Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 171 (2020) 857–867

Procedia Computer Science

www.elsevier.com/locate/procedia

Third International Conference on Computing and Network Communications (CoCoNet'19)

Emotion Recognition in Valence-Arousal Space from Multi-channel EEG data and Wavelet based Deep Learning Framework

Divya Garg[a]*, Gyanendra K. Verma[b]

[a,b]*Department of Computer Engineering, National Institute of Technology Kurukshetra, Kurukshetra-136119, India*

Abstract

The conventional emotion recognition methods are mostly based on the frequency characteristics of electroencephalograph (EEG) signals. However, spatial features are likewise valuable as it contains latent information related to emotional states. In this paper, a wavelet-based Deep Learning framework proposed by considering both frequency and spatial characteristics of multi-channel EEG signal for emotion recognition. The Continuous Wavelet Transform is utilized to produce Scalogram, a function of frequency and time to getting better time localization for short-duration, high-frequency events, and better frequency localization for low-frequency, longer-duration events. Then, the GoogleNet model is presented to recognize emotion states from Scalogram. The experiments performed with benchmark DEAP database having a three-dimensional valence, arousal, and dominance data along with multi-channel EEG data. The experimental results demonstrate that the characteristics contained in the Scalogram were complementary, and GoogleNet is more suitable for emotion recognition in two/ three-dimension space.

Peer-review under responsibility of the scientific committee of the Third International Conference on Computing and Network Communications (CoCoNet'19).

Keywords: EEG, Scalograms, Wavelet Transform, CNN, Affective Computing, DEAP database, GoogleNet

* Divya Garg Tel.: +91-9034239244.
E-mail address: divya_6180091@nitkkr.ac.in

1877-0509
Peer-review under responsibility of the scientific committee of the Third International Conference on Computing and Network Communications (CoCoNet'19).
10.1016/j.procs.2020.04.093

1. Introduction

Emotions act as a vital part of decision making and human interpersonal communication [1]. Moreover, various states of emotions can internally affect communication among humans and personal tendency to memorize the related information. The different emotional states can be analyzed and recognized accordingly from person to person, and this poses a challenge to computational routines. Humans can express their emotions, either vocally or non-vocally. The scope of Automatic Emotion Recognition is enhancing these days in various research fields, and it is also a stimulatory pattern recognition task. But most of the HCI (Human-Computer Interaction) models are not able to identify individual affective states and are not able to decide proper actions for the execution process.

Affective Computing is an active area for the interaction between humans and computers that have been made possible by enhancing the excellence of communication between humans and machines plus refining the brainpower of the network [2]. In medical domains, the emotions of patients might be analyzed and accepted as an indicator of their emotional disorder. Conventionally, "affect" has occasionally been related to lifeless technologies, and psychologists have typically considered this. However, nowadays, new affective features are being apprehended and administered by computer. The extraction of features from EEG (Electroencephalography) signal is essential to recognize emotions. In the human brain, a physical method deviates from the assured form of beliefs and affective actions. Hence, researchers are using signals which mined from the human brain that relates to the change in physical methods for understanding of opinions. Brain typing is one of the standard methodologies for the consideration of physiological practices of the brain that routines EEG signal. EEG signals taken from the mind powerfully interrelated with the central nervous system deliver spatial features of emotional conditions with the change in affective behavior. Traditional affective computing has developed approaches from conventional machine learning algorithms. However, with the advancement of Deep Learning (DL), it is now possible to automatically learn features from various input data. However, DL has become prevalent for the analysis of EEG signals in multiple architectures for the recognition of human emotions.

In this paper we have focused on the classification of emotions using Conventional Neural Networks (CNN). The GoogleNet model is presented to identify affective states from Scalogram. The experiments are performed with benchmark DEAP database having two-dimensional valence and arousal data along with multi-channel EEG data.

The rest of the paper is categories as follows: The literature review given under section 2 is based on the categorization of emotion into three emotion models. Section 3 describes the time-frequency representation of wavelet transform followed by the Deep Learning model under the proposed methodology. Experiments and results are given in section 4, and concluding remarks are given under section 5.

