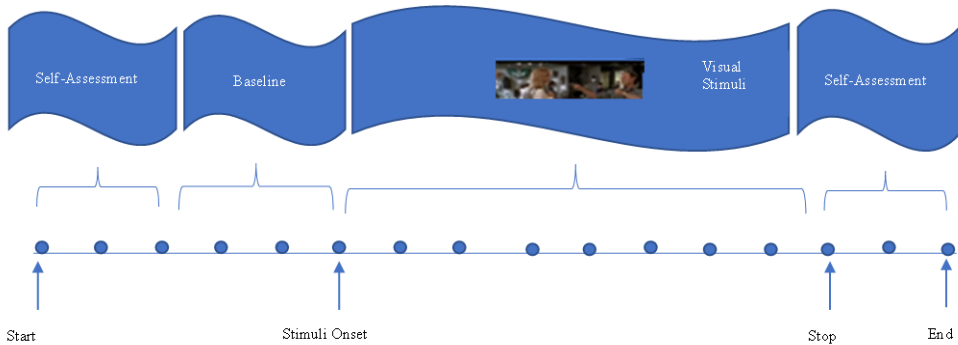
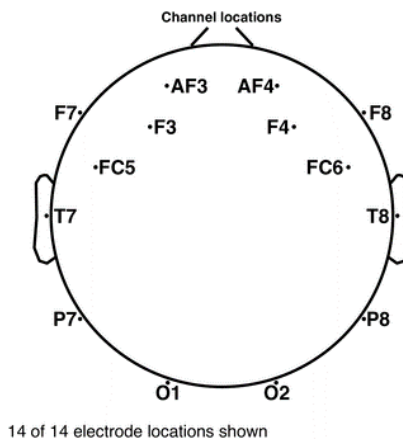


Figure 4. Electrode scalp positions



Analysis (ICA) based blind source separation technique is used to remove eye artifacts from EEG signals at all channel locations. Further common average re-referencing (CAR) of EEG data is done and in the last EEG signals were band-pass filtered between 4-65 Hz frequency range (Miranda-Correa et al., 2021).

In the proposed approach, EEG signals corresponding to short video clips are used, to establish the relationship between video affective content and human attention. Thus, frequency domain features are extracted for further analysis. In literature, it is already discussed that different brain portions play different roles in processing of visual information. Further, human behavior is usually analyzed using five frequency ranges i.e., Delta, Theta, Alpha, Beta, and Gamma. These frequency ranges are also known as Rhythms.

To isolate different frequency rhythms for EEG data, Discrete Wavelet Transform (DWT) based signal decomposition is performed. DWT executes decomposition of time series signals using high pass and low pass filtering with down sampling ratio of two (Akansu & Haddad, 2000; Kehtarnavaz, 2008). At each level i DWT output two types of coefficients: i^{th} level approximation coefficients A_i and detailed coefficients D_i . The decomposition of A_i components is done as depicted in Figure 5 to excerpt local information from each sub band. The decomposition is performed using Daubechies-four (db4) mother wavelet and with da decomposition level of four, as the EEG data used in the experiment is sampled at 128Hz (Kehtarnavaz, 2008). The details of the extracted rhythms using DWT are presented in Table 3.

Feature Extraction and Human Attention Modelling

The most popular method in EEG signals analysis is to analyze the power spectrum of signal to see the responses in different frequency ranges at a particular time. After wavelet-based decomposition of the rhythms, the power in the baseline period and video period is calculated. Further, to analyze the affective response of participants with respect to video, the difference between the power in video period w.r.t the baseline period is calculated, which is further normalized using Eq 1, where symbol

Figure 5. DWT based signal decomposition

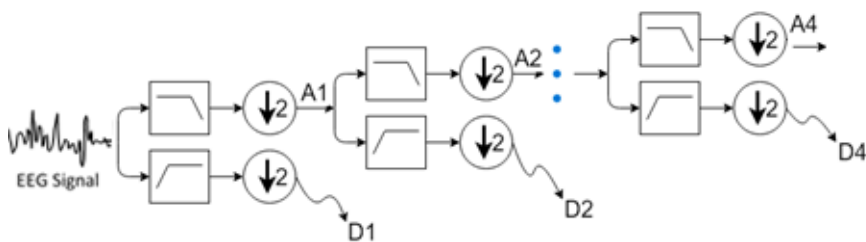


Table 3. Rhythm isolation using DWT decomposition

DWT Coefficients	Frequency Range	Extracted Frequency Bands
D1	32-64 Hz	Gamma
D2	16-32 Hz	Beta
D3	8-16 Hz	Alpha
D4	4-8 Hz	Theta

\dot{P} represents normalized power. These normalized power differences are extracted for all extracted frequency rhythms at 14 channel locations as shown in Table 4 and Figure 4.

$$\dot{P}(c,r) = \frac{P(c,r)^{video} - P(c,r)^{baseline}}{P(c,r)^{baseline}} \quad (1)$$

Where $r \in [1,4]$, $c \in [1,14]$ and k represents different rhythms, channels, and number of wavelet coefficient in the r^{th} rhythm respectively. These extracted power values in different frequency rhythms and channel locations are further used to model the more significant feature for affective tagging of videos.

