

# Emotion Detection from EEG Recordings

Jingxin Liu<sup>1</sup> and Hongying Meng<sup>1</sup>

<sup>1</sup>Dept of Electronic and Computer Engineering  
Brunel University London  
London, United Kingdom

Asoke Nandi<sup>1,2</sup> and Maozhen Li<sup>1,2</sup>

<sup>2</sup>The Key Laboratory of Embedded Systems and Service  
Tongji University  
Shanghai 200092, China

**Abstract**—Human brain behavior is very complex and it is difficult to interpret. Human emotion might come from brain activities. However, the relationship between human emotion and brain activities is far from clear. In recent years, more and more researchers are trying to discover this relationship by recording brain signals such as electroencephalogram (EEG) signals with the associated emotion information extracted from other modalities such as facial expression. In this paper, machine learning based methods are used to model this relationship in the publicly available dataset DEAP (Database for Emotional Analysis using Physiological Signals). Different features are extracted from raw EEG recordings. Then Maximum Relevance Minimum Redundancy (mRMR) was used for feature selection. These features are fed into machine learning methods to build the prediction models to extract the emotion information from EEG signals. The models are evaluated on this dataset and satisfactory results are achieved.

**Keywords**—EEG, Emotion, Emotion Analysis, Affective Computing.

## I. INTRODUCTION

Human emotion is a complex phenomenon which comes from human brain while there is no clear knowledge on its generation mechanism. Psychologists and computer scientists have tried to study it for a long time. Ekman [1], proposed the notion of basic emotions that were universal and found across cultures. Russell [2] proposed a two-dimensional model in which emotions were given co-ordinates denoting the degree of valence (the positive or negative quality of emotion) and arousal (how responsive or energetic the subject is). Other models include Plutchik's wheel of emotions [3], a tree of emotions [4].

Emotion recognition through EEG has a wide variety of practical applications. Chief among these are the use in medicine and scientific research, and the field of affective computing. The latter refers to the incorporation of emotions in human-computer interaction giving machines a degree of emotional intelligence. Possibilities that have been proposed for the use of these machine learning systems include multimedia environments that recognise the emotions of the users, such as recommendation and tagging systems, games and films that respond to the user's emotions, and biofeedback devices that can be worn in the manner of headsets and might help users gain control over their emotional states.

In this paper, we use different machine learning methods to study this relationship from the EEG recordings in a public available dataset DEAP (Database for Emotional Analysis using Physiological Signals) [5]. We will evaluate different

features and different machine learning methods on emotion information extraction from EEG signals.

The rest of the paper is organized as the following. In section 2, related works are reviewed. The proposed method are introduced in details in section 3. Section 4 is the experimental results and section 5 is the conclusion.

## II. RELATED WORKS

Emotion analysis can be treated as a classification problem since the goal of the system is to predict the correct label of emotion. It is thus often a supervised learning task since labels are already assigned to the data by humans, although clustering methods have also been employed [6].

An important part of the study of emotion via machine learning involves the choice of features. Researchers have made use of a variety of features. Jenke [7] surveyed feature selection and extraction across a variety of studies and classified these as time-domain, frequency-domain, time-frequency domain and multi-electrode features. Time-domain features include event related potentials, signal statistics, Hjorth features, non-stationary index, fractal dimension and higher-order crossings; frequency-domain features include band power and higher order spectra; time-frequency domain features include the Hilbert-Huang spectrum and discrete wavelet transforms; multi-electrode features include magnitude squared coherence estimate and differential and rational asymmetries. Frequency domain features are prevalent and appear in the majority of the studies surveyed in the paper, in particular spectral power, but it was found by the authors to have lower performance scores compared to other features.

