Emotion Classification using EEG signals

Divya Gupta
LNM Institute of Information Technology
Jaipur, India
17uec046@lnmiit.ac.in

Kartik Goyal
LNM Institute of Information Technology
Jaipur, India
17uec061@lnmiit.ac.in

Yukti Mittal
LNM Institute of Information Technology
Jaipur, India
17uec142@lnmiit.ac.in

***Abstract*—In contrast to the traditional emotion classification using text, speech and video data, a relatively new concept of using Electroencephalography (EEG) and brain wave patterns is discussed in this paper. EEG is already being used in the medical field to study and cure patients, currently diagnosed with sleep disorders, coma, encephalopathies, epilepsy, and brain death. In this paper the pre-recorded EEG dataset (DEAP) is used. The emotions are classified as Low (i.e. class 0) and High (i.e. class 1) for four parameters. The project was carried out to extract the features from the EEG signals using empirical wavelet transform and entropy computation and remove the biasness of the dataset towards the majority class. Further results are evaluated using most suitable classification techniques and calculating f1 score for evaluation. The achieved score was 78.925 for Valence, 70.04 for Arousal, 74.129 for Dominance and 81.14 for Liking.**

Keywords: Emotion Classification, Electroencephalography (EEG), Empirical Wavelet Transform (EWT), Signal Entropy

I. INTRODUCTION

Sentiment analysis studies of the product reviews, social media comments and chats site back to more than three decades ago [1]. Emotion classification, being a subset of sentiment analysis, has also been a topic of research since a long time. Along with playing an important role in cognitive processing and decision making, emotions are the biological states associated with our body's nervous system which define our feelings, thought process, and response to the surrounding in one or more ways. Traditionally, chats, comments, posts (i.e. text), videos, search and music preference patterns, speech and facial expressions are used for classifying the emotions of a subject. But lately, studies around the brain wave patterns have proved to be a significant and more efficient way to do so. These patterns give more accurate and detailed information about the emotions of the subject, which sometimes, the subject himself/herself is unaware of [2].

Recording and studying brain waves patterns is relatively a new concept used for emotion classification and Electroencephalography (EEG) is used for this purpose. EEG is a typically non-invasive, electrophysiological monitoring method to record the electrical activities of the brain, via electrodes placed on the scalp surface. As the brain cells communicate with each other through electrical impulses, EEG measures the electric field generated when thousands of neurons fire in sync and create this electric impulse. These impulses are strong enough to be measured on the head scalp surface. EEG electrode CAPs are used to carry out the electrode placement, making it easier to affix the electrodes to the scalp precisely having sufficient contact with each other. With the help of it, even the activity within cortical areas even at sub-second timescales can be detected.

EEG based emotion classification is more robust to stimuli changes and psychology factors that result in repression of feelings that leads to oblivious emotions. This method could prove beneficial for medical applications, product development for disabled people like music recommendation systems, smart systems, etc. and brain computer interface applications [3].

In this research work, the pre-recorded and pre-processed EEG signals from the DEAP dataset [4] has been used and the emotions have been classified as Low and High for four labels, named Valence, Arousal, Dominance and Liking. In psychological terms, Valence indicates the intrinsic pleasure or hatred/aversiveness associated with a stimulus, object, or situation. A high valence level maps to positive feelings such as happy, surprised, protected, joyous, satisfied, etc., whereas a low valence maps to the emotions related to being sad, frightened, or angry. Arousal is the quantitative degree of physiological and psychological activation level. On a scale of 0-9, 0 indicates the subject being least excited and 9 indicates the most excited state. Most of the previous works on DEAP dataset use just these two labels, i.e. Valence and Arousal to form two-dimensional models for emotion classification [5]. Further, Dominance indicates the control factor. A high dominance value indicates the emotion of dominating or controlling over the situation, example when a person is aggressive, while a low value indicates being dominated by or controlled by the situation, example when the person is frightened by the stimulus. Lastly, Liking is the scale of simply liking (high value) or disliking (low value) of a stimulus. Analyzing all these four aspects of the emotions give an all-round detailed information about the emotions of the subject based on the changes of the stimuli. In the case of DEAP dataset, as discussed in coming sections, the stimuli are 60-second videos, i.e. the EEG patterns of the subjects were recorded during show of several one-minute videos to the subjects.