2. Related Work

Emotion is a psycho-physiological practice prompted by conscious or unconscious observation of the situation and frequently linked with mood, happiness, nature, personality, cheerfulness, and enthusiasm. For communication among humans, emotions perform a vital role, and these emotions conveyed either vocally or non-verbal signs like voice inflection, expressions of faces, and gesticulations [3]. The study of emotions has been started by 'Charles Darwin' back in the 19th century. He has proposed the theory of evolution, which signifies that emotions in humans are analogous to emotions in other species. In 1978, Ekman and Friesen extended the research on emotions [4]. They developed the FACS (Facial Action Coding System) to measure human facial expressions and proved the universality of spontaneous feelings. John B Waston [5] postulated the concept of primary and secondary emotions. Ekman, [6] and Plutchik [7] encouraged the concept of complex emotions that derived from some basic human emotions.

The categorization of emotions is based on three significant models: i) Basic emotion model ii) Dimensional emotion model and iii) Appraisal based model.

2.1. Basic Emotion Model

In the basic emotion model, all emotions are considered as discrete emotions that can be distinguished in human facial expressions. Ekman (1993) extended the findings on facial expressions of emotions and characterized each emotion based on their unique features [11]. In the literature, theorists have given many theories that differ based on the number and features of basic emotions i.e., fear, anger, sadness, joy, and disgust. The basic emotion model has the following limitations. i) No phenomena to predict basic emotional conditions. ii) The description of basic and non-basic emotions is unspecified, and iii) The mechanisms for finding emotions are unclear.

2.2. Dimensional Emotion Model

The Dimensional model represents the affective states in one or more dimensional space. Wilhelm Max Wundt (1897) categorized human emotions in three dimensions: "pleasurable versus un-pleasurable", "arousing versus subduing," and "strain versus relaxation"[12]. Harold Schlosberg [1954] has given three dimensions named "pleasantness–unpleasantness", "attention–rejection" and "level of activation" [13]. Most of the dimensional models classify affective states in two dimensions that are 'Valence' and 'Arousal.' Russell's (1980) affect model signifies the inter-relationship between affective dimensions [14]. This model analysed emotions in two-dimensional space containing axes representation. Other well-known models are "Vector Model" and "Positive Activation – Negative Activation (PANA) model" that conceptualized affective states in two dimensions [15]. Lang (1984) stated that human emotions can be recognized not only from single-mode i.e., speech, but another cue is also important, like facial expressions [16]. He presented the emotions in three dimensional, namely Valence, Arousal, and Dominance.

2.3. Appraisal based Model

Appraisal based model extracts and differentiates emotions by continuous and cognitive appraisal of human emotions to an event. The appraisal mechanism is divided into two steps - primary appraisal and secondary appraisal. In primary appraisal, an individual evaluates the situation in two aspects, whether it is positive or irrelevant based on appraisal influence, whereas, the secondary appraisal includes people's analysis of resources [17]. One approach to recognize human emotions is to analyze the abrupt changes in the electrical activity of the brain. It is challenging for a human to control these changes in the brain. Physiological signals can attain these electrical activities. In the existing literature, Electroencephalography (EEG) signals are the most commonly used by the researchers. Kolestra et al. (2011) [18] proposed a novel approach for the selection of stimuli. In this paper, a database for the study of spontaneous emotions was offered for the research community. The database is based on the frontal facial video and ratings of participants. The multimodal data used in the dataset includes thirty-two EEG signals and eight peripheral signals. The participants' ratings are recorded on a continuous scale ranging from 0 to 9. Music video clips are used to stimulate human emotions and the classification was done in the form of arousal, valence, liking, and dominance.

Chung and Yoon (2012) [19] focused on classifying DEAP data into Valence classes and Arousal classes by using statistical learning approaches such as Bayesian Classification. They have organized the valence and arousal data into two/ three categories. Two classes include high/low valence and high/low arousal, whereas three-class include low/normal/high valence and arousal, respectively.

Shang et al. (2013) [20] presented a model based on DBNs (Deep Belief Networks). This model extracted features automatically from physiological data having four channels. It predicted the level of valence, liking, arousal, and valence by using three classifiers and helped in learning the features. They claimed 60.9%, 51.2%, and 68.4% accuracies with two classes for valence, arousal and liking respectively.

