Pattern Classification and Analysis of Memory Processing in Depression Using EEG Signals

Kin Ming Puk[1], Kellen C. Gandy[2], Shouyi Wang[1], and Heekyeong Park[2]

[1] Department of Industrial, Manufacturing, & Systems Engineering
[2] Department of Psychology
University of Texas at Arlington
701 S Nedderman Dr, Arlington, TX 76019
{kinming.puk@mavs.uta.edu,kellen.gandy@mavs.uta.edu,
shouyiw@uta.edu,hkpark@uta.edu}

Abstract. An automatic, electroencephalogram (EEG) based approach of diagnosing depression with regard to memory processing is presented. EEG signals are extracted from 15 depressed subjects and 12 normal subjects during experimental tasks of reorder and rehearsal [2]. After preprocessing noisy EEG signals, nine groups of mathematical features are extracted and classification with support vector machine (SVM) is conducted under a five-fold cross-validation, with accuracy of up to 70% - 100%. The contribution of this paper lies in the analysis and visualization of the difference between depressed and control subjects in EEG signals.

Keywords: depression, working memory, long-term memory, EEG, machine learning, signal processing, feature extraction, feature selection, classification, support vector machine

1 Introduction

Depression, according to [3], is a term referring to a disabling and prevalent psychiatric illness, major depressive disorder (also known as clinical depression). Depressed patients tend to feel sad and pessimistic for a long period of time, and they are likely associated with low self-esteem and tendency to commit suicide, among other negative symptoms. It has been reported that depressed patients suffer from poor concentration and memory. For the sake of evaluating the effectiveness of memory retrieval, working memory (WM) and long-term memory (LTM) are the primary research interest in this work [7]. Techniques of brain signal analysis of EEG with classification framework in data mining is adopted in this regard. Indeed computer-aided diagnosis using EEG signals is a popular field of research. Classification framework consists of EEG preprocessing, feature extraction, feature selection and classifier. Depending on the disorder/disease and framework, classification accuracy may vary from 80% [4] to 97% [6]. Interested readers are referred to [4], [5], [6], [11], [12] for more details.

2 Experimental Design and Data Acquisition

Participant Participants were recruited from the University of Texas at Arlington. Each participant completed a pre-screen questionnaire which included the Center of Epidemiological Studies Depression Scale (CES-D) for separating groups of individuals with high and low depressive symptomatology. Individuals who scored 25 or above qualified to be part of the high group, below 15 to be part of the low group, and 15-25 to be part of the moderate group. A total of 60 individuals - 20 with low depression, 20 with moderate depression, and 20 with high depression - were recruited for the purposes of this experiment. In this EEG analysis, only data from 15 high depression individuals (4 males, 11 females; Age: 20.3 ± 3.21) and low depression individuals (4 males, 8 females; Age: 20.5 ± 2.66) is used due to cleanliness of data and for the sake of binary classification (so that data from moderately depressed subjects is not used).

Procedure Tasks of varying cognitive difficulty in semantic processing are designed with reference to [2]. During the entire experimental procedure, EEG signal is measured using the Brain Vision 32 channel system and recorded using the Pycorder software.

The assessment of working memory ("WM procedure") consists of "reorder" and rehearsal". For reorder tasks, participants are instructed to mentally reorder the sequence of three pictorial items based on their physical weight in an arrangement from lightest to heaviest, with 1 representing the lightest item, 2 representing the mid-weight item, and 3 representing the heaviest item ("reorder"). For rehearsal tasks, participants were instructed to remember the sequence of three pictorial items based on their serial order from top to bottom, with 1 representing the top item, 2 representing the middle item, and 3 representing the bottom item ("rehearsal").

On the other hand, long-term memory is assessed by differentiating "new" from old images ("LTM procedure"). After the WM procedure, participants are asked to continue with LTM procedure by indicating whether each image appears using their right hand ("new"), followed by a confidence rating of that decision which ranged from 1 to 3, with 1 representing low confidence, 2 representing medium confidence, and 3 representing high confidence using their left hand.