In this paper, classification approaches like fusion of Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest for classifying these four labels as 0 (i.e. Low) and 1 (i.e. High), using techniques like Empirical wavelet transform (EWT), entropy based feature extraction, changing threshold, oversampling of minority class to remove the biasness have been proposed. Rest of the paper is organized as: Section II briefs about the previous works in the field. Section III describes the DEAP dataset. Section IV presents the proposed methodology for the research that includes Data Pre-processing and model selection. Section V presents the results, Section VI compares the proposed results with that of the previous works. Finally, the conclusions are presented in Section VII followed by references in Section VIII.

II. Related Work

A lot of previous research have been successfully done in the field of emotion classification VIII [6] [7] [8] [9] [10] [11] [12] [13] [14]. And the data used for these works range widely from text, videos, facial expressions, speech signals to the browser search and other online behavioral patterns [15] [16] [17]. Classification using EEG based technologies recently gained popularity in research field, and many works have been published since the release of DEAP dataset, justifying the reliability of the dataset.

One of the widely used previous works is the bipolar model, where only the Arousal and Valence labels are considered for the classification because degree of pleasure/displeasure and arousal accounted for almost all of the variances for other commonly used scales of measuring affect (e.g., scales of happiness, elation, anger, fear, anxiety, and depression) of emotions. This emotion classification approach is advocated by Russell [5]. Whereas, Plutchik defines eight basic emotion states: anger, fear, sadness, disgust, surprise, anticipation, acceptance, and joy, all other emotions derivable from these basic emotions [18]. Liu and Sourina's work explore real-time EEG based emotion recognition algorithm using Higuchi Fractal Dimension (FD) Spectrum, taking into account the non-linearity of EEG signals along with them being multi-fractal [19]. They explored these characteristics and proposed using FD spectrum, further classifying with SVM technique.

Many other machine learning techniques have been used by researchers to analyze and classify EEG data. According to one of the surveys conducted for this purpose by Rani [20], KNN and SVM are two of the most widely used techniques for classifying EEG data. This claim is supported by the works of others as well [21] [22] [23] [24]. Further researches propose for other widely used techniques, Deep Neural Network, Bayesian Network and Regression tree being some of them [25] [26].

To compare these results solely based on accuracy would not be justified, as it can result into biasness towards the majority class if data is not balanced. Different evaluation metrics, feature selection methods and classification models account for different results, and all these factors are chosen according to requirement and application of research. In this paper, f1 score is used as evaluation metric and the fusion of classification models such as SVM, KNN, Random Forest, feature extraction based on entropies, varying and finding the optimal threshold value have been analyzed.

III. Dataset

In this paper, the DEAP dataset [4] is used. Since the release of this dataset, a lot of research have been conducted on emotion classification using EEG which accounts for the ease of availability and reliability of the dataset. The DEAP dataset presents a multimodal dataset for the analysis of human affective states. The EEG signals of 32 participants were recorded as each watched 40 one-minute-long videos. Then the participants rated each video in terms of the levels of Valence, Arousal, Dominance and Liking on a scale of 0-9.

The original unprocessed data file contains 32 .bdf files (Bio Semi's data format), each with 48 recorded channels at 512Hz for 40 trials. We have used the pre-processed data files which were available in .dat format (pickled python/numpy), well suited for applying classification and regression techniques. Pre-processing involved down sampling to 128 Hz, removal of EOG artifacts, and passing through a bandpass filter of 4-45 Hz. Each trial, i.e. video, is of 63-seconds (including 3 seconds pretrial) (down)sampled at 128Hz so we have 8064 samples for each trial.

The final dataset contains 32 .dat files, one for each subject, containing a hash data structure with keys namely, 'data' and 'labels' and their values in form of arrays. The key 'data' has 8064 numerical values of samples recorded for 40 trials (videos) from each of the 40 channel/electrode, whereas the 'labels' key has values of the 40 trials for the four labels, namely Valence, Arousal, Dominance and Liking.