Rozgic et al. (2013) [21] represented the EEG signal as a sequence of overlapping segments. The features of the segment level with characteristics of response level centred on an innovative non-parametric Nearest-Neighbor model. As a pre-processing step, they have used Kernel PCA dimensionality reduction. The classification was done using multiple algorithms such as NBN Neighbours, Nearest Neighbours Voting, and RBF SVMs.

Candra et al. (2015) [22] examined the effect of window size on the analysis of EEG signals. Authors have summarized that a large window will result in information load that reasons the feature to be varied up with alternative information. Likewise, the information related to emotion won't be adequately mined if the time window is too short. For extracting time-frequency domain features, they used the standard discrete wavelet transforms (DWT) in electroencephalogram signals.

We have also proposed a technique for the analysis of multimodal physiological signals in our previous study [23]. In this technique, multimodal data decomposed using wavelet transform to extract features. The experiments were conducted with various classification methods i.e. SVM, MLP, and K-NN and achieved 63.7% and 69.62% accuracies for three classes in two-dimensional spaces.

Chunmei Qing et al. (2019) [24], proposed a coefficient-based algorithm for emotional simulation through machine learning. This method worked on EEG signals that represented the relationship between human emotions. Machine learning classifiers were used to extract features. They constructed correlation and entropy curves, which are indicators of emotional activation progression. The model achieved better accuracy in the classification process. In divergence to recognition of emotions through facial expression, we might tend to claim that a variety of emotions can be recognized precisely via physiological signals.

3. Proposed Methodologies

Our proposed methodology involves two main steps. In the first step, the pre-processed multimodal EEG signal is represented by using wavelet analysis. The second step follows the process of classification of emotions by using deep learning. The wavelet-based Deep Learning framework is proposed by considering both frequency and spatial characteristics of multi-channel EEG signals for emotion recognition. The EEG signals are used to perform experiments that are further used to classify human emotions. As we know, that single channel is not sufficient to distinguish various classes of emotions; we have used multi-channel EEG signals. The experiments are performed with benchmark DEAP database having three-dimensional valence, arousal and dominance data along with thirty-two channel EEG data for analysing the emotional states of humans. Initially, the data is down-sampled to 128Hz and Band pass filtered in the range of 4.0 Hz - 45.0 Hz. We are classifying three emotional states given as sadness, cheerfulness, and happiness on two-dimensional affective space [25]. The valence and arousal values are used to create ground truth data in VAD space and fragmentation is experimented on VAD data to segment V-A values into various classes as shown in Table 1. The detailed description of the proposed methodology is given in subsections.

Table 1: Valence and Arousal Class with range

Range of Valence Class	Range of Arousal Class
LV (1-4.5)	LA (1-4.5)
MV (4.5-5.5)	MA (4.5-5.5)
HV (5.5-9)	HA (5.5-9)

3.1. Wavelet Transform

The Continuous Wavelet Transform is utilized to produce Scalogram, a function of frequency and time to get better time localization for short-duration, high-frequency events, and better frequency localization for low-frequency, longer-duration events. It helps in distinguishing noise from the signals by having time-frequency representation. In context to image categorization, images obtained from scalograms are taken as inputs to a Neural Network model. The representation of EEG signals for Valence and Arousal classes is demonstrated in Fig 1.

3.1.1. Time Frequency Representation:

It is challenging to distinguish noise from the signals simply by having time-domain representation. They are composed of slowly varying components interspersed with abrupt adjustments and are regularly buried in noise. Wavelets are handy in analysing these kinds of signals.