A very high level of performance was achieved in the study by Valenzi [6] who analysed EEG data from nine participants in response to video stimuli intended to induce the emotional states of amused, disgusted, sad and neutral. A key difference in the use of video stimuli in this study was that in between stimuli a distraction task rather than a relaxation task was used to neutralise the emotional state of the participant and considered to be more effective than a relaxation task. Data was obtained from 32 electrodes. The features extracted from the EEG were spectral power in delta (0.16-4 Hz), alpha (8-13 Hz), lower beta (14-21 Hz), upper beta (21- 30 Hz), and gamma (30-40 Hz) bands. Linear discriminant analysis was used to reduce dimensionality of the feature space. Both supervised and unsupervised learning methods were used. Supervised learning methods were Error Back Propagation, Learning Vector Quantization and SVM. Unsupervised learning algorithms used were Vector Quantization, Fuzzy C-

Means Clustering (FCM), K-Means and K-Medians. A maximum average accuracy of 97.2% was achieved for supervised learning (for SVM) and a maximum average accuracy of 95% for unsupervised learning (for FCM). Average EEG power was computed across the stimuli for the different electrodes and showed larger frontal right symmetry for negative emotions. Electrode reduction was attempted, using only 8 electrodes (6 frontal and 2 temporal) yielding a best rate (using SVM) of 92.5% for individual classification and an average classification rate of 87.5%. It was noted however that the method in its current state was designed only to work offline.

Another study using a different kind of visual stimuli, namely images of natural scenes, labelled positive and negative, was used by Zhang et al [8]. EEG as well as fMRI data was collected from 20 participants. EEG signals were recorded only from the frontal electrodes F3 and F4. The authors noted that prefrontal asymmetry was one of the best known correlates of emotionality with EEG activity. Classification was done using SVM and the highest accuracy achieved was 76.32% for the F4 electrode.

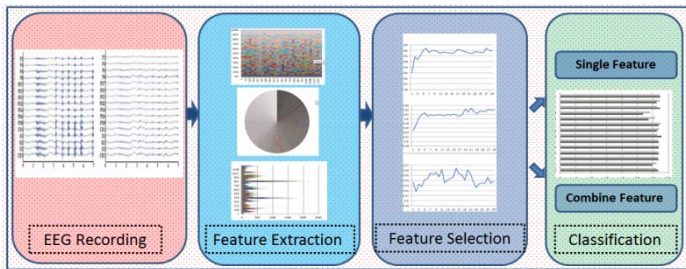
In summary, a considerable amount of research has been done for the prediction of emotion information from EEG signals. However, different datasets were recorded in different situations and the results were not consistent from one to another.

### III. EEG DETECTION SYSTEM

#### A. System Overview

Figure 1 show the proposed system in our study. Firstly, feature extraction from time domain, frequency domain, time-frequency domain and multi electrode were proposed to capture the different information from different methods. Subsequently, different machine learning methods were used to model this relationship between emotions and features. For discrete emotion labels, different classification methods were used. The modelling methods were evaluated and compared.

Figure 1. Overview of EEG Detection System



#### B. Feature Extaction

A widely range of features used for emotion recognition from EEG that have been proposed in the past. We generally distinguish the features extraction methods into time domain, frequency domain, time-frequency domain and Multi-electrode features. Overall, 12 different features have been used.

##### a) Time Domain Features

- Statistical Features (STA)

A total of 7 features are obtained per channel recording leading to 224 features in total. They are straightforward to calculate according to the formulas given in [7], as shown below:

$$\text{Power: } P = \frac{1}{T} \sum_{t=-\infty}^{\infty} |\mathbf{S}(t)|^2. \quad (1)$$

$$\text{Mean: } \mu = \frac{1}{T} \sum_{t=1}^T \mathbf{S}(t) \quad (2)$$

$$\text{Standard deviation: } \sigma = \sqrt{\frac{1}{T} \sum_{t=1}^T (\mathbf{S}(t) - \mu)^2} \quad (3)$$

$$\text{1<sup>st</sup> difference: } \delta = \frac{1}{T-1} \sum_{t=1}^{T-1} |\mathbf{S}(t+1) - \mathbf{S}(t)| \quad (4)$$

$$\text{Normalized first difference: } \bar{\delta} = \frac{\delta}{\sigma} \quad (5)$$

$$\text{2<sup>nd</sup> difference: } \gamma = \frac{1}{T-2} \sum_{t=1}^{T-2} |\mathbf{S}(t+2) - \mathbf{S}(t)| \quad (6)$$

$$\text{Normalized second difference: } \bar{\gamma} = \frac{\gamma}{\sigma} \quad (7)$$