Table 1 STRUCTURE OF .DAT FILE FOR SUBJECT 1 (SAME FOR ALL 32 SUBJECTS)

	Key	Value (type=array)	Array shape
S01. dat	Data	Trial x channel x samples	40 x 40 x 8064
	Labels	Trial x label rating (valence, arousal, dominance, liking)	40 x 4

The ratings of each label were given between the range of 0-9. So, for classifying the labels as low (class 0) or high (class 1), binary encoding is used based on a set threshold value. In next sections the results with different such values are analyzed, and the best threshold value accounting for maximum accuracy has been proposed.

.

IV. Methodology

Between data collection and finally making prediction using machine learning models a lot is done to get good results. Data cleaning, data preprocessing, choosing a suitable model for the task then training and evaluating the model, tuning hyperparameters are some of the basic steps included. The choice of data preprocessing steps, feature selection procedure, classification models all play a major role in determining better results for emotion classification using EEG as compared to the previous works. Below is the detailed explanation of the methods chosen for this paper.

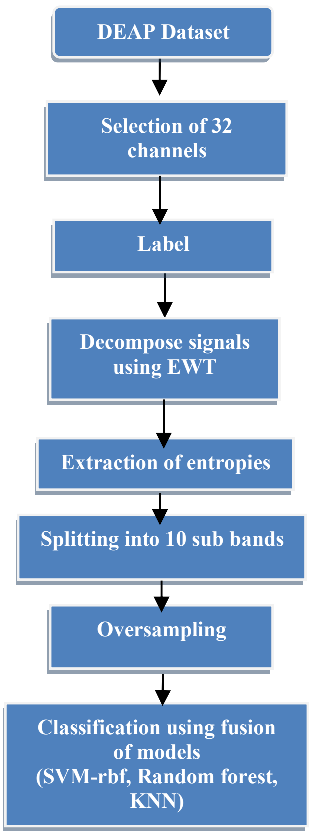

Figure 1 Flow-chart of the experiment

A. Data Preprocessing

The preprocessed files, each for 32 subjects, of DEAP dataset consisted data for 8064 samples from 40 channels each for 40 different videos. Out of those 40 channels, 8 were peripheral channels, so data was taken from left 32 channels for analyzing the emotions. Final structure of dataset is detailed in below table.

Table 2 STRUCTURE OF .DAT FILE AFTER TAKING 32 CHANNELS

	Key	Value (type=array)	Array shape
S01. dat	Data	Trial x channel x samples	40 x 32 x 8064
	Labels	Trial x label rating (valence, arousal, dominance, liking)	40 x 4

1) Changing Threshold Values:

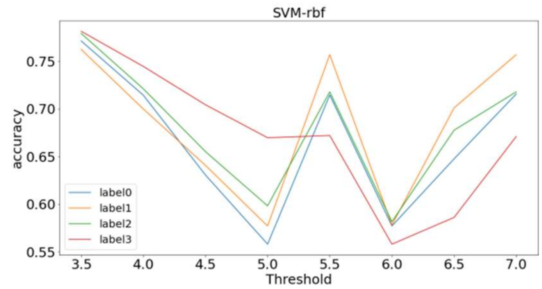

Figure 2 Accuracy score of SVM (rbf kernel) model with varying threshold values

The value of labels, i.e., Valence, Arousal, Dominance and Liking, are a score on the scale of 0-9. 0 signifying lowest and 9 signifying highest value for the label. For the purpose of classifying the value of these labels as Low (i.e., 0) and High (i.e. 1), these scores were converted to Binary encoding, based on comparing these scores with a threshold value, scores greater than this set threshold were encoded as ‘1’ denoting High/1-class and scores less than the threshold were encoded as ‘0’ denoting Low/0-class. In most the previous works, this threshold value was strictly set to a single value, in most cases 4.5, i.e., mean of 0-9. In this paper the recorded results of the optimal model were recorded with varying threshold values ranging from 3.5 to 7.