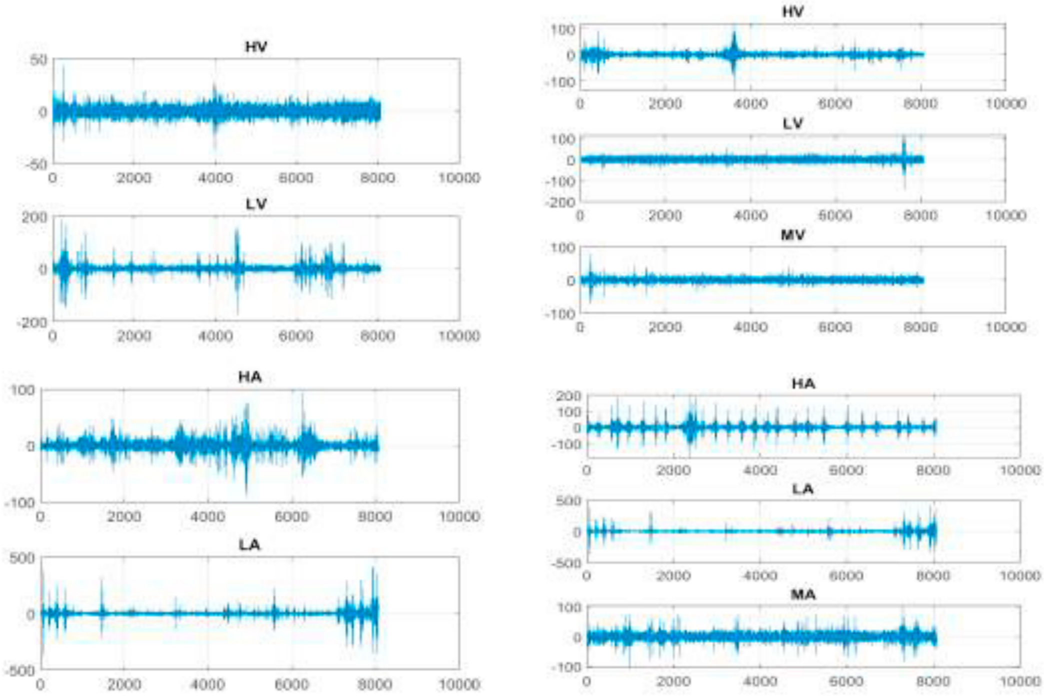

Fig 1: Representation of EEG signals: a) Valence-2 classes: HV, LV b) Valence-3 classes: HV, LV, MV c) Arousal-2 classes: HA, LA d) Arousal-3 classes: HA, LA, MA

Wavelet transform is excellent in representing locally transitory features in both time and frequency domain. The wavelet transform is a joint function of wavelet and time series of a function. Compared with continuous wavelet transform (CWT), 'Short Time Fourier Transform,' has a fixed window size. It is used for stationary signals and is incapable of distinguishing two or more features of the event, plus the frequency information is not very well localized. We use the CWT because it produces fine joint time-frequency analysis and helps in the localization of frequency information. It involves the creation of a time-frequency representation of the EEG signals. These representations are called scalograms.

A scalogram is an absolute value of coefficients obtained using a continuous wavelet transforms of a signal. It shows how much each component of the frequency band has contributed to the energy of the signal over time intervals. As compared to conventional SR methods, like 'Sparse Coding', the wavelet analysis is not examined via learning; nonetheless, it is pre-determined by a mother wavelet ψ. After scaling s and translation u of the mother wavelet, a group of wavelet basis functions ψ s,u can be attained, as equation 1 obtained from [25].

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi((t-u)/s), \quad u \in R,\ s>1 \tag{1}$$

3.1.2. Selection of frequency band

In the proposed work, the multi-channel EEG and peripheral physiological signals are used for feature extraction. The extraction of features in affective computing is essential. Affective states can be analysed using multimodal signals. The EEG signal represents brain waveforms used to classify three emotions like wisely, happy, sad, and cheer. There are five types of frequency bands [26].

We have applied continuous wavelet transform with 'Generalized Morse Wavelet family.' Morse wavelets is a part of analytic wavelets and is defined by two parameters, symmetry and time-bandwidth product [27]. These wavelets have three significant properties: high frequency/time concentration, unique analysis of frequency, and minimized bias for analyzing oscillatory signals. We plotted the scalogram as a function of frequency and time for getting better time localization for short-duration, high-frequency events, and better frequency localization for low-frequency, rather than raw EEG signal. The scalograms are illustrated in Fig 2.

