

# Density Based Mode Within Radius and Its Application to EEG based Emotion Recognition

Parampuneet Thind,<sup>1</sup> Vaibhav Katturu,<sup>1</sup> Teryn Cha,<sup>2</sup> Sung-Hyuk Cha,<sup>1</sup>

<sup>1</sup>Computer Science Department, Pace University, New York, NY, USA

<sup>2</sup>Computer Sciences, Essex County College, Newark, NJ, USA  
vk38221n@pace.edu, pt54854n@pace.edu, yan@essex.edu, scha@pace.edu

## Abstract

Statistical parameters such as center points are widely used to represent patterns in machine learning. There are three popular center points: mean, median, and mode. Mean and median are pervasively used as features to represent patterns, but mode is not because of its weakness. One of the major problems in the conventional definition of mode occurs when the distribution is multimodal with tied multiple maximum points. Hence, in this paper we propose a modified version of mode, which utilizes the density within a radius. The effectiveness of proposed mode is demonstrated in the emotion recognition application using the Electroencephalogram (EEG). An experimental result of amalgamating statistical parameters with a new modified parameter suggests the superiority of the proposed statistical parameter over the traditional measures.

## Introduction

Statistically, central tendency measures provide an ingenious idea of how the entity of interest looks as a whole. The center of a can be a very valuable piece of information indicating how exactly the distribution is biased, since whichever value the data revolves around is essentially a bias. Center points are prevalently used as features to signify patterns. Mean and median are frequently used as features or machine learning algorithms however mode is seldom used.

Two possible reasons for why mode is often ignored in machine learning and pattern recognition include the multimodality and quantization problems. Suppose that quantized data are multimodal, which means that they may contain more than one value for mode. Most conventional statistical or machine learning software packages and built-in libraries of programming languages select the first maximum frequency element in case of multimodal data. The second problem is that the mode value varies significantly depending on how data are quantized. To deal with the mentioned setbacks, a modified version of mode is proposed, which utilizes, density within a radius to evaluate the value of mode.

So as to demonstrate the superiority of the proposed modified version of mode, EEG brain dataset focusing on recognition of human emotions is used. To achieve accurate brain signal readings, a neurologist requires accurate observation

skills so as to identify characteristic signals. Variations in diagnostic observations indicated by EEG signals may lead to incorrect readings. Modern techniques are required for accurate readings and effective observations. For a long time now, it has been a practice to develop machine learning techniques to help study brain signals for emotion recognition.

The performance of the machine learning model in a given task is calculated by a performance metric that is improved with training over time. At the end of the training period, the model can be used to predict the given emotion. For this research, DEAP, a widely used dataset for the purpose of studying emotion recognition using brain signals is considered. A total of 7 statistical parameters are computed for performing statistical feature extraction with the goal for achieving maximum information with every parameter. For the purpose of extracting features from the dataset, a feature vector 40 STFT ranging from  $f = 4\text{Hz}$  to  $f = 45\text{Hz}$  is considered for each electrode, the resultant was passed through each statistical function to get their corresponding output. A 1D CNN is effectively used for deriving features from a fixed-length segment of the overall dataset and implemented for learning purposes.

The rest of the paper is organized as follows. Section “Related Work” provides the literature review on the previous works pertaining to several application conducted utilizing the study of statistical parameters to identify patterns in data. In section “Mode”, the statistical parameter, *mode* is reviewed with illustration and *density based mode* is introduced followed by the algorithm used in the process. Section “EEG Emotion Dataset” explains the DEAP dataset and EEG emotion dataset. Section “Experimental Results and Discussion” discusses the three different experimental results and compares studies based on using Mode as a statistical measure. Finally, section “Conclusion” concludes this work.

## Related Work

Statistical parameters are used exceedingly for extracting useful information from data in many applications such as recognizing patterns in drilling time series (B.Esmael et al. 2012). In this paper, a time series data is transformed into statistical features from which selected features are used to train a classifier used for pattern recognition. An overall of 22 statistical functions were calculated and used to reduce

the dimensionality of the data by building a high-level representation where a set of significant features are calculated.