Affective tagging of videos somewhere depends on the emotional as well as interesting content in a video, thus the features corresponding to these two aspects need to be modeled. Thus, a rigorous analysis has been performed to establish the relationship between affective response of participants and brain regions to categorize the two categories of videos in Sharma et al. (2021). According to the presented relationship, activation response, and significance of brain regions, responses under the following brain regions are merged to extract meaningful features. The merging is performed after the analysis of their active response and according to their position on the scalp.

Frontal Region (F): $F7, FC5, FC6, F8$

Visual cortex Region (V) $O1, O2, P7, P8$

Further to measure the interestingness of the video content, viewer attentiveness needs to be modeled. Information present in different frequency ranges represent a certain kind of cognitive state (Table 1). Previous analysis also shows that different frequency rhythms have shown a significant response under certain brain regions in. Alpha range usually represents the idle situation of the brain, whereas Beta waves represents the attentiveness, thus to formulate the interestingness of video content, the ratio of normalized power in the Alpha and Beta band is calculated at the Frontal and Visual Processing Region using Eq 2 resulting into two features. Here A_R denotes attentiveness ratio, and F , and V represents combined frontal and visual processing region.

$$A_{R1} = \frac{\dot{P}_{F,V}^{\alpha}}{\dot{P}_{F,V}^{\beta}} \quad (2)$$

Further, another attentive feature is modelled by taking the ratio of Theta and Beta Power values using Eq 3, as theta waves represent the working memory and emotional process, the ratio is calculated at the both Frontal and Visual Processing Region only.

$$A_{R2} = \frac{\dot{P}_{F,V}^{\theta}}{\dot{P}_{F,V}^{\beta}} \quad (3)$$

To model the affectiveness of the video content, the well-known asymmetric features were extracted by taking spectral power difference between symmetrical pair electrodes in all frequency ranges using Eq 4.

$$AS_{(r,m)} = \dot{P}_{(r,L)} - \dot{P}_{(r,R)} \quad (4)$$

Where $r \in [1, 4], m \in [1, 7]$ represents rhythms and mentioned electrode pairs in Table 4. L and R used to represent left and right electrodes numbering. Further, analysis on different brain regions is done to find the most effective Asymmetry based features. As the frontal region is usually involved in emotional process, thus Frontal Asymmetry Score is calculated by combining the responses at frontal Region (F): F7, FC5, FC6, F8

Additional important feature i.e., Relative Wavelet Energy of a particular rhythm w.r.t to the total energy of the signal is also calculated using Eq 5. It provides information about the similarity between two signals.

$$RWE_{F,V}^i = \frac{E_{F,V}^i}{E_{F,V}^{total}} \quad (5)$$

relative wavelet energy to assist us to choose an effective wavelet in our technique. RWE gives information about relative energy with associated frequency bands and can detect the degree of similarity between segments of a signal where, $i \in [1, 4]$ for four frequency rhythms, F, V represents Frontal and Visual Cortex region respectively. The Energy in specific frequency rhythm and the total energy of the signal is calculated using Eq 6 and Eq 7 respectively. relative wavelet energy to assist us to choose an effective wavelet in our technique. RWE gives information about relative energy with associated frequency bands and can detect the degree of similarity between segments of a signal

$$E_i = \sum_k |D_{i(k)}|^2 \quad (6)$$

$$E_{total} = \sum_i \sum_k |D_{i(k)}|^2 \quad (7)$$

Table 4. Symmetric pairs of electrodes

Asymmetric Pairs (AS)	Left Electrodes (L)	Right Electrodes (R)
	AF3	AF4
	F3	F4
	F7	F8
	FC5	FC6
	P7	P8
	O1	O2

RESULTS AND DISCUSSION

The proposed prototype is based on the modelling of human attention and affect using EEG signals for affective tagging of videos. Various results during the implementation of the proposed model along with their explanation are presented in this section.

As discussed in the previous section, the model is developed using the EEG data recorded corresponding to the videos from two quadrants of the valence/arousal dimension i.e., $H_A H_V$ and $L_A L_V$.

EEG data of 40 participants/subjects for 8 video clips (4-4 video clips from $H_A H_V$ and $L_A L_V$, Table 2) at 14 scalp locations is taken for the experimentation. The time-domain view of the EEG signal of one participant at one of the channels at the Frontal Region i.e., F7 is presented in Figure 6.