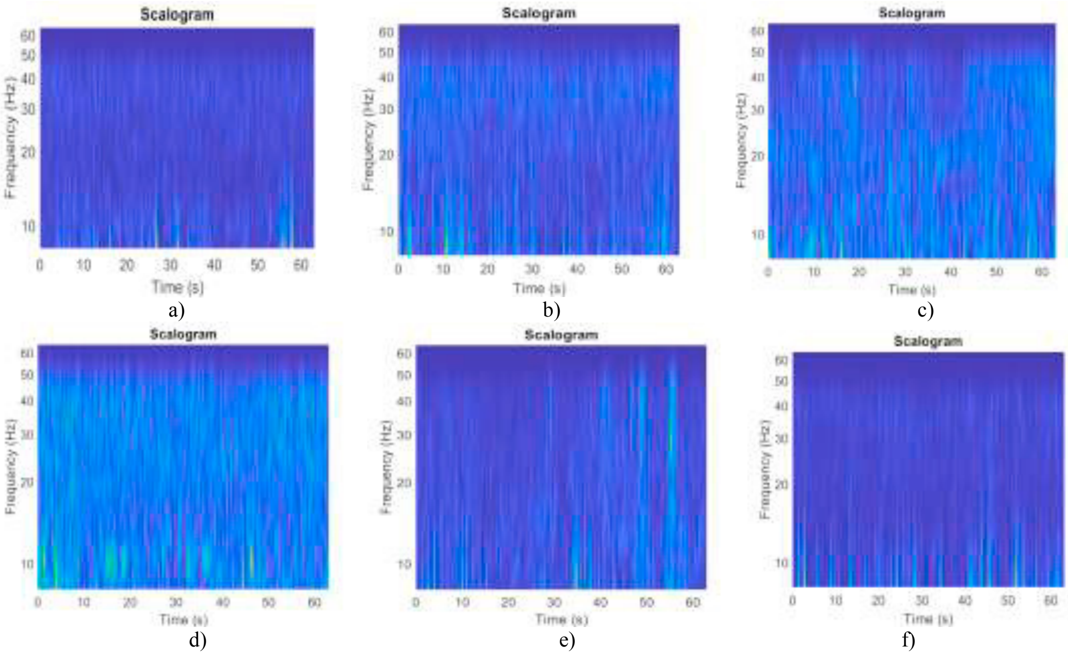

Fig 2: Time-frequency representation of the EEG signals: a) LV at scale 2.51 b) MV at scale 5.5 c) HV at scale 9 d) LA at scale 1 e) MA at scale 5.46 f) HA at scale 9

Frequency is the prime feature for defining EEG signals. In a healthy awake adult, frequency of 8 Hz and higher than this determines the normal brain functioning. The rate of 7 Hz or below in brain waveform is observed in kids during relaxing and sleeping periods. But in adults, these frequencies are considered to be abnormal. Here, specific EEG waveforms with appropriate frequencies of interest are taken. In our work, the signal is filtered above 8 Hz to trace the significant activities and features present. The other reason for the selection of the frequency scale is to remove noise and to satisfy the 'Nyquist criterion'. For example, if the sampling frequency is 128 Hz, then the maximum 64 frequency components can be obtained (according to 'Nyquist's sampling theorem').

3.2 Convolutional Neural Networks

The Convolutional Neural Network (CNN) is one of the most standard models of deep learning. The task of CNN is to learn various high-ordered attributes through convolution. It has been successfully applied in multiple applications, including face recognition, object detection, localization, image categorization, and so on. CNN works very well in the classification and identification of objects in an image, which is the main reason for adopting a deep learning approach all over the world. In context with image categorization, images are taken as inputs to the neural network model. The model acts as a function and provides the probability of features that are contained in images. The CNN is decomposed into a set of functions of layers to extract possible features of images. Each layer of CNN is dedicated to performing a specific task.