- Higher Order Crossings (HOC)

The aim of the HOC feature is to try to capture the oscillatory pattern of the EEG waveform. The crossings are calculated by subtracting from the mean from the time series and then counting the number of sign changes. It is calculated only for the alpha and beta range so the signal is first filtered through a 10th order Butterworth band-pass filter. The highest order for which the number of crossings was calculated was 10. The first order is the original signal. For subsequent orders, the new signal is obtained by taking the difference between consecutive points of the previous signal and the number of crossings is then computed for this signal. When taking a difference, one point is lost so in order to retain the same number of points at each level, it is necessary to start (in this case) 10 points from the start of the signal [9].

- Fractal Dimension (FD)

This feature also seeks to capture information about the shape of the signal. The formal way of defining dimension is to consider the scaling relationship between units of measurement and the number of such units required to measure a shape, as shown in the following equation:

$$N \propto \epsilon^{-D} \quad (8)$$

In the above  $\epsilon$  denotes the amount by which the unit of measurement is increased or reduced,  $N$  denotes the number of the newly scaled units of measurement required to measure the same shape and  $D$  is the fractal dimension.

The method used in [7] was the Higuchi algorithm. The method is described in more detail in [10]. To use the method for the signal  $S(t)$ , a new time series is constructed as follows

$$S_k^m = S(m), S(m+k), S(m+2k), \dots, S(m + \text{floor}[(N-m)/k]*k) \quad m = 1, 2, \dots, k \quad (9)$$

In the above the function floor rounds down the value of the argument to the nearest integer. The length of the curve is then given by

$$L_m(k) = \left\{ \left( \sum_{i=1}^{\text{floor}\left[\frac{N-m}{k}\right]} |S(m+ik) - S(m+(i-1)k)| \right) \frac{N-1}{\text{floor}\left[\frac{N-m}{k}\right]*k} \right\} / k \quad (10)$$

In the above the term  $\frac{N-1}{\text{floor}\left[\frac{N-m}{k}\right]*k}$  is a normalisation factor for the curve length of subset time series. For each time interval  $k$ , the value of the length of the curve is averaged over all values of  $m$  to obtain the value  $\langle L(k) \rangle \propto k^{-D}$ , which is analogous to  $N \propto \epsilon^{-D}$  above. We can estimate the fractal dimension  $D$  by calculating  $\langle L(k) \rangle$  for different values of  $k$  plotting it against  $k$  and fitting a line to the points. The magnitude of the gradient of this line is the estimate for  $D$ .

- Hjorth Feature (Hjorth)

These are simple statistical features computed using the following expressions:

$$\text{Mobility: } \sqrt{\frac{\text{variance}(S(t))}{\text{variance}(S(t))}}, \text{ Complexity: } \sqrt{\frac{\text{mean}(S(t))}{\text{mean}(S(t))}}$$

As in [7], an additional feature, ‘Activity’, was omitted because it is just square of the standard deviation (i.e. the variance) and the standard deviation is already included among the statistical features above.

- Non-stationarity Index (NSI)

NSI is a measure of fluctuation dynamics that is used to evaluate the change in time of the local average [11], independent of the magnitude of the fluctuation.

b) *Frequency Domain Features*

- Power Spectral Density (PSD)

It is usually computed for a number of frequency bands and used as an indicator of the extent of brain activity within each of these bands. The DEAP data already been down-sampled to 128 Hz and low-pass filtered to remove frequencies above the desired range. The power spectral density can be regarded as either of the following [12]:

The averages of the square of the magnitude of the Fourier transform:

$$S_x(f) = \lim_{T \rightarrow \infty} E \left\{ \frac{1}{2T} \left| \int_{-T}^T x(t) e^{-j2\pi ft} dt \right|^2 \right\} \quad (11)$$