3 Extraction and Classification of EEG Features

Preprocessing of EEG Signals EEG data of 32 electrodes (FP1, FP2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T9, T7, C3, Cz, C4, T8, T10, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, Oz, Oz, O2; see Fig. 2) are imported into Matlab with software package "EEGLAB" [8]. EEG signals will then be re-referenced at channels T9 and T10 since these two channels are least influenced by cognitive processing, resampled from 1000 Hz to 256 Hz for reducing data size, and bandpass filtered at 1-35 Hz for removing unnecessary signal noise.
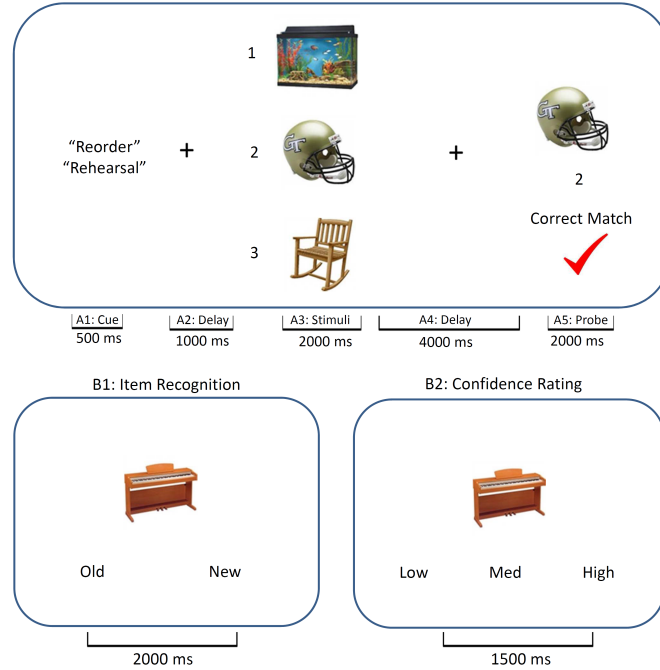

Fig. 1: i) WM, Top: On rehearsal trials, participants are instructed to maintain the serial order of the three presented items (top to bottom), whereas on reorder trials, participants were instructed to mentally rearrange the items according to their physical weight (lightest to heaviest). ii) LTM, Bottom: Participants are instructed to recognize whether the image appeared in WM procedure or not.

Epoching After preprocessing, EEG signal is partitioned into different epochs according to the experiment procedure for working memory and long-term memory. Each participant were shown 504 images, which correspond to 504 trials.

For working memory (Fig. 1), each trial (either reorder or rehearsal) consists of a cue on the center of the screen (A1, 500 ms), inter-stimulus interval (A2, 1000 ms), showing of the stimuli (A3, 2000 ms), delay (A4, 4000 ms) and probing for the answer (A5, 2000 ms). Therefore, EEG

signal of 336 trials, with each trial lasting for 10500 ms (including baseline of 1000 ms before the start of each trial) are extracted for working memory.

As for long-term memory (Fig. 1), the procedure consists of item recognition (B1, 2000 ms) and rating the confidence of the recognition (B2, 1500 ms). Therefore, EEG signal of 168 trials, with each trial lasting for 4500 ms (including baseline of 1000 ms before the start of each trial) are extracted for long-term memory.

Please note that baseline removal can only be done after epoching with the availability of baseline signal.

Artifact Removal Artifact is then removed from EEG signal with EEGLAB plugin - ADJUST (An Automatic EEG artifact Detector based on the Joint Use of Spatial and Temporal features) [10]. Artifact features including eye blinks, (vertical and horizontal) eye movements and generic discontinuities are accounted for. The four stages are i) Epoched EEG signal is first decomposed into different independent components using independent component analysis (ICA); ii) artifact features for each component are computed; iii) the value of each artifact feature for each component is to be checked against threshold value computed by Expectation-Maximization [23] in order to determine whether that component is an artifact; and iv) EEG signal is reconstructed using independent components which are not rejected.

Feature Extraction Nine groups of mathematical features - statistical features, time-frequency features, signal power, Hjorth parameters, Hurst exponent, band power asymmetry and spectral edge frequency - are extracted from each trial of subjects as in [4], [6] and [11]. These nine groups of features are extracted from EEG signal at each of the 30 channels (2 channels removed after rereferencing). They are then concatenated as a feature vector. This procedure is applied to all trials of EEG data for all participants. In the following, $X = \{x_1, x_2, ..., x_m\}$ denote a single-channel signal with m time points.

1) Statistical Features: Mean, variance, skewness, kurtosis at theta, alpha, beta and low gamma bands are computed. More specifically, mean is the averaged signal amplitude, variance measures the signal variability to the mean, skewness quantifies the extent to which the distribution leans to one side of the mean, and kurtosis measures the 'peakedness' of the distribution.