CNN model has three major categories:

- Input layer
- Feature-extraction layers
- Fully connected Layers

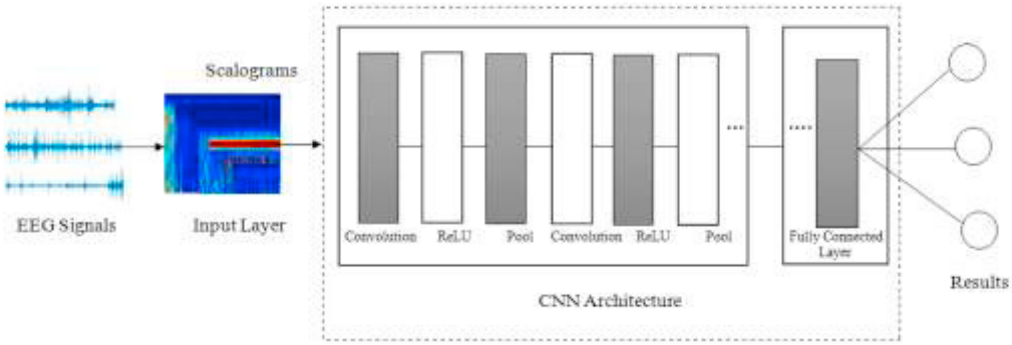

Fig 3: The mechanism of Convolution Neural Network for emotion recognition

3.2.1 Input Layer

This layer accepts three-dimensional data as input in the form of size (width* height) of an image and has a depth that represents the three RGB colors.

3.2.2 Feature-extraction layers

- Convolution layer: The first layer of Feature extraction is convolution layer. A Convolution layer is similar to multi-layered neural network. Instead of going from multiple single neurons to one neuron, the convolution layer transforms input data from the previous layer by using a group of locally connecting nodes (neurons).
- *Pooling Layer*: This layer extracts number of features in the images and constructs high-ordered attributes. It helps in reducing the dimensionalities and also controls over-fitting. Max pooling is referred to as a down-sampling operation and does not have hyper-parameters. This operation is used to resize the input data spatially; however, it does not change the depth dimension. In our work, we used max-pooling layer, with a dropout probability of 0.6 to prevent over-fitting.

3.2.3 Fully connected Layer

It is considered as output of the network. It "flattens" the networks' spatial features and computes class probabilities of higher-ordered features in images.

In our work, we used GoogLeNet architecture of CNN. GoogLeNet is a pre-trained network and it was first proposed by [28], which is used for classification of images, prediction, and recognition. Each layer in CNN is treated as a filter. The mechanism of CNN for emotion recognition is shown in Fig. 3. To retrain the GoogLeNet model, we reconfigured the last two layers at the end of the network model for making it adaptable to emotion classes or labels. The main functions of these layers are to combine features extracted from scalograms into class probabilities and predicted classes. Retrain the new fully connected layer with the number of filters equal to the number of unique labels (classes). New configured layer learns faster than the previous layers by increasing the learning rate parameter. The classification layer gives the output of the model. Reconfigure this layer with a new classification layer with no labels of class. After the whole process, train the network on multimodal data. The training process takes approximately 45-55 minutes on desktop CPU as the data size is large.

4. Experiments and Results

4.1. Database

The DEAP dataset [18] is a multimodal dataset used to analyse human emotional, includes multichannel EEG signals and peripheral physiological signals from 32 subjects. The users' feedback is recorded in terms of valence, arousal, dominance, and liking on a continuous 9-point scale by showing a music video clip. A total of 8064 samples were recorded from 32 individual subjects with 40 trials for each subject. The data format is 32×40×40×8064 that includes 32 subjects, 40 trials, 40 channels and 8064 samples of data.

In our experiment, we classified three emotional states, happiness, sadness and cheerfulness on two-dimensional affective space of DEAP dataset that are Valence and arousal which is proposed by [29]. The data is reorganized on the basis of the emotional states. Therefore, the total data size of the experiment is 128×40×8064 of 128 trials, 40 channels, and 8064 samples.

4.2. Training Model

Training a neural network is an iterative process that includes reducing a loss function. A stochastic gradient descent algorithm is applied to minimize the loss function. The training of the system is carried out by configuring various parameters. 'Learning Rate' controls the network's learning time of a problem. 'MiniBatchSize' acts as the fourth dimension of CNN that states how much of a batch of the training data is to be used in each process. One epoch is one full pass of all the images that are processed over the model. 'MaxEpochs' tells the maximum number of epochs to be used for training and opting for the right number of epochs is not an easy task. Decreasing and increasing the number of epoch results in under-fitting and over-fitting problem respectively. Our model uses the hyper-parameters to train the network.