The best results were seen when threshold value was taken as 3.5, followed by 5.5 and 7. For the rest of the paper, when not specified, the threshold value was set to 3.5 to optimize the results.

The best results were seen when every 32^{nd} feature was taken; thus 8064 features were selected for each 1280 videos.

2) Feature Extraction:

As the data in DEAP dataset consists of amplitude recorded at different time samples, these amplitude values couldn't be directly used as features. Thus, signal processing technique Empirical Wavelet Transform (EWT) was applied to finally extract signal entropies, which were the features that were used to train the model. For this, first the recorded signals were decomposed using EWT, and then Permutation entropies (PE), Spectral entropies (SE), Singular value decomposition entropies (SvdEn), Approximate entropies (ApEn) and Sample entropies (SampEn) were calculated.

3) Empirical Wavelet Transform:

The empirical wavelet transform is a technique that creates segments of signal spectrum using an adaptive wavelet subdivision scheme. This technique was used to create 10 sub-bands, according to frequency, which helped in further analysis. By analyzing each band, the results were found to be different for different models that were applied. To get better results oversampling of minority class was done. Finally, top three performers among these models were used to create a fusion of models.

B. Entropies

Entropy of a signal describe the distribution of signal components [27]. Extracting entropies for a signal is a nonlinear method that has been proved to study EEG signals. After dividing the signals into ten subsequent sub bands, different entropies were calculated for each sub band. Following are the entropies which are used in the proposed model.

1) Permutation entropy (PE):
It is a robust time series tool which provide a quantification measure of the complexity of a dynamic system by capturing the order relations between values of a time series and extracting a probability distribution of the ordinal patterns.

2) Spectral entropy (SE):
It is a measure of signal irregularity, which sums the normalized signal spectral power.

3) Singular value decomposition entropy (SvdEn):
It characterizes information content or regularity of a signal depending on the number of vectors attributed to the process.

4) Approximate entropy (ApEn):
It is a technique used to quantify the amount of regularity and the unpredictability of fluctuation over time-series data.

5) Sample entropy (SampEn):
It is a modification of approximate entropies used for accessing the complexity of physiological time series signals, diagnosing diseased states.

C. Oversampling

Oversampling is the process of duplicating data samples from minority class, and it is done to adjust the class of distribution of a dataset and reduce biasness without loss of actual data. The DEAP dataset is highly biased with most of the data samples accounting to positive values of labels (Valence, Arousal, Dominance and Liking). So, by training and testing on such dataset, traditional models result in predicting positive for all the test cases, eventually leading to large number of False Positives (FP). This is a classic case when a model predicts True value without considering the features, although this could result in high accuracy, but model is not reliable. To make a robust model, oversampling of minority class was carried out. This resulted in a more reliable model with better F1 score.

D. Models

After preprocessing the data, extracting entropies and applying oversampling, the model was trained using classification algorithms like support vector machine(SVM) with RBF kernel, K-nearest neighbors(KNN), Naive Bayes, Logistic Regression and Random Forest. As mentioned earlier, each video was rated for all the four labels on a scale of 0-9, and to classify it as either High (label 0) or Low (label 1), a threshold value was set that fetched more accurate results for future predictions. So, the threshold was varied from 3.5 to 7 in the interval of 0.5 and the results achieved were plotted. In Figure 3, it can be clearly seen that most of the models give their best results for the threshold of 3.5.

SVM with radial basis function (RBF) kernel outperforms other models and results in the highest accuracy of 77.1% and KNN performs the same. Other SVM kernels also followed with the nearly same score of 76.9% and 75.9%.75.9%.