The Fourier transform of the autocorrelation function:

$$S_x(f) = \int_{-T}^T R_x(\tau) e^{-j2\pi f\tau} dt \quad (12)$$

$$\text{Where, } R_x(\tau) = E\{x(t) * x(t + \tau)\} \quad (13)$$

c) *Time-Frequency Domain Features*

- Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) is an alternative method to power spectral density for measuring the prominence of different frequencies in the EEG. The discrete wavelet transforms involves a cascade of processing steps. At each stage there is a high-pass and a low-pass filtering process. From the high-pass filter detail coefficients are recovered for the higher frequency range whilst the low-pass filter results in what are called approximation coefficients, which are passed onto the next stage. The stage whose coefficients include the desired frequency bands (alpha, beta and gamma in this case) depend on the sampling frequency of the original signal, in this case 128 Hz. Using the detailed coefficients, three feature vectors were created comprising values related to the energy, Root Mean Square (RMS) and entropy [13, 14]:

Energy values were calculated using the following expressions: Band energy:

$$E_{band} = \sum_{-\infty}^{\infty} |S(t)|^2 \quad (14)$$

Total band energy across alpha, beta and gamma bands:  $E_{total} = E_{alpha} + E_{beta} + E_{gamma}$  (15)

From the energy values, the Recursive Energy Efficiency (REE) was obtained for each of the alpha, beta and gamma bands:  $REE = \frac{E_{band}}{E_{total}}$

Two further values,  $\log(REE)$  and the absolute value of  $\log(REE)$  were also computed from the above. These three values were included together in a single feature vector.

$$\text{Root mean square: } RMS(j) = \sqrt{\frac{\sum_{i=1}^j \sum_{n_i} D_i(n)^2}{\sum_{i=1}^j n_i}} \quad (16)$$

The REE, RMS and entropy features were obtained as three feature vectors to be used separately in the classifiers.

d) *Multi Electrode Features*

It is important that the existence of interconnections between different parts of the brain is also considered.

- Differential Asymmetry and Rational Asymmetry (DA,RA)

These two features are calculated based on difference of power spectral density and ratio of symmetrical electrodes.

- Magnitude Squared Coherence Estimate (MSCE)

This feature takes into account the cross PSD between pairs of electrodes according to the equation [15]:

$$C_{ij}(f) = \frac{|P_{ij}(f)|^2}{P_i(f)P_j(f)} \quad (17)$$

Only magnitude of the cross power spectral density is required and since  $P_{ji} = P_{ij}^*$ ,  $|P_{ji}| = |P_{ij}|$  so that  $C_{ij} = C_{ji}$ . Also the value of  $C_{ij}$  for  $i = j$  is simply  $P_i$ , the power spectral density, which is considered separately, so that is also neglected from this set of features.

### C. Feature Selection

We proposed that use Maximum Relevance Minimum Redundancy (mRMR) feature selection method for the combination feature of all these 12 kinds of features. The combine feature dimension is 1952. This method use mutual information to characterize the suitability of features proposed by Ding and Peng [16]. Mutual information between two variables is shown above. The mRMR method is used to optimize two criteria simultaneously: Maximal-relevance criterion D, which means to maximize average mutual information between each feature and the specific label. The Minimum-redundancy criterion R means to minimize the average mutual information between two features [17]. The algorithm finds near-optimal features using forward selection. Given an already chosen set  $S_k$  of  $k$  features, the next feature is selected by maximizing the combined criterion D-R:

$$\max_{x_j \in X - S_k} [I(x_j; y) - \frac{1}{k} \sum_{x_i \in S_k} I(x_j; x_i)] \quad (18)$$

We adopt mRMR on each feature and combine feature. However, we don't know how many features to choose the results are reaching the best. So we proposed that pick from 1 to the last dimension for each features and combine feature under mRMR[18].

### D. Classification

We adopt K-Nearest Neighbour (KNN) and Random Forest as the classification method.