We performed the classification process epoch-by-epoch. The total 5120 scalograms are divided into training and validation set with 4096 and 1024 scalograms, respectively. Both sets are used to train the GoogleNet model. The details of parameters used to train the system are given in table 2.

Table 2: Training Model Hyper parameters

Name of Hyper parameters	**Parameters**
MiniBatchSize	64
MaxEpochs	10
Total Iterations	640
Learning Rate	1e-4
Dropout Probability	0.6
Loss Function	Cross entropy ex
Activation function in convolution layer	ReLU
Pooling	MaxPooling

4.3. Results Analysis

The aim of analysis and recognition of emotions is twofold. Initially, the verification of the two-dimensional emotion model is done, and after that, the evaluation and prediction of various emotions from EEG signals through scalograms is carried out. The identification of emotions is a multi-task classification problem. The experiments are performed with benchmark DEAP database having two-dimensional valence and arousal data along with multi-channel EEG data. The DEAP dataset is using 32 EEG signals and 8 Peripheral signals.

All the experiments are carried in MatLab 2018a with 64-bit Intel I5 processor with 8 GB RAM. The wavelet transform is done by using a tool named wavelet toolbox, which is available in MatLab.

In this section, we compared our results with the previous research work. This work achieved the best results in the classification of valence class as compared to other several studies. In the case of Valence dimensional space, we attained 92.19% accuracy for two classes named High Valence and Low Valence and 83.59% for three classes: Low, Medium, and High Valence. In the case of Arousal dimensional space, we attained 61.23% accuracy for two classes: High and Low Arousal and 55.56% for three classes: Low, Medium and High Arousal. The performance comparison with existing literature for two (low/high) and three classes (low/medium/high) are shown in table 3 and table 4 respectively.

Table 3: Comparison of performance with existing study (two class: low/high)

Study	**Accuracy (in %)**	
	Valence	**Arousal**
Kolestra et al. [18]	62.0	62.0
Chung and Yoon [19]	66.6	66.4
Shang et al. [31]	51.2	60.9
Proposed model	**92.19**	**61.23**

Table 4: Comparison of performance with existing study (three class: low/medium/high)

Study	**Accuracy (in %)**	
	Valence	**Arousal**
Kolestra et al. [18]	62.0	62.0
Tripathi et al. [30]	58.44	55.70
Gyanendra K Verma et al. [23]	63.47	69.62
Proposed model	**83.59**	**55.56**

The accuracy achieved with the proposed framework is shown in Fig 4.

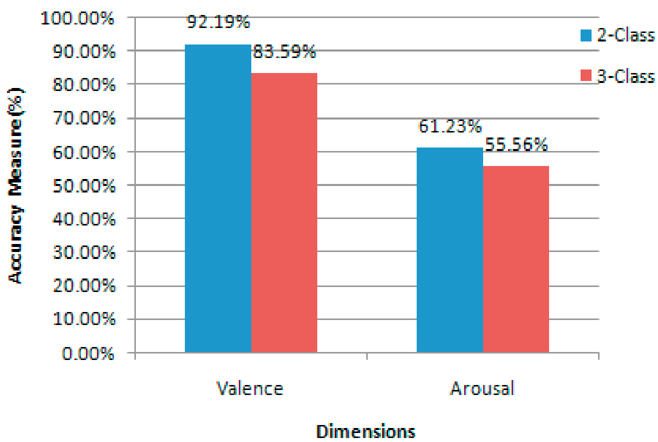

Fig 4: Accuracy for two and three class in 2D space

5. Conclusion

In this paper, we have classified the human emotions in 2D space that is Valence and Arousal. The proposed methodology used multimodal data to analyse affective sates. We presented the GoogLeNet model for the classification of human emotions and extracted features of emotions through scalograms. The experiments are performed with the DEAP dataset and concluded that the proposed framework outperforms in Valence dimensional space as compared to arousal dimensional space. We find that the characteristics of scalograms play a vital role in the classification of emotion classes, and GoogLeNet is more suitable for emotion recognition in two/ three-dimension space. As a future task, we plan to evaluate our proposed model in 3D with other similar datasets with complex emotions.