As only short video clips are used to measure the attentiveness and affectiveness, the time domain analysis of EEG signals is not useful for finding the required features. A frequency-domain representation of EEG responses of a participant for $H_A H_V$ and $L_A L_V$ classes of videos at all channels is presented in Figure 7, which shows a distinguishable conduct of the signal.

Further, to extract the frequency domain features, different frequency rhythms i.e., Theta, Alpha, Beta, and Gamma have been extracted by Wavelet-based disintegration of EEG signals. The frequency representation of different rhythms corresponding to the EEG signal response at channel location F7 shown in Figure 6 is presented in Figure 8.

After extraction of frequency rhythms, various frequency domain features were extracted, and considering the significance of different rhythms and brain regions, modeling of the extracted features is done to model the human attention and affect. As per the description provided in the methodology section, the Band power, Asymmetry, and Attention Ratio based features were extracted at the Frontal and Visual Processing Areas. Various extracted features are listed in Table 5.

The extracted twenty-one features were used to train the model for affective labeling of videos. Since the dataset contains only 8 trials/video clips for $H_A H_V$ and $L_A L_V$ categories of videos, the EEG data was windowed at every ten seconds of the data and further advance it by one second. In this way, the dataset was augmented to get more than 200 trials for one subject. EEG features were extracted from each trial. The model is trained and tested in two ways: Single Subject Video Classification Model and Multi-Subject Video Classification Model. In a single-subject classification model, the EEG data of one subject is used to train and test the model, whereas in the second one the trials of

Figure 6. EEG response of one participant at channel F7 for $H_A H_V$ and $L_A L_V$ Video

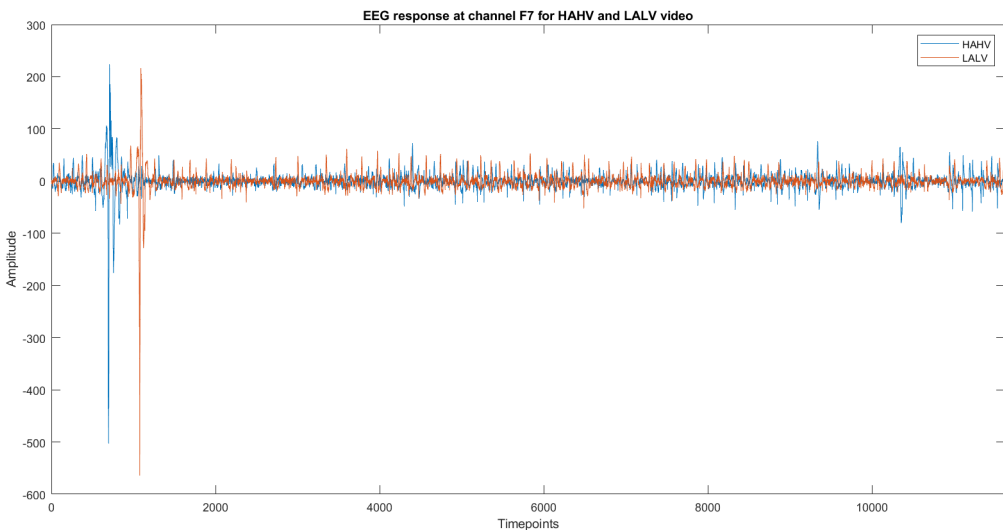


Figure 7. Frequency Domain representation of EEG responses for (a) HA HV, and (b) LALV Video