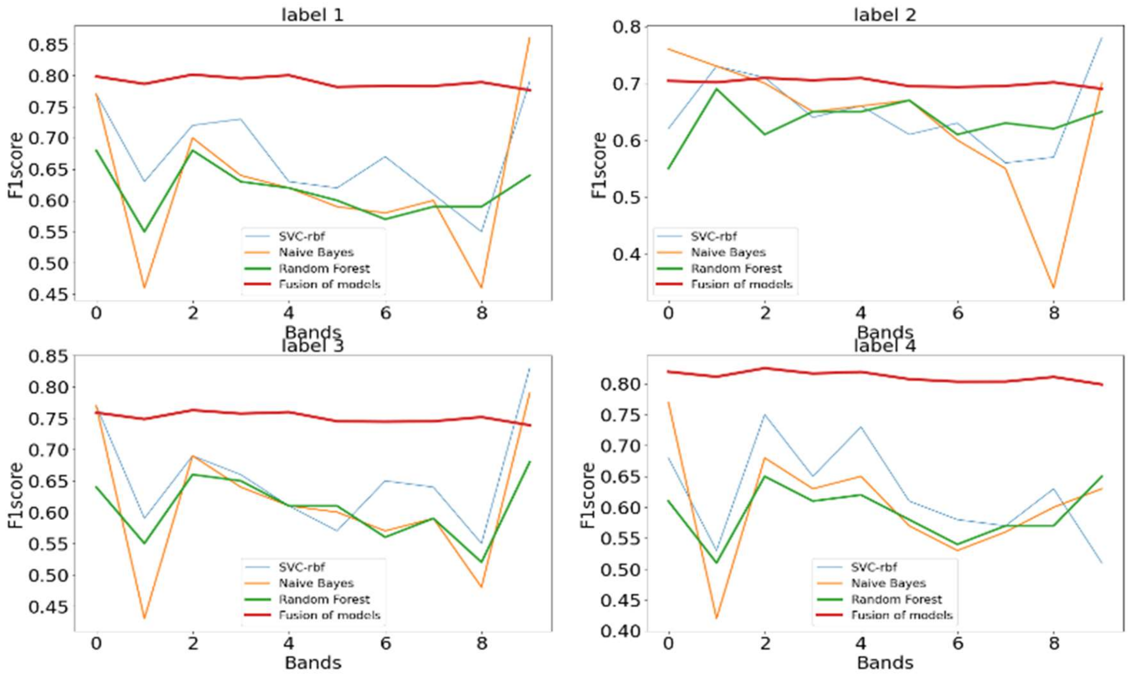

Figure 3 F1 score vs Band for four labels applying different models

SVM classifier plots the data points in n-dimensional space and finds a hyperplane to differentiate between the two classes with maximum margin. Support vectors are the data points that are closest to the hyperplane and help in maximizing the margin of the classifier which means maximizing the distance between the data points of different classes so as to make future predictions with more accuracy. This algorithm also has a feature of ignoring the outliers. It performs well when there is a greater number of features compared to the number of examples which are in our case. SVM uses kernels that take low dimensional input data space and convert it to high dimensional space to convert non-separable data into separable.

All the kernels of SVM use a different equation to determine the hyperplane. RBF kernel function returns the value of the inner product of two points which defines similarity in high dimension space.

$$K(X1, X2) = exp(-\gamma \parallel X1 - X2 \parallel^2)$$

$\parallel X1 - X2 \parallel$ = Euclidean distance between X1 & X2

In contrast, the KNN algorithm, which also performs with the same accuracy as SVM, is a supervised machine learning algorithm that works on the principle of feature similarity. It computes the distance between the new data points for every training example. For computing the distance measure such as Euclidean distance, Hamming distance or Manhattan distance is used. The model picks K entries which are closest to the data point. Then with the help of majority voting, the most common label is considered as a class for the new data point. In our work, we took K value to be 13, to pick 13 topmost similar data points and predict with the most common label.

Another algorithm used in this work is Random Forest. Random forest is also a supervised learning algorithm. The term forest represents that it is an ensemble of decision trees. In this, data is trained using the bagging method in which combination of learning models increases the overall result. Each individual decision tree gives a class prediction and the class with the majority becomes the model's prediction. Trees are taken to be relatively uncorrelated so that they protect each other from their individual errors. While splitting a node in the tree instead of looking for the most important feature it looks for best feature among a random subset of features and it is easy to measure the relative importance of each feature on prediction so it can prevent the overfitting by dropping some not required features.