References

[1] Ekman, P. (1999). Basic emotions. Handbook of cognition and emotion. Wiley, New York, 301-320.

[2] Li, X., Song, D., Zhang, P., Yu, G., Hou, Y., & Hu, B. (2016, December). Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network. In *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 352-359). IEEE.

[3] Liu, Y., &Sourina, O. (2014). Real-time subject-dependent EEG-based emotion recognition algorithm. In *Transactions on Computational Science XXIII* (pp. 199-223). Springer, Berlin, Heidelberg.

[4] Friesen, E., & Ekman, P. (1978). Facial action coding system: a technique for the measurement of facial movement. Palo Alto, 3.

[5] Watson, J. B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of experimental psychology*, *3*(1), 1.

[6] Ekman, P. (1992). An argument for basic emotions. Cognition & emotion, 6(3-4), 169-200.

[7] Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33). Academic press.

[8] Tomkins, S. S. (1962). Affect, imagery, consciousness: Vol. 1. The positive affects. New York: Springer.

[9] Tomkins, S. S. (1963). Affect, imagery, consciousness. Vol. 2: The negative affects. New York: Springer.

[10] Darwin, C. (1872/1998). The expression of the emotions in man and animals (3rd ed.). New York: Oxford University Press.

[11] Ekman, P. (1992). An argument for basic emotions. Cognition and Emotion, 6, 169–200.

[12] Wundt, W. M., & Judd, C. H. (1897). *Outlines of psychology* (Vol. 1). Scholarly Press.

[13] Schlosberg, H. (1954). Three dimensions of emotion. *Psychological review*, *61*(2), 81.

[14] Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178

[15] Rubin, D. C., &Talarico, J. M. (2009). A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory*, *17*(8), 802-808.

[16] Lang, P. J. (1984). Cognition in emotion: Concept and action. In C. Izard, J. Kagan, & R. Zajonc (Eds.). Emotion, cognition, and behavior. New York: CambridgeUniversity Press.

[17] Lazarus, R. S. (1991). Emotion and adaptation. New York: Oxford University Press.

[18] Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T.,&Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18-31.

[19] Chung, S. Y., & Yoon, H. J. (2012, October). Affective classification using Bayesian classifier and supervised learning. In *2012 12th International Conference on Control, Automation and Systems* (pp. 1768-1771). IEEE.

[20] Wang, D., & Shang, Y. (2013). Modeling physiological data with deep belief networks. *International journal of information and education technology (IJIET)*, *3*(5), 505.

[21] Rozgić, V., Vitaladevuni, S. N., & Prasad, R. (2013, May). Robust EEG emotion classification using segment level decision fusion. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 1286-1290). IEEE.

[22] Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., & Su, S. (2015, August). Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. In *2015 37th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 7250-7253). IEEE.

[23] Verma, G. K., & Tiwary, U. S. (2017). Affect representation and recognition in 3d continuous valence–arousal–dominance space. *Multimedia Tools and Applications*, *76*(2), 2159-2183.

[24] Qing, C., Qiao, R., Xu, X., & Cheng, Y. (2019). Interpretable emotion recognition using EEG signals. *IEEE Access*, *7*, 94160-94170.

[25] Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., & Bhattacharya, S. (2017, February). Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In *Twenty-Ninth IAAI Conference*.

[26] Stickel, C., Fink, J., &Holzinger, A. (2007, July). Enhancing universal access–EEG based learnability assessment. In *International conference on universal access in human-computer interaction* (pp. 813-822). Springer, Berlin, Heidelberg.

[27] Lilly, J. M., and S. C. Olhede. "Higher-order properties of analytic wavelets." *IEEE Transactions on Signal Processing*, Vol. 57, No. 1, 2009, pp. 146-160.

[28] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., &Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[29] Gerber, A. J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., & Peterson, B. S. (2008). An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. *Neuropsychologia*, *46*(8), 2129-2139.

[30] Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., & Bhattacharya, S. (2017, February). Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In *Twenty-Ninth IAAI Conference*.

[31] Wang, D., & Shang, Y. (2013). Modeling physiological data with deep belief networks. *International journal of information and*

education technology (IJIET), *3*(5), 505.