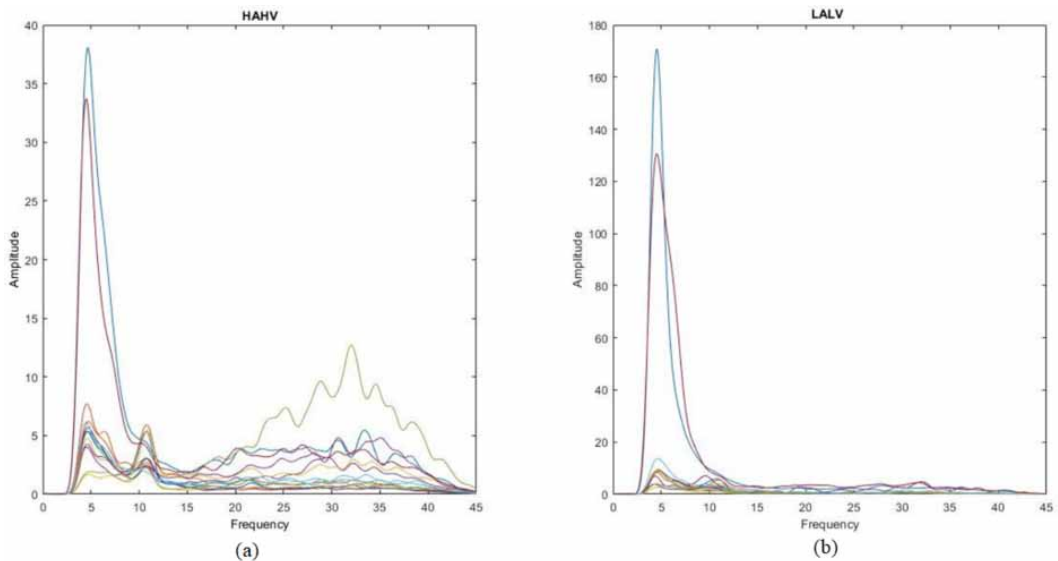
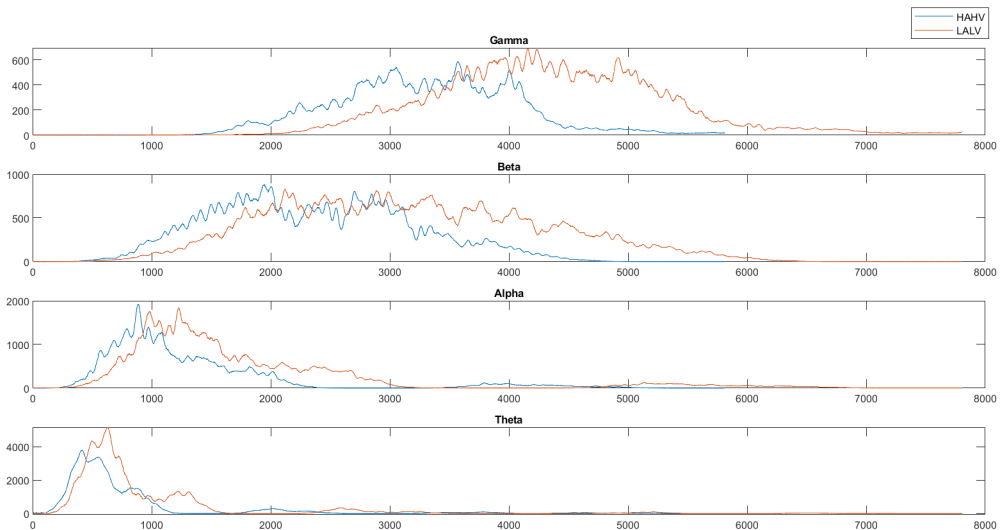


Figure 8. Rhythms extraction from EEG signal of one participant at channel F7 for HA HV and LALV Video



forty participants are used to train and test the model, thus EEG features were extracted from around 8000 trials.

Further, the Support Vector Machine (SVM) based classification methodology is used to test the performance of the proposed affect and attention-based features. SVM is a popular method to classify the data using an N-dimensional hyperplane. The results of the SVM classification method on proposed extracted features are presented in Table 6.

The result presented here shows that the extracted human attention-based features are generating promising results with an average accuracy of 93.2% using the SVM-based classification model which

Table 5. Summary of extracted features

Extracted Feature	Explanation	No. of Features
Combined Normalized Power (\dot{P})	Normalized power of all frequency rhythms at the combined frontal (F) and Visual Cortex (V) region F: F7, FC5, FC6, F8 V: O1, O2, P7, P8	Eight (4 Rhythms * 2 Regions)
Attention Ratio A_{R1}	The Ratio of Combined normalized power in Alpha and Beta Rhythms at Frontal (F) and Visual Cortex (V) region	Two
Attention Ratio A_{R2}	The Ratio of combined normalized power in Theta and Beta Rhythms at Frontal and Visual Cortex region	Two
Asymmetry Score AS	The asymmetry Score is calculated at Frontal Region	One
Relative Wavelet Energy	Relative Wavelet Energy of rhythms at combined frontal (F) and Visual Cortex (V) Region	Eight (4 Rhythms * 2 Regions)

Table 6. Performance of proposed work

Model Type	Accuracy (%)	
	$H_A H_V$	$L_A L_V$
Single Subject Classification Model	92.7	91.5
Multi Subject Classification Model	96.3	92.4

supports the applicability of the model for various BCI-based applications for automatic classification of multimedia content, and consumer-oriented content assessment, etc.

KEY OBSERVATIONS AND FINDINGS

In this paper, the work is presented to model the affective tagging of videos using EEG signals. The proposed work is aimed to excerpt the meaningful features from the EEG signals, to aid an affordable BCI-based approach for mouse-free tagging of videos. The key findings of the proposed work are summarized here.