2) Morphological features: Three morphological features at theta, alpha, beta and low gamma bands were extracted to describe morphological characteristics of a single-channel signal as in [24] and [25].

- Curve length is the sum of distances between any two pair of consecutive points. Intuition behind this feature is that curve length increase with the signal magnitude, frequency and amplitude variation. It is mathematically calculated as follows:

$$\frac{1}{m-1} \sum_{i=1}^{m-1} |x_{i+1} - x_i| \tag{1}$$

- Number of peaks measures the overall frequency of a signal. It is mathematically calculated as follows:

$$\frac{1}{2} \sum_{i=1}^{m-2} \max(0, sgn(x_{i+2} - x_{i+1}) - sgn(x_{i+1} - x_i)) \tag{2}$$

- Average nonlinear energy, according to [26], is sensitive to spectral changes and is calculated as:

$$\frac{1}{m-2} \sum_{i=2}^{m-1} x_i^2 - x_{i-1}x_{i+1} \tag{3}$$

3) Time-Frequency Features: Wavelet transform (WT) is a powerful tool to perform time-frequency analysis of signals. The fundamental idea of WT is to represent a signal by a linear combination of a set of functions obtained by shifting or dilating a particular function called mother wavelet [13]. The WT of a signal $X(t)$ is defined as:

$$C(a,b) = \int^R X(t) \frac{1}{\sqrt{a}} \Psi(\frac{t-b}{a}) dt \tag{4}$$

where Ψ is the mother wavelet, $C(a,b)$ are the WT coefficients of the signal $X(t)$, a is the scale parameter, and b is the shifting parameter. Continuous wavelet transform (CWT) has $a \in R^+$ and $b \in R$ and discrete wavelet transform (DWT) has $a = 2^j$ and $b = k2^j$ for all $(j,k) \in Z$ given the decomposition level of j. Since CWT explores every possible scale a and shifting b, it is generally a lot more computationally expensive than DWT. As a result, DWT is often used to perform time-frequency analysis of a signal at different decomposition levels [14]. The DWT coefficients provide a non-redundant and highly efficient representation of a signal in both time and frequency domain. At each level of decomposition, DWT works as a set of bandpass filters to divide a signal into two bands called approximations and details signals. The details (D) are the high-frequency components. Among different wavelet families, we employed Daubechies wavelet as it is frequently used in physiological signal analysis due to its orthogonality property and efficient filter implementation [15]. A 4-level discrete wavelet transform (DWT) decomposition was applied to the collected signals with the sampling rate of 256 Hz. Table 1 lists the decomposed signals D1, D2, D3 and D4, which roughly corresponded to the commonly recognized brain signal frequency bands theta, alpha, beta, and gamma, respectively.

Decomposed Level	Frequency Range (Hz)	Approximate Band
D1	4-8	Theta
D2	8-12	Alpha
D3	12-25	Beta
D4	25-40	Gamma

Table 1: Frequency bands of signals by discrete wavelet decomposition.

After the four-level DWT decomposition, a set of wavelet coefficients can be obtained for each decomposed signals. To further decrease feature dimensionality, we employed a measure of wavelet coefficients called wavelet entropy (WE), which indicates the degree of multi-frequency signal order/disorder in the signals [27]. To obtain WE, the first step is to calculate relative wavelet energy for each decomposition level as follows:

$$p_j = \frac{E_j}{E_{total}} = \frac{E_j}{\sum_{j=1}^n E_j} \tag{5}$$

where j is the resolution level, and n is the number of decomposed signals ($n = 5$ in this study). E_j is the wavelet power, the sum of squared wavelet coefficients, of decomposed signal j. The relative wavelet energy p_j can be considered as the power density of the decomposed signal level j. Similar to Shannon entropy [28] for analyzing and comparing probability distributions, the WE is defined by

$$WE = -\sum_{j=1}^n p_j * ln(p_j)] \tag{6}$$

WE characterizes the order/disorder of signals powers in the five brain signal frequency bands (theta, alpha, beta and gamma) during the experiment.

4) Signal Power: Adopting the signal features used in a previous work [16], "band power" of EEG signals for each channel in commonly used frequency bands of brain signal including theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), and gamma bands (25-40 Hz) is computed. On the other hand, "relative band power" at each channel is computed as the ratio of the band power of the individual band over the sum of band power of all four bands.