In (Lambrou et al. 1998), an ensemble of 8 statistical features was collected from music dataset to distinguish between the sounds of 3 instruments. A study is conducted to establish which measures clearly distinguish between three different musical styles in conjunction with statistical pattern recognition.

Statistical parameters have shown promising results when used to extract features from Images. The technique has been applied for classification of Image spam in (Soranamageswari and Meena 2010) where artificial neural networks was employed to achieve best results with minimum false positives. Image statistical feature histogram and mean value of a block of an image are considered for image spamming detection.

In (Zazzi, Ulivi, and Prosperi 2007), statistical validation is used to assess the robustness of results where different machine learning, and feature selection methods are applied to classification of treatment success and predicting the in-vivo effect of genotypic drug resistance of Human Immunodeficiency Virus type-1 (HIV-1) on response to antiretroviral therapies represents a major clinical issue.

In (Islam et al. 2013), Islam et. al. uses several statistical measures including skewness and kurtosis which indicate the largest dispersion in different mental states and help to evaluate different human emotions following which frequency analysis shows how the ranges of magnitude vary with different frequency components and on the basis of magnitude ranges different emotional states are identified.

In (Yuen et al. 2009), Yuen et. al. proposed an emotion stimulation experiment using visual stimuli in which a total of six statistical features are computed from the brain signals and back-propagation neural network is applied for the classification of human emotions achieving an overall classification accuracy of 95%. Multiple statistical parameters have been studied in (Yu and Zuo 2015) for Traffic flow detection. The variance in gray value of current image and constructed background is employed as the principal statistical parameter to identify the automobile. The texture-based statistical parameter is applied as an enhancement to augment the robust in contradiction of the moving shadow.

## Mode

The mode of a sample is the element that occurs most often in the collection. This definition is only valid if the samples are discrete valued data. Let  $n$  be the number of samples in a set  $X$ ;  $n = |X|$ . Let  $V(X)$  be the set of unique values in  $X$  and  $m$  be the number of possible discrete values in  $X$ ;  $m = |V(X)|$ . The discrete valued data means that  $\forall x \in X, x \in V(X) = \{v_1, v_2, \dots, v_m\}$ . The mode in a discrete valued data can be defined formally as follows:

$$\text{mode}(X) = \underset{x \in X}{\operatorname{argmax}} \sum_{v \in V(X)} \delta(x, v)$$

$$\text{where } \delta(x, v) = \begin{cases} 1 & \text{if } x = v \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

For example, if  $X = \{4, 4, 4, 5, 5, 5, 6, 7, 7, 7, 7, 8, 8\}$ , then  $V = \{4, 5, 6, 7, 8\}$ ,  $n = 14$ ,  $m = 5$ , and  $\text{mode}(X) = 7$  because the value '7' occurs most frequently.

## Problems in Conventional Mode

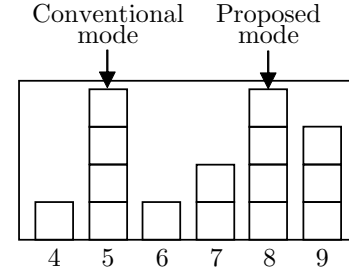


Figure 1: Problem of Mode when data is multimodal.

One of the obvious problems of the mode definition in eqn (1) occurs when data are multi-modal as shown in Figure 1. There are two values which have the highest frequency: 5 and 8. Most statistical packages or built-in library functions in most programming languages choose the highest frequency variable occurring first as the mode, which often may not be the case.

When the data is real valued numbers, i.e.,  $m = \infty$  and  $V(X) = [s, e] = \{v \in \mathbb{R} \mid s \leq v \leq e\}$ , the mode definition in eqn (1) cannot be applied. It is limited to the discrete valued data only and this limitation is another problem.  $s$  and  $e$  are starting and ending possible real values for the elements in  $X$ , respectively.