- Human brain involves the complex processing of information, thus most effective brain regions are identified and modeled using the analysis presented by Sharma et al. (2021) for generating meaningful features to tag the videos.
- The different frequency rhythms of EEG signals contain certain information about the cognitive state of the human, and their responses to $H_A H_V$ and $L_A L_V$ categories of videos were also discussed and explained in previously published work in Sharma et al. (2021). Here, the significance of these rhythms was successfully established with the help of extracted features.
- The affective content of the video is mapped with the human affective state by modeling the frequency domain features i.e., Combined Normalized Power, Combined Relative Wavelet Energy, and Frontal Asymmetry Score in the most effected brain regions and frequency rhythms.
- Human attention is somewhere directly or indirectly related to the interestingness of the video content. Attention Ratio based features are proposed to model the attentiveness of the participant by considering the significance of the rhythms.

- The extracted features were successfully used to generate a classification model with reasonable average accuracy of 93.2% for affective tagging of videos.

In literature, EEG signals have been explored in a combination of applications such as emotion recognition, psychological state analysis, disorder or motor rehabilitation-related activities, etc. Researchers are now discovering the use of EEG signals for audiovisual content analysis. The most related research to our effort is done by Mutasim et al. (2017), where researchers modeled the classification of three types of videos using EEG responses by testing different classifiers on extracted features at five-channel locations. Highest accuracy is achieved at AF8 channel position, but the proper justification of selecting a particular channel and feature is not justified. The proposed work is motivated by the work presented by authors in Sharma et al. (2021). Here authors performed a rigorous analysis of all extracted frequency bands and brain regions to find the most operative brain regions and frequency bands to distinguish the two classes of videos. Another supported work is done by Bezugam (2021) and Singhal et al. (2018), where authors present an application of EEG signals to excerpt the important highlights in the video by human attention modelling. In this paper, the interesting use of EEG is presented which can be used to provide the prospect of implicit video tagging without the user’s conscious effort. To show the significance of the proposed work, its comparative analysis with some of the most related work is presented in Table 7.

Table 7. Relative analysis of the projected work

Paper	Task	Data Used	Methodology	Performance and Analysis
(Soleymani & Pantic, 2013)	Two Tasks: <ul style="list-style-type: none"> • Emotional Tagging of videos. • Matched and Unmatched tags classification using EEG signals 	<ul style="list-style-type: none"> • MANHOB-HCI Database • 24 Participants • 20 short video clips • 32 Channel Locations 	Power spectral features are extracted from frequency rhythms alpha, slow alpha, theta, beta, and gamma considering all 32 brain locations and a subset of locations.	F1 score in the range of .55 to .64 and .83 to .89 was claimed by the authors for emotional tagging of videos considering features from all brain locations and a subset of locations respectively for videos selected from three-dimension on Valence and Arousal. It is noticed that reduced no. of electrodes has increased the F1 value of SVM-based classification, whereas the reason and significance of combining the specific electrodes are not mentioned.
(Jang et al., 2018)	Video Identification using EEG signals	<ul style="list-style-type: none"> • DEAP database • 32 Participants • 40 video clips • 32 Channel Locations 	A Graph Convolution Neural Network (GCNN) approach is used to process EEG signals. Power and Entropy features are extracted at eight frequency rhythms	Accuracy in the range of 48.25% – 62.94% is reported by authors using different network parameters. Reported accuracy is very low to be used as a model for real-time application.
(Mutasim et al., 2017)	Video Classification of 3 types of videos using EEG signals	<ul style="list-style-type: none"> • Self-Made Data • 13 Participants 	Testing of different classification methods is done using extracted features through DWT, STFT, and FFT, etc., methods at all channel locations.	An accuracy of 80% is achieved at channel location AF8, although the selection of a particular, channel and feature is not justified.
(Bezugam, 2021)	Video Summarization EEG and Eye-tracking Signals	<ul style="list-style-type: none"> • Self-collected data • 15 participants • 64 Channel Locations 	Power Spectral Density (PSD) based features are extracted from High-frequency components using the Empirical Mode Decomposition method.	The work reported F1 measure of 75.95 using the combination of high-frequency components and different brain regions such as Frontal and Occipital lobes to extract the attention-based keyframes from videos.
Proposed Work	Video Tagging of two categories of videos using affective and interesting content	<ul style="list-style-type: none"> • AMIGOS Database • 40 participants • 14 channel locations • 8 short video clips 	A DWT-based frequency rhythms extraction followed by modeling of attention and affect-based feature extraction after analysis of most effective brain regions and frequency rhythms.	Average accuracy of 93.2% is achieved while testing the model. The proposed work presented a combination of sophisticated features to model the affective and interesting content of the video using EEG signals.