5) Hjorth Parameters: Hjorth parameters, namely activity, mobility, and complexity, are frequently used in signal processing since its introduction by Bo Hjorth in 1970. These time-domain features are commonly used in brain signal analysis as in [6], [17].

6) Hurst Exponent: It is a statistical measure used to detect autocorrelation in time-series data such as EEG signal (usually notated as H) . If the value of H is 0.5, it indicates that the time-series data is a random series, whereas $H > 0.5$ indicates a trend reinforcing series [18].

7) Band Power Asymmetry: Asymmetry of power in theta, alpha and beta bands between different regions (inter-hemispheric) and within the same region (intra-hemispheric) of the brain, as in Fig. 2, are computed as features [21].
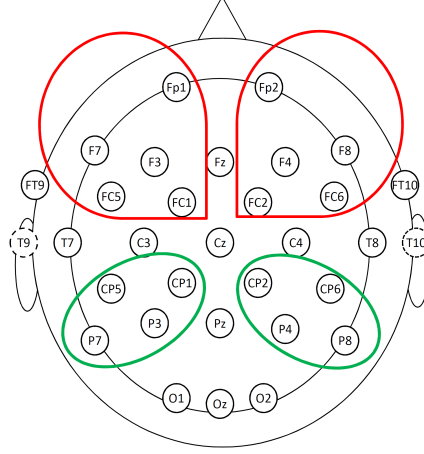

Fig. 2: Illustration of 4 groupings of channels (out of 32 according to 10-20 system) for inter- and intra-hemisphere band-power asymmetry. They correspond to left frontal, right frontal, left parietal and right parietal areas of the brain. T9 and T10 are not available after re-referencing.

8) Spectral Edge Frequency: It measures the frequency below which a certain percentage of total power of the EEG time-series signal [19]. In this project, percentage values of 50%, 90% and 95% are considered.

9) Zero Crossing: It is the number of points where the sign of the EEG signal changes from positive to negative (or vice versa).

Table 2 summarizes the features extracted in this work.

No	Group Name	Features (Generated at 30 channels and 4 bands except for groups 2-3, 5-9)	Count
1	Statistical	Mean, variance, skewness, kurotsis (at each channel)	120
2	Morphologi-cal	Curve length, number of peaks, average nonlinear energy (at each channel)	90
3	Time-frequency	Wavelet Entropy (power ratio of theta, alpha+low beta, beta and low gamma, alpha to beta ratio)	180
4	Signal Power	Band power and Relative band power	240
5	Hjorth	Activity, mobility, complexity (at each channel)	90
6	Hurst	- (at each channel)	30
7	Band Power Asymmetry	Asymmetry of inter-hemisphere and intra-hemisphere band power	68
8	Spectral Edge Frequency	Percentage values of 50%, 90% and 95% (at each channel)	90
9	Zero Crossing	- (at each channel)	30

Table 2: Summary of Groups of Mathematical Features Employed in this Classification Study

Feature Selection Feature selection method "minimal-redundancy-maximal-relevance criterion" (mRMR) [20] is used. mRMR aims at selecting a subset of feature set based on the statistical property of a target classification variable, subject to the constraint that the features are mutually dissimilar to each other but at the same time marginally similar to the target classification variable. Because of its first-order incremental nature, mRMR selects features very quickly without sacrificing classification performance. The number of features chosen in this classification study is 100.

Classifier Support vector machine with radial basis function (RBF) kernel from Matlab (name of the function is fitcsvm) is used.

In binary classification, SVM basically seeks a separating hyperplane which maximizes the distance between two classes of data points in order to differentiate data point of one class from another. To ease the use of kernel trick, the dual formulation (7) of SVM is usually considered, in which x is the feature vector (or data point in machine learning terminology) selected from the last step, and y is the class of that feature vector:

$$
\begin{aligned}
\underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\
\text{subject to} \quad & \sum_{i=1}^{n} \alpha_i y_i = 0, \ i = 1, \ldots, n \\
& C \geq \alpha_i \geq 0, \ i = 1, \ldots, n.
\end{aligned}
\tag{7}
$$