One can compute the mode after discretizing or quantizing the values. Let  $H^q(X)$  be the frequency histogram of  $X$  after quantizing the values by the quantization size of  $q$ . Let  $H_i^q(X)$  be the  $i$ th bin frequency in the histogram,  $H^q(X)$  as defined in eqn(2).

$$H_i^q(X) = \sum_{x \in X} f_i(x) \quad \text{where}$$

$$f_i(x) = \begin{cases} 1 & \text{if } s + q \times (i - 1) \leq x < s + q \times i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Let  $m_q$  be the number of bins in the  $H^1(X)$  histogram.

$$m_q = \left\lceil \frac{e - s}{q} \right\rceil \quad (3)$$

The mode for the real valued data can be defined as follows:

$$\text{mode}(X) = \text{mode}(H^q(X)) = k \times q - \frac{q}{2}$$

$$\text{where } k = \underset{i \in [1, m_q]}{\operatorname{argmax}} H_i^q(x) \quad (4)$$

Consider the example in Figure 2 where  $X = \{0.12, 0.51, 0.95, 2.45, 2.51, 3.12, 3.52, 3.89, 4.01, 4.49, 5.49\}$  and

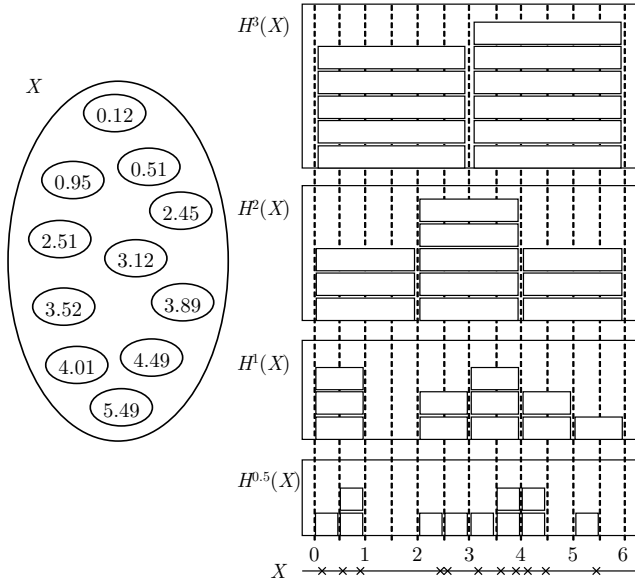


Figure 2: Different Mode Values for Different Quantization Values.

$V(X) = [0, 6]$ . There are different mode values for different quantization values;

$$\begin{aligned} \text{mode}(H^{0.5}(X)) &= 0.75, \text{mode}(H^1(X)) = 0.5, \\ \text{mode}(H^2(X)) &= 3, \text{and } \text{mode}(H^3(X)) = 4.5. \end{aligned}$$

The mode definition for the real valued data in eqn (4) also suffers the multimodal case problem;

$$\begin{aligned} \text{mode}(H^{0.5}(X)) &= \{0.75, 3.75, 4.25\} \\ \text{mode}(H^1(X)) &= \{0.5, 3.5\} \end{aligned}$$

### Density Based Mode

To deal with the aforementioned setbacks, we propose a modified version of mode which is the density-based mode within a radius.

First, the user defined radius,  $r$ , must be given. One possible radius can be given as follows.

$$r = \frac{e - s}{20} \quad (5)$$

Next, the density of an element,  $x \in X$  is defined as follows.

$$\begin{aligned} \text{dens}(x, r) &= \sum_{y \in X} N(x, y, r) \\ \text{where } N(x, y, r) &= \begin{cases} 1 & \text{if } d(x, y) \leq r \\ 0 & \text{otherwise} \end{cases} \quad (6) \end{aligned}$$

Euclidean distance can be used in  $d(x, y)$ .

A pseudo code to compute the proposed density based mode within radius is stated in Algorithm 1.

### Algorithm 1 Density based mode within radius

$X$  and the user defined radius increment factor,  $\alpha$  are defined globally and it is called with  $\text{rmode}(X, \alpha)$  initially.

```

rmode( $C, r$ )
   $C' = \text{argmax}_{x \in C} \text{dens}(x, r)$       1
  if  $|C'| = 1$ ,                          2
    return  $x \in C'$                     3
  else                                    4
    return rmode( $C', r + \alpha$ )      5

```

Line 1 computes the density of elements in the candidate set only; it counts how many samples are within the radius  $r$ . Note that the argmax function returns a set not a scalar value. If there is a unique element with the highest density, it is returned in line 2 and 3. Otherwise, it increases the radius  $r$  and reevaluates the tied candidates, recursively until it becomes unique.