Random Forest is based on decision tree. The tree consists of nodes and branches. Each node corresponds to a feature and each branch to a value taken by the feature. For continuous valued features, as here, the features need to be divided into discrete ranges. At each branch the feature having the highest information gain is selected for the next node. The information gain is calculated as follows:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (19)$$

$S$  denotes the set of all training example,  $A$  denotes the feature for which the gain is being calculated,  $S_v$  denotes the subset of training examples for which the feature  $A$  takes on the value of  $v$ . The entropy term is calculated as follows:

$$Entropy(S) \equiv \sum_{i=1}^c -p_i \log_2 p_i \quad (20)$$

Here  $p_i$  denotes the proportion of training examples belonging to the  $i$ -th class out of  $c$  classes. The tree starts off with a root node which is the feature that has the highest information gain out of all the features based on all the training examples. Random Forest method generates a number of decision trees. Each tree is trained on a subset of the training examples generated by sampling with replacement so there tends to be overlap between the subsets for each tree. During classification, each tree assigns a label and the final classification is given as:

$$(c|\mathbf{v}) = \frac{1}{T} \sum_{t=1}^T P_t(c|\mathbf{v}) \quad (21)$$

Averaging the decision boundaries that result from each tree can produce a superior result to that generated by single trees [19].

## IV. EXPERIMENTAL RESULTS

### A. The DEAP dataset

In DEAP dataset is a multimodal dataset for the analysis of human affective states [5]. EEG recordings from 32 channels, peripheral physiological signals and frontal face videos were obtained from 32 participants whilst watching 40 music videos. The videos were selected to evoke one of the four of the following categories of emotion: high valence, high arousal; high valence, low arousal; low valence, high arousal; low valence, low arousal. The EEG data was processed by average referencing, down-sampling to 256 Hz and high-pass filtering to 2 Hz cut-off frequency. Changes in power relative to the pre-stimulus period was computed and averaged over the theta (3-7 Hz), alpha (8-13 Hz), beta (14-29 Hz) and gamma (30-47 Hz) bands. Figure 2 shows a part (5 seconds) of an EEG recording (63 seconds) with 32 channels.

TABLE I. FEATURES

	Features	
	Feature name	No. of features
<b>Time Domain</b>	Statistical features	224 (7x 32 channels)
	Higher order crossing	320(10x32 channels)
	Fractal Dimension	32 (1x32 channels)
	Hjorth features	64 (2x32 channels)
	Non-stairarity Index	32 ( x32 channels)
<b>Frequency Domain</b>	Power Spectral Density	128 (4x32 channels)
<b>Discrete Wavelet Transform</b>	REE	288 (9x32 channels)
	RMS	128 (4x32 channels)
	entropy	128 (4x32 channels)
<b>Multi Electrode features</b>	Differential Asymmetry	56
	Rational Asymmetry	56
	MSCE	496

### B. Data Processing

In this paper we address the binary classification problem that results after we threshold the self-assessments following [5]. The affective label will be set to high if the rating is above 5. If the rating is equal or lower than 5, the corresponding affective label will be set to low. Thus for each trial, two labels were generated. HV (high valence) or LV (low valence) was to describe the affective level in valence space, and HA (high arousal) or LA (low arousal) was to describe the affective level in arousal space. Identifications of valence and arousal levels are addressed as two independent tasks in this paper. Both of the two tasks are posed as two-class problems.

### C. Feature Extraction

EEG signals with whole duration except the 3 seconds (use last 60s) prior to representing are used for extracting connectivity features. Then we use KNN to calculate the accuracy of each feature.

### D. Results

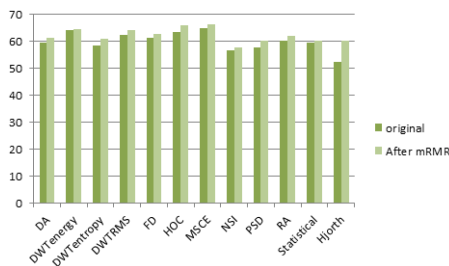
In what follows we will present the classification results on the DEAP database under different settings of our algorithm. RF is setting as 1000 trees. All the methods are running under 10-folds Cross validation. Table II shows the classification results for each feature original and after mRMR by KNN, (K =5).