If $\sum_{i=1}^{l} \alpha y_i K(x_i^T x) + b \geq 0$ (where l is the number of features), the data point is classified to be depressed subject; otherwise, it is classified to be control subject. The RBF kernel on two samples x and y is given by:

$$
K(x, y) = exp(-\frac{||x - y||^2}{2\sigma^2})
\tag{8}
$$

Cross Validation Data will be divided into five (5) folds for cross validation (CV). In each fold of CV, 80% of data are used for training the classification model by tuning the hyper-parameters with grid search (namely C in equation (7) and σ in equation 8 for SVM), whereas the remaining 20% will be used for testing the trained model. Testing accuracy is the main measure of classification performance, which is calculated as the number of correctly predicted class of subjects in each trial (i.e. depressed or control) over the total number of trials available from all subjects in an epoch:

$$
\text{acc} = \frac{\text{no. of correctly predicted class of subjects from all trials}}{\text{no. of trials available from all subjects in an epoch}}
\tag{9}
$$

4 Experimental Result and Analysis

	WM					LTM	
	A1	A2	A3	A4	A5	B1	B2
Reorder	92%	100%	98%	86%	86%	73%	77%
Rehearsal	86%	91%	83%	86%	89%	86%	75%
New	-	-	-	-	-	57%	68%

Table 3: Classification Accuracy with SVM for Epoches in working memory (WM) and long-term memory (LTM). A means WM (A1: Cue, A2: Inter-stimulus interval, A3: Stimuli, A4: Delay, A5: Probe), whereas B stands for LTM (B1: Item Recognition, B2: Confidence Rating).

Table 3 shows the accuracy of classifying subjects with high depression from those with low depression under different experimental tasks and epoches. The higher the accuracy, the greater

the difference between the depressed and control subjects in performing experimental tasks. On the other hand, the epoch with high classification accuracy is investigated in order to better identify the difference between the two groups of subjects.

As mentioned before, each trial of WM procedure consists of a cue (A1), inter-stimulus interval (A2), showing of the stimuli (A3), delay (A4) and probing for the answer (A5). In these epoches, A3 is the time at which memory encoding takes place, and A4 is the time at which memory processing happens. As for long-term memory procedure, each trial consists of item recognition (B1) and rating the confidence of the recognition (B2). Therefore, B1 is the time at which retrieval of long-term memory takes place. Accuracy in "reorder" tasks is generally higher than those of "rehearsal" as a result of greater difficulty of cognitive processing required by "reorder" tasks [9]. Accuracy in "new" tasks is lowest among three kinds of tasks because of the same reasoning.

Epoches A3, A4 and B1 are worth investigating. It is because memory encoding and retrieval of working memory take place at epoches A3 and A4 respectively, whereas memory retrieval of long-term memory takes place at epoch B1. One way to investigate these epoches would be to consider the top 3 features (out of 100 selected by the mRMR algorithm) used for classification. There are 27 subjects, and therefore each feature (vector) under consideration consists of 27 values, with each one extracted from the EEG signal of one subject. Figures 3, 4 and 5 are the boxplots plotted with these 27 values of each feature.

An observation in this regard is that the higher the classification accuracy in that epoch, the farther the distance between the boxes of the depressed and control subjects. An example supporting this notion is epoch A3 for reorder tasks (figure 3), having accuracy of 98%. Its top feature is relative band power at alpha band (FC5) - the box of depressed does not overlap with that of the control completely.
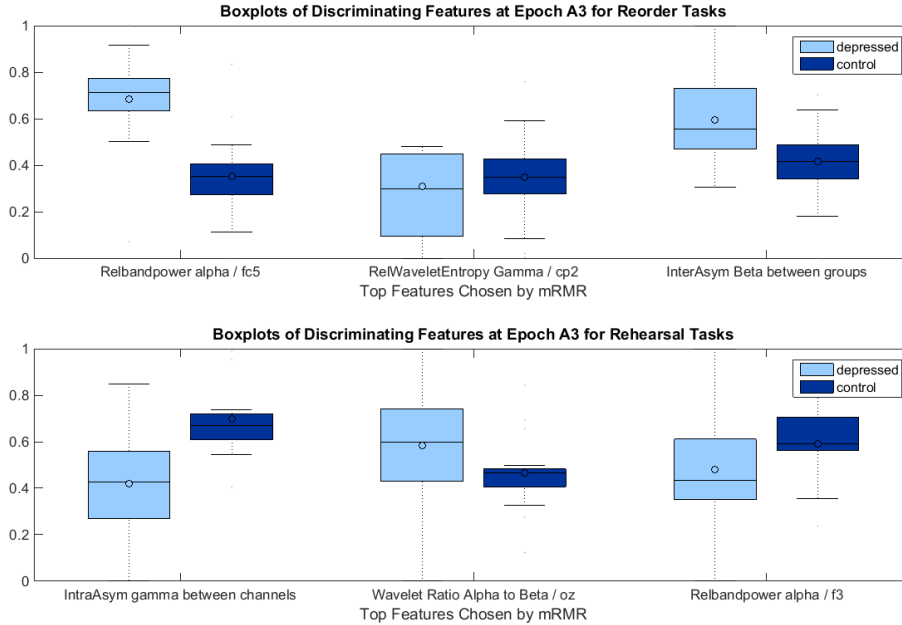