Recall the example with multimodal case in Figure 1. There are two values with highest density:  $\{5, 8\}$ . Hence, the radius can be increased to 1 and they are reevaluated. Since  $\text{dens}(5, 1) = 6$  and  $\text{dens}(8, 1) = 9$ ,  $\text{mode}(X) = 8$  in the proposed mode definition as depicted in Figure 3 (a).

Figure 3 (b) illustrates resolving the problems in the example in Figure 2. When the radius is  $r = 0.45$ , there are three candidates with the same density. After increasing the radius, the candidate set reduces to two in size. Eventually, only one is left to be the mode.

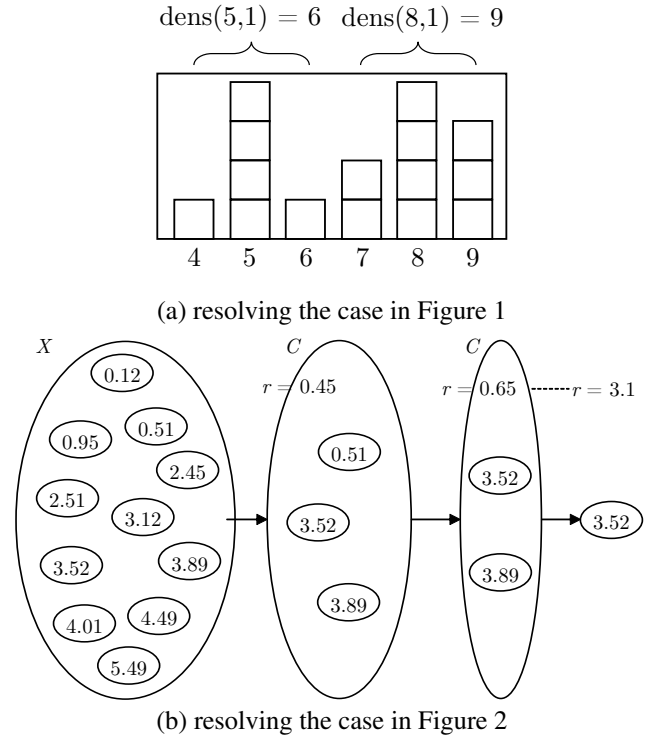


Figure 3: Density based Mode Illustration.

## Statistical Feature Extraction on EEG Signals

This section emphasizes on the various methods of extracting features according to a standard EEG database. Using Electroencephalography (EEG) brain signals has recently received increasing attention for the analysis of human affective states (Jahankhani, Kodogiannis, and Revett 2006). Different from the traditional applications of emotion recognition, studied by Zeng et al (Zeng et al. 2009) such as image, text, facial expressions or gestures, human brain signals have high abnormality natures as they are intrinsic and not easy to forge. Since brain signals unlike conventional approaches are impossible to be suppressed, EEG signals have become a possible approach to identifying human emotions. This research investigates the effectiveness of combining various statistical parameters and tests its effectiveness to identify human emotions through the EEG signal patterns. Two-fold steps of feature extraction are performed on the EEG signals: a feature set used in EEG emotion classification extracted from 40 frequency using Short-time Fourier Transform and statistical parameters are combined with the proposed mode to define a feature vector.

### DEAP dataset

Preprocessed EEG signals were gathered from a public database DEAP (see (Koelstra et al. 2012) for details), which recorded observation from 32 subjects who each watched 40 one-minute videos. The selected videos were of different emotional contents, media trials for the purpose of deriving the active emotional state of brain signals. During the watching procedures, 40 valued channels of physiological signals (32 sensors of EEG signals, and 8 peripheral physiological signal sensors) were gathered. For this research, 32 EEG channels (Fp1, AF3, F3, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, F4, F8, FC6, FC2, C4, T8, CP6, CP2, P4, P8, PO4, O2) were studied, according to the 10-20 international system. The placement of the EEG electrode positions is illustrated in Figure 4.