CONCLUSION

Nowadays every event in the world appears to be captured in the form of images or videos. These multimedia contents are usually designed in a way such that they can hit the user's cognizance affectively and this can be achieved through its affective content. Thus, the modelling of human emotion and attentiveness can provide significant information about the affectiveness of the video content. In this paper, an attempt is made to find the association between a human's cognitive state measured using neurophysiological signals (EEG) and video content. An extensive simulation has been performed to find the sophisticated features for the creation of an effective model. According to the significance of the effective brain regions and rhythms, human attention and affect-based features are presented here, to measure the affectiveness and interestingness of the participant corresponding to a particular video. The significance behind all the extracted features is thoroughly discussed in the Methodology section, which was supported by the literature presented in the Background Section. Step by step experimental outcomes are presented in the Results and Discussion Section followed by major conclusions and a comparative examination of the projected work. The results show the usefulness of the extracted features and explore new applicabilities of EEG signals to develop a variety of Brain-Computer Interface-based devices for automatic video content assessment in the near future.

ACKNOWLEDGMENT

The authors would like to thank the developers of the publicly available dataset "AMIGOS" (Miranda-Correa et al., 2021), for supporting this research by providing the access to this dataset.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Transactions on Affective Computing*, 6(3), 209–222. doi:10.1109/TAFFC.2015.2392932
- Akansu, A. N., & Haddad, R. A. (2000). Chapter 6 - Wavelet Transform. In *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets (Series in Telecommunications)* (2nd ed., pp. 391–442). Academic Press. doi:10.1016/B978-012047141-6/50006-9
- Alarcao, S. M., & Fonseca, M. J. (2019). Emotions Recognition Using EEG Signals: A Survey. *IEEE Transactions on Affective Computing*, 10(3), 374–393. doi:10.1109/TAFFC.2017.2714671
- Baveye, Y., Chamaret, C., Dellandrea, E., & Chen, L. (2018). Affective Video Content Analysis: A Multidisciplinary Insight. *IEEE Transactions on Affective Computing*, 9(4), 396–409. doi:10.1109/TAFFC.2017.2661284
- BezugamS. S. (2021, January 27). *Efficient Video Summarization Framework using EEG and Eye-tracking Signals*. <https://arxiv.org/abs/2101.11249>
- Bigdely-Shamlo, N., Vankov, A., Ramirez, R. R., & Makeig, S. (2008). Brain Activity-Based Image Classification From Rapid Serial Visual Presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5), 432–441. doi:10.1109/TNSRE.2008.2003381 PMID:18990647
- Caviedes, J. E. (2012). The Evolution of Video Processing Technology and Its Main Drivers. In *Proceedings of the IEEE (vol. 100, Issue 4, pp. 872-877)*. Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/JPROC.2011.2182072
- Cudlenco, N., Popescu, N., & Leordeanu, M. (2020). Reading into the mind's eye: Boosting automatic visual recognition with EEG signals. *Neurocomputing*, 386, 281–292. doi:10.1016/j.neucom.2019.12.076
- Dimitrova, N., Zhang, H.-J., Shahrray, B., Sezan, I., Huang, T., & Zakhor, A. (2002). Applications of video-content analysis and retrieval. *IEEE MultiMedia*, 9(3), 42–55. doi:10.1109/MMUL.2002.1022858
- Duan, L., Bao, M., Cui, S., Qiao, Y., & Miao, J. (2017). Motor Imagery EEG Classification Based on Kernel Hierarchical Extreme Learning Machine. *Cognitive Computation*, 9(6), 758–765. doi:10.1007/s12559-017-9494-0
- Gawali, B., Rokade, P., Mehrotra, S., Rao, S., & Abhang, P. (2012). Classification of EEG signals for different emotional states. In *Fourth International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom2012)* (pp. 177-181). Institution of Engineering and Technology. doi:10.1049/cp.2012.2521
- Ghaemmaghani, P. (2017, March). *Information Retrieval from Neurophysiological Signals*. University of Trento. <http://eprints-phd.biblio.unitn.it/1848/1/PhD-Thesis.pdf>
- Hassanien, A. E., & Azar, A. T. (Eds.). (2015). *Brain-Computer Interfaces: Current Trends and Applications* (1st ed., Vol. 74). Springer Cham. doi:10.1007/978-3-319-10978-7
- Hiyoshi-Taniguchi, K., Kawasaki, M., Yokota, T., Bakardjian, H., Fukuyama, H., Cichocki, A., & Vialatte, F. B. (2013). EEG Correlates of Voice and Face Emotional Judgments in the Human Brain. *Cognitive Computation*, 7(1), 11–19. doi:10.1007/s12559-013-9225-0
- Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., & Hild, K. E. II. (2011). A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing*, 74(12–13), 2041–2051. doi:10.1016/j.neucom.2010.12.025
- Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What Makes a Photograph Memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1469–1482. doi:10.1109/TPAMI.2013.200 PMID:26353315
- Jang, S., Moon, S. E., & Lee, J. S. (2018). Eeg-Based Video Identification Using Graph Signal Modeling and Graph Convolutional Neural Network. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. (pp. 3066-3070). Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/ICASSP.2018.8462207