Fig. 3: Boxplot of Top 3 Features Used for Classification in Epoch A3. The top 3 features for reorder tasks are relative band power at beta (FC5), relative wavelet entropy of gamma band (CP2) and asymmetry inter-hemisphere band power at beta, whereas those for rehearsal tasks are asymmetry of intra-hemisphere band power at gamma, wavelet ratio of alpha to beta (Oz) and relative band power at alpha (F3).

In addition to the above, topographical plots (figure 6) at epoches A3, A4 and B1 for theta (4-8 Hz) and alpha (8-12 Hz) bands are plotted with average EEG signal of trials after preprocessing. With topographical plots, activation in different frequency bands can be considered over the 30 EEG channels.

In epoch A3 (both theta and alpha bands), left prefrontal area is observed to have greater activation for control over depressed subjects for reorder task. Surprisingly, no distinct difference
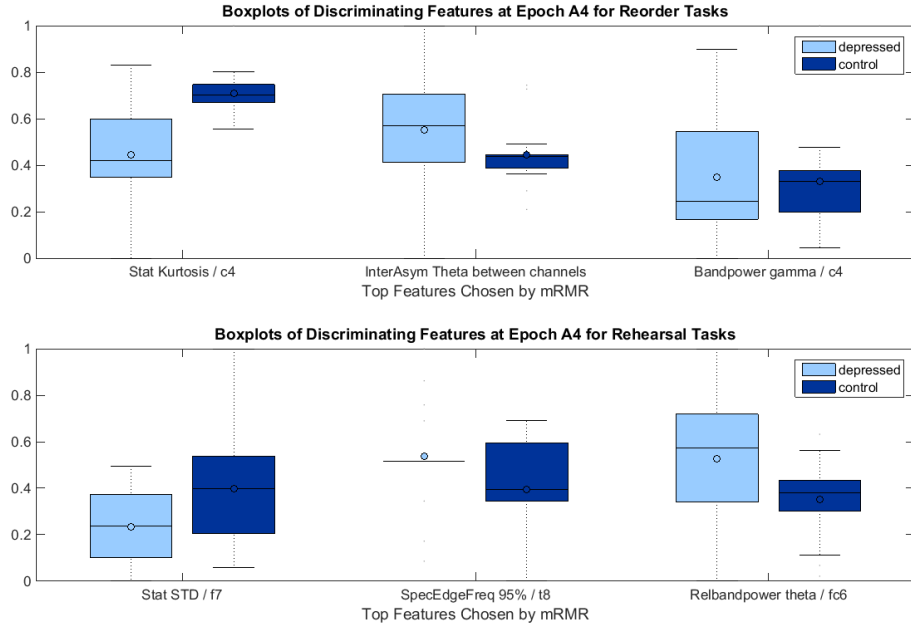
Fig. 4: Boxplot of Top 3 Features Used for Classification in Epoch A4. The top 3 features for reorder tasks are kurtosis (C4), asymmetry of inter-hemisphere band power at theta and band power at gamma (C4), whereas those for rehearsal tasks are standard deviation (F7), spectrum edge frequency at 95% (T8) and relative band power at theta (FC6).
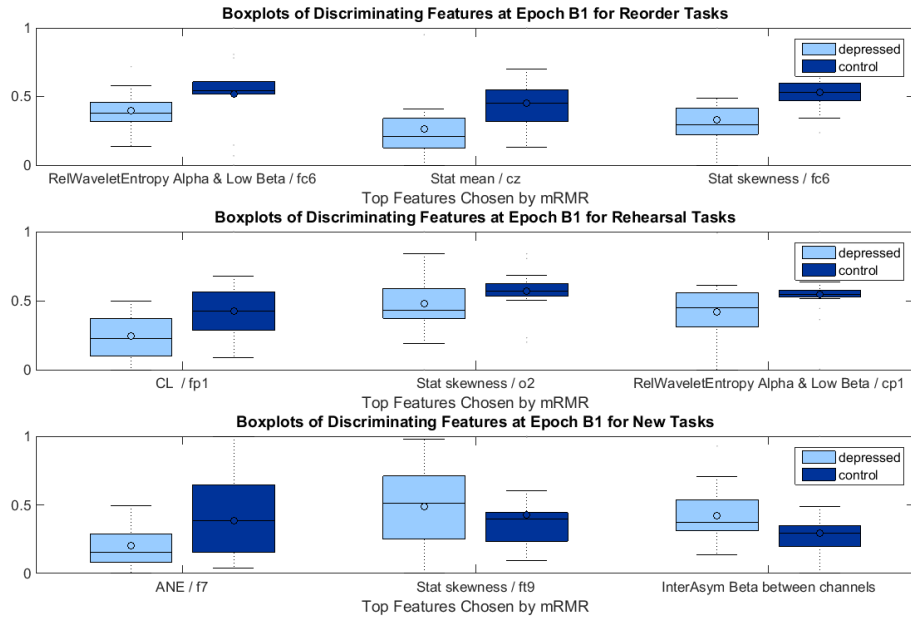