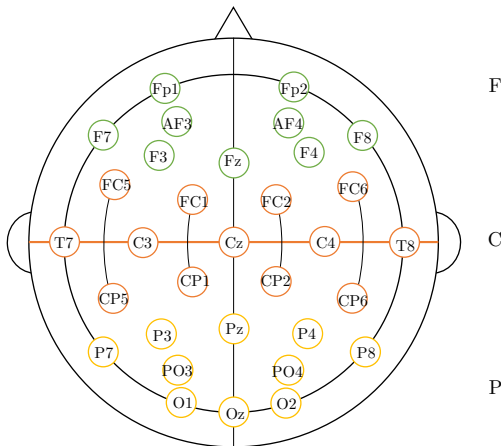


Figure 4: 32 channels (positions placements) according to the 10 - 20 system. Different colors represent the three different cerebral zones: frontal (F), central (C), and parieto-occipital(P)

The signal data was successively down sampled to 128 Hz after applying a bandpass frequency filter to restrict the available frequency range from 4 Hz to 45 Hz. In the DEAP database, each single EEG channel contains 8064 physiological or EEG signal segments; totaling to 40,960 subject samples and each sample is represented by 8064 dimensions vector. Five main frequency rhythms can be detected from EEG recordings: Delta (0.5-4Hz), Theta (4-8Hz), Alpha (8-14Hz), Beta (14- 30Hz), and Gamma (over 30Hz).

### Statistical features applied on the channels of EEG

In (Li, Cha, and Tappert 2018; Picard, Vyzas, and Healey 2001), six statistical features in equations (7 ~ 12) were selected to extend a range of standard statistical measurements. Following statistical features of each channel are computed where the  $V$  denotes to the value of the  $k$ th segment of the EEG raw data, the  $i = 1 \sim n$ , and  $n = 8064$  for the DEAP dataset. Let  $\bar{V}$  be the normalized value of the raw signals.

1. Mean

$$\mu(V) = \frac{1}{n} \sum_{i=1}^n V_i \quad (7)$$

2. Standard Deviation

$$\sigma(V) = \sqrt{\frac{1}{n} \sum_{i=1}^n (V_i - \mu(V))^2} \quad (8)$$

3. Mean of absolute values of first difference

$$\mu(d_1(V)) = \frac{1}{n-1} \sum_{i=1}^{n-1} |V_{i+1} - V_i| \quad (9)$$

4. Mean of normalized absolute values of first difference

$$\mu(d_1(\bar{V})) = \frac{1}{n-1} \sum_{i=1}^{n-1} |\bar{V}_{i+1} - \bar{V}_i| \quad (10)$$

5. Mean of absolute values of second difference

$$\mu(d_2(V)) = \frac{1}{n-2} \sum_{i=1}^{n-2} |V_{i+2} - V_i| \quad (11)$$

6. Mean of normalized absolute values of second difference

$$\mu(d_2(\bar{V})) = \frac{1}{n-2} \sum_{i=1}^{n-2} |\bar{V}_{i+2} - \bar{V}_i| \quad (12)$$

The features were extracted and computed from each of the EEG raw signals. The corresponding feature vector is defined by  $FV_{stati} = [\mu(V), \sigma(V), \mu(d_1(V)), \mu(d_1(\bar{V})), \mu(d_2(V)), \mu(d_2(\bar{V}))]$ . For the purpose of this study, a particular frequency  $f$  is utilized, and the spectrogram is reduced to 10 bins in the time domain. For each electrode, a feature vector  $FV_{stft}$  consisting of 40 STFT values ranging from  $f = 4\text{Hz}$  to  $f = 45\text{Hz}$  is considered. The mean frequency for each bin of the spectrogram is computed.

In this paper, the 7th statistical parameter, which is the density based mode within radius computed by Algorithm 1, is added.