- Kehtarnavaz, N. (2008). Frequency Domain Processing. In *Digital Signal Processing System Design: LabVIEW-Based Hybrid Programming (Digital Signal Processing SET)* (2nd ed., pp. 175–196). Academic Press. doi:10.1016/B978-0-12-374490-6.00007-6
- Kossaifi, J., Tzimiropoulos, G., Todorovic, S., & Pantic, M. (2017). AFEW-VA database for valence and arousal estimation in-the-wild. *Image and Vision Computing*, 65, 23–36. doi:10.1016/j.imavis.2017.02.001
- Kumar, S., Riddoch, M. J., & Humphreys, G. (2013). Mu rhythm desynchronization reveals motoric influences of hand action on object recognition. *Frontiers in Human Neuroscience*, 7. Advance online publication. doi:10.3389/fnhum.2013.00066 PMID:23471236
- Lees, S., Dayan, N., Cecotti, H., McCullagh, P., Maguire, L., Lotte, F., & Coyle, D. (2018). A review of rapid serial visual presentation-based brain–computer interfaces. *Journal of Neural Engineering*, 15(2), 021001. doi:10.1088/1741-2552/aa9817 PMID:29099388
- Li, J., Zhang, Z., & He, H. (2017). Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition. *Cognitive Computation*, 10(2), 368–380. doi:10.1007/s12559-017-9533-x
- Miranda-Correa, J. A., Abadi, M. K., Sebe, N., & Patras, I. (2021). AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. *IEEE Transactions on Affective Computing*, 12(2), 479–493. doi:10.1109/TAFFC.2018.2884461
- Mishra, A., Ranjan, P., & Ujlayan, A. (2020). Empirical analysis of deep learning networks for affective video tagging. *Multimedia Tools and Applications*, 79(25–26), 18611–18626. doi:10.1007/s11042-020-08714-y
- Mohedano, E., Healy, G., McGuinness, K., Giró-i-Nieto, X., O'Connor, N. E., & Smeaton, A. F. (2014). Object Segmentation in Images using EEG Signals. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 417–426). ACM. doi:10.1145/2647868.2654896
- Mohedano, E., McGuinness, K., Healy, G., O'Connor, N. E., Smeaton, A. F., Salvador, A., Porta, S., & Giró-i-Nieto, X. (2015). Exploring EEG for Object Detection and Retrieval. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval (ICMR '15)* (pp. 591–594). ACM. doi:10.1145/2671188.2749368
- Müller, V. C. (2008). Margaret A. Boden, *Mind as Machine: A History of Cognitive Science*, 2 vols. *Minds and Machines*, 18(1), 121–125. doi:10.1007/s11023-008-9091-9
- Mutasim, A. K., Tipu, R. S., Raihanul Bashar, M., & Ashraf Amin, M. (2017). Video Category Classification Using Wireless EEG. In *International Conference on Brain Informatics, BI 2017. Lecture Notes in Computer Science* (vol. 10654, pp. 39–48). Springer. doi:10.1007/978-3-319-70772-3_4
- Padfield, N., Zabalza, J., Zhao, H., Masero, V., & Ren, J. (2019). EEG-Based Brain-Computer Interfaces Using Motor-Imagery: Techniques and Challenges. *Sensors (Basel)*, 19(6), 1423. doi:10.3390/s19061423 PMID:30909489
- Pham, T. H., Vicesh, J., Wei, J. K. E., Oh, S. L., Arunkumar, N., Abdulhay, E. W., Ciaccio, E. J., & Acharya, U. R. (2020). Autism Spectrum Disorder Diagnostic System Using HOS Bispectrum with EEG Signals. *International Journal of Environmental Research and Public Health*, 17(3), 971. doi:10.3390/ijerph17030971 PMID:32033231
- Sharma, S., Dubey, A. K., Ranjan, P., & Rocha, A. (2021). Neural correlates of affective content: Application to perceptual tagging of video. *Neural Computing & Applications*. Advance online publication. doi:10.1007/s00521-021-06591-6
- Sharma, S., Mishra, A., Kumar, S., Ranjan, P., & Ujlayan, A. (2018). Analysis of Action Oriented Effects on Perceptual Process of Object Recognition Using Physiological Responses. In U. Tiwary (Ed.), *Lecture Notes in Computer Science: Vol. 11278. Intelligent Human Computer Interaction. IHCI 2018* (pp. 46–58). Springer. doi:10.1007/978-3-030-04021-5_5
- Siddharth, S., Jung, T. P., & Sejnowski, T. J. (2019). Impact of Affective Multimedia Content on the Electroencephalogram and Facial Expressions. *Scientific Reports*, 9(1), 16295. Advance online publication. doi:10.1038/s41598-019-52891-2 PMID:31705031
- Singhal, A., Kumar, P., Saini, R., Roy, P. P., Dogra, D. P., & Kim, B. G. (2018). Summarization of videos by analyzing affective state of the user through crowdsourcing. *Cognitive Systems Research*, 52, 917–930. doi:10.1016/j.cogsys.2018.09.019