Naive Bayes is the classifier that works based on Bayes theorem. It takes continuous values associated with each feature that is assumed to be distributed according to a Gaussian distribution. Bayes theorem is defined as

$$P\left(\frac{A}{B}\right) = P\left(\frac{B}{A}\right) * P(A)/P(B)$$

Here, Naive assumption is that A and B are independent sets which imply:

$$P(A, B) = P(A) \times P(B)$$

So, $y = argmax_y(P(y) \prod_{i=1}^{n} \quad P(x_i|y))$

Here, P(y) is class probability and $P(x_i/y)$ is called conditional probability.

V. RESULTS

The result of our fusion of models with unbiased oversampled data is shown below in the table for each sub band and their corresponding label.

Table 3 Results for 10 bands and four labels

	Label 1	Label 2	Label 3	Label 4
Band 1	79.811	70.407	75.895	81.932
Band 2	78.617	70.156	74.859	81.122
Band 3	80.113	70.931	76.287	82.505
Band 4	79.481	70.516	75.732	81.646
Band 5	80.002	70.909	75.975	81.897
Band 6	78.120	69.479	74.514	80.729
Band 7	78.319	69.317	74.443	80.320
Band 8	78.274	69.527	74.513	80.331
Band 9	78.896	70.156	75.184	81.089
Band 10	77.616	69.009	73.888	79.870

VI. Comparisons of Results of previous publication

In this section, proposed results of this paper have been compared with different previous works for the experiment of emotion classification using EEG data. In table 4, V, A, D, L represent labels Valence, Arousal, Dominance and Liking respectively.

Table 4 RESULTS OF PREVIOUS WORK VS PROPOSED

Author	Valance F1-score	Arousal F1-score
Koelstra et al [4]	63.10	65.20
Deniala Girardi [28]	60.5	56.3
Chen et al [29]	68.96	67.83
Proposed	78.90	70.04

VII. Conclusion

In this research work, we presented the observations on the classification of emotions with EEG brain waves using the DEAP dataset. The aim was to improve the f1 score of the previous research that has been done on the same. Although the accuracy is almost the same, F1 score was improved and thus this work was able to reduce False Positives resulting in a more robust model. In this paper we discussed starting with selection of threshold values for labels to extraction of features after applying EWT and finally oversampling the dataset to feed into the classification models. The best classification model for this classification can be the fusion of models SVM, KNN and Random Forest. To compare the results, we took reference of previous works. These analyses will help us in our future work of implying different techniques to get a working application.

VIII. References

[1] M. Mäntylä, D. Graziotin, M. Kuutila, "The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers," *Computer Science Review,* vol. 27, 2016.

[2] N. Shaari, M. Syafiq, M. Amin, O. Mikami,, "ELECTROENCEPHALOGRAPHY (EEG) APPLICATION IN NEUROMARKETING-EXPLORING THE SUBCONSCIOUS MIND," *JAMT,* vol. 13, 2019.

[3] M. Dawn, S. Ramzan , "A Survey of Brainwaves using Electroencephalography(EEG) to develop Robust Brain-Computer Interfaces(BCIs): Processing Techniques and Algorithms," in *9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2019.

[4] S. Koelstra, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing,* vol. 3, no. IEEE, pp. 18-31, 2017.

[5] J. A. Russell, "Affective space is bipolar," *Journal of Personality and Social Psychology,* vol. 37, pp. 345-356, 1979.

[6] S. M. Alarcao, "Emotions recognition using EEG signals: A survey," *IEEE Transactions on Affective Computing,* no. IEEE, 2017.

[7] S. Tripathi, "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset," in *Twenty-Ninth IAAI Conference*, 2016.

[8] G. Chen, "A New Machine Double-Layer Learning Method and Its Application in non-Linear Time Series Forecasting," in *2010 International Conference on Mechatronics and Automation (ICMA)*, 2007.