## Russells circumplex Model

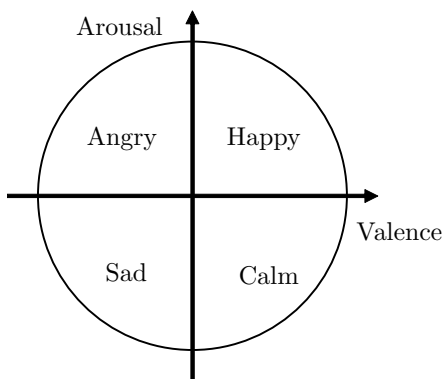


Figure 5: Russell’s Circumplex Model based on 2 dimensions: Valence and Arousal

Given the values in the dataset, Russells Circumplex model was used. The circumplex model of emotion which was developed by James Russell (Posner, Russell, and Peterson 2005), suggests that emotions are distributed in a two-dimensional circular space, containing arousal and valence dimensions. These labels were then extracted from the DEAP dataset as illustrated in Figure 5. 277 samples of low arousal and high valence are evaluated as calm. 272 samples of high arousal and low valence as angry. 187 samples of high arousal and high valence as happy. Finally, 285 samples of low arousal and low valence as sad.

## Experimental Results and Discussion

This section discusses and reports the experimental results of the analysis of emotion recognition based on EEG signals from DEAP database. The flow of the experiment is as shown in Figure 6.

### Experimental Setup

Three experimental trials were conducted to demonstrate the effectiveness of the proposed method. 3 CNNs with identical architectures were trained. For training and testing the model, dataset was split into 714 samples for training and 307 samples for testing. The first CNN was trained by using all 6 statistical parameters except Mode. This trial was conducted to see the nature of the traditionally used statistical parameters. In the second trial, the conventional formula of Mode available as functions on online platforms and coding environments was included as a 7th statistical parameter to observe its performance on the EEG dataset. The last and final trial was conducted using the Proposed Density based Mode as a final parameter along with the six other parameters.

### Experimental results and Discussion

Here we discuss the experimental results of the Density based mode within radius on EEG signals to recognize emotions. From the DEAP EEG signal database we extracted 32 trials of 32 participants. There were 1024 EEG signal

samples with 32 channels and 7 features extracted from 7 statistical parameters in a single experiment. The labels of 1024 EEG signal samples were distributed among 4 classes (Calm, Happy, Sad, Angry) using the Russell Circumplex model (Posner, Russell, and Peterson 2005). The data was then divided into testing, training and validation datasets.

To show the superiority of our density-based mode within radius we conducted three experiments each with four classes. Using the training dataset, we trained the 1-D Convolutional Neural Network and tested its accuracy which is shown in Table 1.

Table 1: Experimental Results

Experiment	Result
Accuracy without mode	78.00%
Accuracy with conventional mode	87.11%
Accuracy with proposed mode	89.78%

Table 1 shows three results, the first experiment was with 6 statistical features namely the mean, Standard deviation, mean of absolute values of first differences, mean of normalized absolute values of first difference, mean of absolute values of second differences and mean of normalized absolute values of second differences which showed an accuracy of 78.00%. The second experiment was conducted with 7 statistical features including the conventional mode which gave an accuracy of 87.11% showing that mode is an essential feature and can be considered for classification problems. The last experiment which showed the highest accuracy was done with 7 statistical features including our proposed mode which gave an accuracy of 89.78%.

Overall, experiment 3 which included our proposed mode as the 7th statistical parameter outperforms the other two models.

## Conclusion

Exploring new features to represent patterns is important in pattern recognition and machine learning. Statistical parameters have been widely utilized as features to represent patterns. This paper identified some weakness of a mode, which is pervasively used in statistics and pattern recognition and suggested an improved definition of mode using the density within radius concept. The primary purpose of this research was to explore how well density-based mode performs as a parameter on the statistical feature vector to recognize patterns in machine learning. The EEG database was used to show experimental superiority of the proposed method.

There are many methods in computer science especially using artificial intelligence for emotion detection and classification process, but still, research in this field is lacking. Even popular convolutional neural networks did not provide a good result. However, an ensemble of relevant and inferential statistical parameters on the feature vector is calculated and good performance was observed.