TABLE II. ACCURACY FOR EACH FEATURE (KNN)

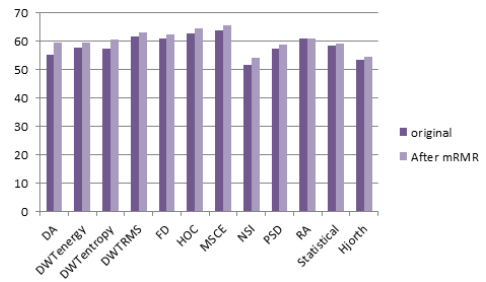
FeatureS	Emotions			
	Arousal	Valence	Arousal	Valence
DA	59.5	55.2	61.4	59.5
DWTenergy	64.1	57.6	64.5	59.7
DWTentropy	58.4	57.5	61.0	60.6
DWTRMS	62.3	61.6	64.1	63.2
FD	61.3	60.9	62.9	62.3
HOC	63.4	62.6	66.1	64.7
MSCE	64.9	64.0	66.2	65.7
NSI	56.8	51.6	57.9	54.0
PSD	57.6	57.5	60.4	58.9
RA	60.1	61.0	62.1	61.1
Statistical	59.4	58.3	60.1	59.3
Hjorth	52.5	53.5	60.2	54.4

As shown above in Figure II both arousal and valence results are increased after mRMR. The best performance of all features is Magnitude squared coherence estimate, reaches 66.17% for arousal and 65.7% for valence which after the mRMR selection. Since the 12 kinds of feature extracted from different area, consider that, we proposed time-domain combine, time frequency-domain combine, and all features combine. Random forest and KNN are used as the classification methods for all combine features. The results are shown below for this page. From the result we can see that after mRMR selection both arousal and valence reach the highest performance 71.23% and 69.97% as expected.

Additionally, the accuracy of valence levels identification is higher than accuracy of arousal. Finally, there are several works that use DEAP database for various purposes. This is consistent with the results presented in. Koelstra's [5] result is 62.0% and 57.6% for valence and arousal with 32 participants. Zhang's [20] result is 75.19% and 81.74% on valence and arousal with selected 10 participants. Chung's [21] result is 66.6% and 66.4% for valence and arousal with 32 participants and use. It should be mentioned here that our work is based on 10-folds Cross Validation and 32 participants. This is different from others' tasks so that comparison is unavailable.



(a) Arousal



(b) Valence

Figure 2. Accuracy of Original and after mRMR on Each Feature

TABLE III. ACCURACY ON COMBINED FEATURES

Combine Features		Arousal		Valence	
		KNN	RF	KNN	RF
Original	Time domain	63.9	64.2	62.5	63.2
	Time-Frequency	64.8	66.1	63.6	65.1
	All	65.2	67.8	64.9	67.0
After mRMR Selection	Time domain	65.4	66.3	64.7	65.1
	Time-Frequency	65.2	67.9	64.3	66.1
	All	68.4	<b>71.2</b>	66.3	<b>69.9</b>

### V. CONCLUSION

In this paper, different features have been extracted from EEG recordings and machine learning methods such as KNN and RF have been used for emotion level (Arousal and Valence) detection. In addition, mMMR feature selection method was used for the dimension reduction of the feature vectors. From the experimental results, it can be found that the proposed method is useful to detect the emotion information from EEG recordings in a good accuracy. Within the results, it can be seen that feature selection can improve the performance by removing some irrelevant feature vectors. The combined feature produces the best results.

For future improvements, more features can be added into the system and the performance might be improved further. In addition, based on different kinds of features, fusion method might be used in the future. Furthermore, the contribution of each channel can be further analyzed and channel selection might make further improvement.

### Acknowledgement

This research is also partially supported by the 973 project on Network Big Data Analytics funded by the Ministry of Science and Technology, China. No. 2014CB340404.