[9] A. Sohaib, "Evaluating Classifiers for Emotion Recognition Using EEG," in *Found. Augment. Cognit. Lecture Notes Comput. Sci.. 8027*, 2013, pp. 492-501.

[10] Q. Wang, O. Sourina, M. K. Nguyen, "Fractal dimension based neurofeedback in serious games," *The Visual Computer,* vol. 27, pp. 299-309, 2011.

[11] Y. Liu, O. Sourina and M. K. Nguyen, "Real-Time EEG-Based Human Emotion Recognition and Visualization," in *2010 International Conference on Cyberworlds, Singapore*, 2010.

[12] H. Xu, "Affective states classification using EEG and semi-supervised deep learning approaches," in *2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP)*.

[13] Y. Liu, O. Sourina, "EEG-based Valence Level Recognition for RealTime Applications," in *Proc. 2012 Int. Conf. on Cyberworlds*, 2012.

[14] Yisi Liu, Olga Sourina, EEG Databases for Emotion Recognition.

[15] Li, Weiyuan, Hua Xu, "Text-based emotion classification using emotion cause extraction.," *Expert Systems with Applications,* vol. 41, no. 4, pp. 1742-1749, 2014.

[16] Bhaskar, Jasmine, K. Sruthi, and Prema Nedungadi, "Hybrid approach for emotion classification of audio conversation based on text and speech mining.," *Procedia Computer Science,* pp. 635-643, 2015.

[17] G. K. Verma, and U. S. Tiwary, "Multimodal Fusion Framework: A Multiresolution Approach for Emotion Classification and Recognition from Physiological Signals," *NeuroImage,* 2014.

[18] R. Plutchik, Emotions and life : perspectives from psychology, biology,and evolution, 1st ed. Washington, DC: American Psychological Association, 2003.

[19] O. Sourina, "A Fractal-Based Algorithm of Emotion Recognition from EEG Using Arousal-Valence Model.," *Proceedings of Biosignals,* pp. 26-29 , 2011.

[20] B.V.V.S.R.K.K.Pavan, "Electro Encephalogram (Eeg) Recording Systems And The Signal Processing Techniques: A Review," *IJAST,* vol. 29, pp. 1272 - 1281, 2020.

[21] W. Y. Huang, X. Q. Shen, Q. Wu, "Classify the number of EEG current sources using support vector machines," in *Machine Learning and Cybernetics*, 2002.

[22] H. Parvin, "MKNN: Modified k-nearest neighbor," in *World Congress on Engineering and Computer Science (WCECS)*, 2008.

[23] Y. Yu, P. Lai, L. Ko, C. Chuang, B. Kuo, C. Lin, "An EEG-based classification system of Passenger's motion sickness level by using feature extraction/selection technologies," in *The 2010 International Joint Conference on Neural Networks (IJCNN),*, Barcelona, Spain, 2010.

[24] M. Tangermann, I. Winkler, S. Haufe, and B. Blankertz, "Classification of artifactual ICA components," *International Journal on Bioelectromagnetism,* vol. 11, p. 110–114, 2009.

[25] Russell, S. Downey, M. J., "A Decision Tree Approach to Task Independent Speech Recognition," in *Inst Acoustics Autumn Conf on Speech and Hearing*, 1992.

[26] L. E. Brown, I. Tsamardinos, C. F. Aliferis, "A novel algorithm for scalable and accurate bayesian network learning," in *11th World Congress on Medical Informatics (MEDINFO)*, 1992.

[27] W. Y. S. X. L. D. V. L. J. S. J. W. H. S. L. X. Liang Zhenhu, "EEG entropy measures in anesthesia," *Frontiers in Computational Neuroscience,* vol. 9, p. 16, 2015.

[28] Daniela Girardi, Filippo Lanubile, Nicole Novielli, "Emotion detection using noninvasive low cost sensors," in *IEEE*, 2017.

[29] Jing Chen, Bin Hu, Philip Moore, Xiaowei Zhang, Xu Ma, "Electroencephalogram-based emotion assessment system using ontology and data mining techniques," *Applied Soft Computing,* 2015.