The highest performance achieved using the proposed density-based mode is 89.78% which seems to be low. It should be noted that detecting emotions using machine

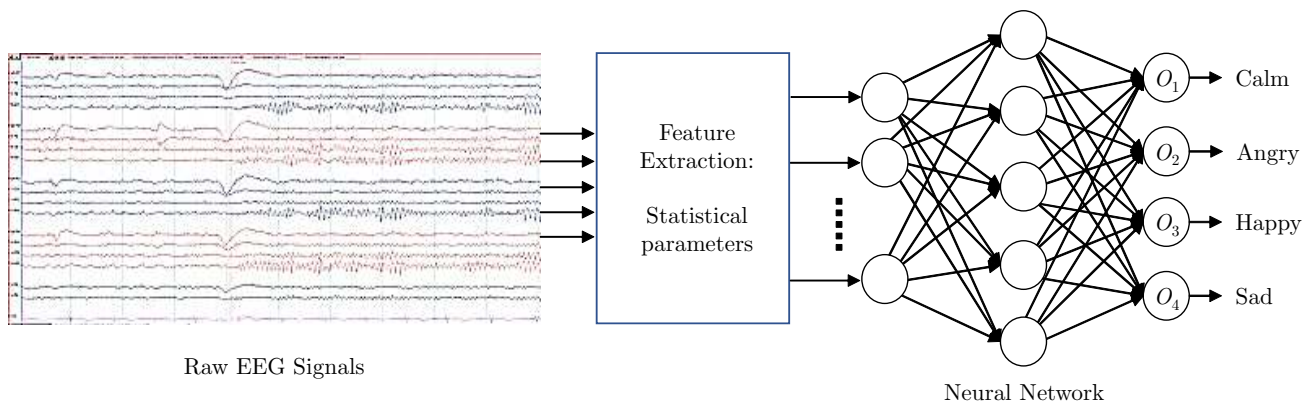


Figure 6: The flow of the experiment

learning techniques is often precision-sensitive process. Very accurate emotion recognition in applications such as social security, human driving health care especially mental health monitoring is important. The proposed system is prompt and economic as it utilizes brain signal information.

There are several future works and open problems. The first one is exploring other datasets to support the superiority of the proposed statistical parameter. The second one is to explore various other statistical parameters.

## References

- B.Esmael; Arnaut, A.; Fruhwirth, R.; and Thonhauser, G. 2012. A statistical feature-based approach for operations recognition in drilling time series. In *CISIM 2012*.
- Islam, M.; Ahmed, T.; Mostafa, S. S.; Yusuf, M. S. U.; and Ahmad, M. 2013. Human emotion recognition using frequency statistical measures of eeg signal. In *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, 1–6.
- Jahankhani, P.; Kodogiannis, V.; and Revett, K. 2006. Eeg signal classification using wavelet feature extraction and neural networks. In *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*, 120–124.
- Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; and Patras, I. 2012. Deap: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing* 3(1):18–31.
- Lambrou, T.; Kudumakis, P.; Speller, R.; Sandler, M.; and Linney, A. 1998. Classification of audio signals using statistical features on time and wavelet transform domains. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, volume 6, 3621–3624 vol.6.
- Li, S.; Cha, S.; and Tappert, C. C. 2018. Biometric distinctiveness of brain signals based on eeg. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1–6.
- Picard, R. W.; Vyzas, E.; and Healey, J. 2001. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(10):1175–1191.
- Posner, J.; Russell, J.; and Peterson, B. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. In *Dev Psychopathol*, volume 17, 715–734.
- Soranamageswari, M., and Meena, C. 2010. Statistical feature extraction for classification of image spam using artificial neural networks. In *2nd International Conference on Machine Learning and Computing (ICMLC 2010)*, 101–105. Los Alamitos, CA, USA: IEEE Computer Society.
- Yu, J., and Zuo, M. 2015. A novel traffic flow detection method using multiple statistical parameters. In *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 51–54. Los Alamitos, CA, USA: IEEE Computer Society.
- Yuen, C. .; San, W. S.; Mohamed.Rizon; and Seong, T. 2009. Classification of human emotions from eeg signals using statistical features and neural network. In *International Journal of Integrated Engineering*.
- Zazzi, M.; Ulivi, G.; and Prosperi, M. 2007. Statistical comparison of machine learning techniques for treatment optimisation of drug-resistant hiv-1. In *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, 427–432. Los Alamitos, CA, USA: IEEE Computer Society.
- Zeng, Z.; Pantic, M.; Roisman, G. I.; and Huang, T. S. 2009. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(1):39–58.