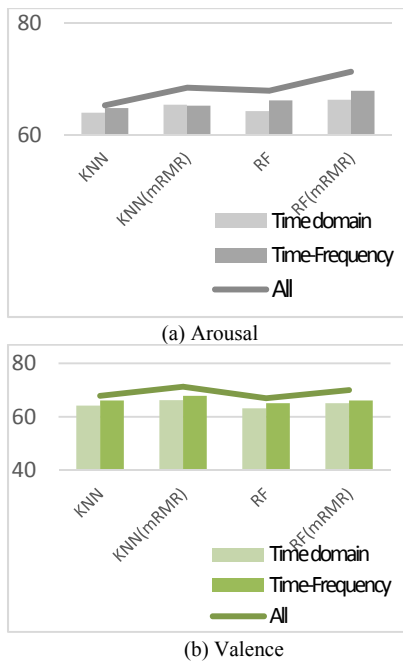


Figure 3. Accuracy of Combine Features for KNN and RF

#### REFERENCES

[1] P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn and P. E. Ricci-Bitti, "Universals and cultural differences in the judgments of facial expressions of emotion." *J. Pers. Soc. Psychol.*, vol. 53, pp. 712, 1987.

[2] J. Posner, J. A. Russell and B. S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Dev. Psychopathol.*, vol. 17, pp. 715-734, 2005.

[3] R. Plutchik, "The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *Am. Sci.*, vol. 89, pp. 344-350, 2001.

[4] P. Shaver, J. Schwartz, D. Kirson and C. O'connor, "Emotion knowledge: further exploration of a prototype approach." *J. Pers. Soc. Psychol.*, vol. 52, pp. 1061, 1987.

[5] S. Koelstra, C. Mühl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *Affective Computing, IEEE Transactions On*, vol. 3, pp. 18-31, 2012.

[6] S. Valenzi, T. Islam, P. Jurica and A. Cichocki, "Individual classification of emotions using EEG," *Journal of Biomedical Science and Engineering*, vol. 7, pp. 604, 2014.

[7] R. Jenke, A. Peer and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *Affective Computing, IEEE Transactions On*, vol. 5, pp. 327-339, 2014.

[8] Q. Zhang and M. Lee, "Emotion recognition in natural scene images based on brain activity and gist," in *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on*

*Computational Intelligence). IEEE International Joint Conference On*, 2008, pp. 3050-3057.

[9] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *Information Technology in Biomedicine, IEEE Transactions On*, vol. 14, pp. 186-197, 2010.

[10] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, vol. 31, pp. 277-283, 1988.

[11] P. H. Franses, "Seasonality, non-stationarity and the forecasting of monthly time series," *Int. J. Forecast.*, vol. 7, pp. 199-208, 1991.

[12] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Prentice hall Upper Saddle River, 1997.

[13] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi and D. Hazry, "EEG feature extraction for classifying emotions using FCM and FKM," *International Journal of Computers and Communications*, vol. 1, pp. 21-25, 2007.

[14] M. Murugappan, N. Ramachandran and Y. Sazali, "Classification of human emotion from EEG using discrete wavelet transform," *Journal of Biomedical Science and Engineering*, vol. 3, pp. 390, 2010.

[15] G. C. Carter, C. H. Knapp and A. H. Nuttall, "Estimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," *Audio and Electroacoustics, IEEE Transactions On*, vol. 21, pp. 337-344, 1973.

[16] H. Peng, F. Long and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions On*, vol. 27, pp. 1226-1238, 2005.

[17] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 3, pp. 185-205, 2005.

[18] L. Jack and A. Nandi, "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 16, pp. 373-390, 2002.

[19] J. Shotton, T. Kim and B. Stenger, "Boosting and randomized forests for visual recognition," in *Tutorial at Int'l Conf. Computer Vision*, 2009, .

[20] X. Zhang, B. Hu, J. Chen and P. Moore, "Ontology-based context modeling for emotion recognition in an intelligent web," *World Wide Web*, vol. 16, pp. 497-513, 2013.

[21] S. Y. Chung and H. J. Yoon, "Affective classification using bayesian classifier and supervised learning," in *Control, Automation and Systems (ICCAS), 2012 12th International Conference On*, 2012, pp. 1768-1771.