Smith, M. A., & Chen, T. (2005). Image and Video Indexing and Retrieval. In A. L. Bovik (Ed.), *Handbook of Image and Video Processing* (2nd ed.). Academic Press. doi:10.1016/B978-012119792-6/50121-2

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Transactions on Affective Computing*, 3(1), 42–55. doi:10.1109/T-AFFC.2011.25

Soleymani, M., & Pantic, M. (2013). Multimedia implicit tagging using EEG signals. In *IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). Institute of Electrical and Electronics Engineers (IEEE). doi:10.1109/ICME.2013.6607623

Suhaimi, N. S., Mountstephens, J., & Teo, J. (2020). EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities. *Computational Intelligence and Neuroscience*, 2020, 1–19. doi:10.1155/2020/8875426 PMID:33014031

Tauscher, J. P., Mustafa, M., & Magnor, M. (2017). Comparative Analysis of Three Different Modalities for Perception of Artifacts in Videos. *ACM Transactions on Applied Perception*, 14(4), 1–12. doi:10.1145/3129289

Wang, J., Pohlmeier, E., Hanna, B., Jiang, Y. G., Sajda, P., & Chang, S. F. (2009). Brain state decoding for rapid image retrieval. In *Proceedings of the 17th ACM International Conference on Multimedia* (pp. 945-954). ACM Press. doi:10.1145/1631272.1631463

Shanu Sharma is currently working as an Assistant Professor in Department of CSE at ABES Engineering College, Ghaziabad (Affiliated to A.K.T University, Lucknow). She is having 11 years of teaching and research experience. Her research area includes Cognitive computing, Computer Vision, Pattern Recognition and Machine Learning. She has published and presented her work in various National and International Conferences and Journals and currently associated with various reputed International Conferences and journals as Reviewer. She is a senior member of IEEE and also an active member of other professional societies like ACM, Soft Computing Research Society and IAENG.

Ashwani K. Dubey received the M.Tech. degree in Instrumentation and Control Engineering from Maharshi Dayanand University, Rohtak, Haryana, India, in 2007, and Ph. D degree from the Department Electrical Engineering, Faculty of Engineering and Technology, Jamia Millia Islamia (A Central Govt. University), New Delhi, India, in 2014. Currently, he is a Professor in the Department of Electronics and Communication Engineering, Amity School of Engineering, Amity University, Noida, Uttar Pradesh, India. He has chaired many IEEE international conferences and workshops. His research interest includes wireless communications, energy optimization of sensor networking, image processing, computer vision, AI, Machine Learning and Deep learning.

Priya Ranjan graduated from IIT Kharagpur (EE, 1997), West Bengal, India and earned his advanced degrees of MS (EE) and Ph.D. (ECE) at the University of Maryland, College Park, USA in 1999 and 2003 respectively. His latest interest is in biological computation, oncological multimedia processing, MRI, EEG and ECG processing in a lab titled Health Analytics and Data Visualization Environment (HAVE) he established jointly with Prof. Rajiv Janardhanan in 2018. Some of their recent papers have been accepted in Multimedia and Tools (MMT) by Springer. Neural Computing and Applications (NCAA) and in conferences on cancer related issues all around the world included highly rated conference HCI and ISBI 2018. In particular his emphasis is on developing software tools for affordable early detection of different reproductive cancers, heart malignancies, mental anomalies using EEG and EcoG signals. He has been helping biologists and pathologists by designing and implementing user interfaces friendly to their way of working and generating innovative results for next generation of analysis for bio-professionals. He truly believes that a multi-way productive collaboration with biologists, public health professionals and clinical scientists with bio-informatics and bio-computational suite of tools will lead to next knowledge revolution for humanity towards better health facilitation. He has also written a search tool for exploring Ayurved related documents under a DST training program which was well appreciated by Vaidyas for Gurukul Kangri. He has received many large project awards from some of the most reputed agencies like National Science Foundation (NSF), DARPA (Credited with development of the Internet), ICMR, Delhi University, DBT etc. He has authored more than fifty research articles in different international conferences around the world, many heavily cited journal papers and many book chapters in heavily cited books published by IEEE and Springer Press. He believes that information, communication, and visualization technology (ICVT) has a huge role to play in making traditional societies more and more transparent in their decision making and providing new capabilities hitherto denied to women and children.