Fig. 5: Boxplot of Top 3 Features Used for Classification in Epoch B1. The top 3 features for reorder tasks are relative wavelet entropy at alpha and low beta (FC6), mean (Cz) and skewness (FC6), those for rehearsal tasks are curve length (Fp1), skewness (O2) and relative wavelet entropy at alpha and low beta (CP1), and those for new tasks are average non-linear energy (F7), skewness (FT9), and asymmetry of inter-hemisphere band power at beta.

between depressed and control subjects can be found in epoch A4. Last but not least, there is stronger activation found in occipital area for control over depressed subjects at epoch B1 for new task.
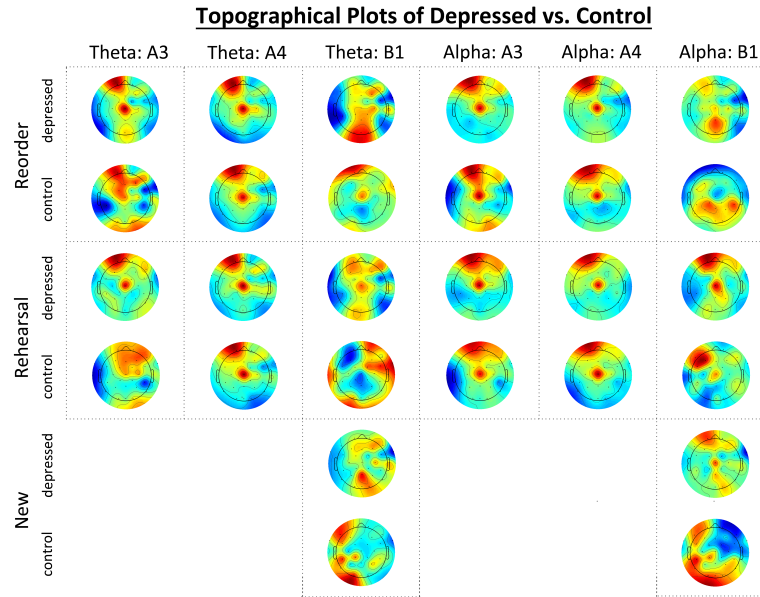

Fig. 6: Plots of relative topographical distribution of mean log power spectrum of theta and alpha bands at epoches A3 (showing of stimuli), A4 (Delay) and B1 (Item Recognition). These plots are generated with EEGLAB function "spectopo" and "topoplot".

5 Conclusion

This study investigated the difference of memory processing between depression and control groups using EEG signals. An extensive EEG feature study has been performed using the most popular techniques of feature extraction in the up-to-date literature. The popular technique of feature selection, minimal-redundancy-maximal-relevance criterion (mRMR), has been employed to identify the most discriminative EEG features among the two groups of subjects. Classification using support vector machine with RBF kernel showed that the depressed subjects indeed exhibited different patterns of brain activity in the processing of both working and long-term memory, with classification accuracies higher than 80%. The top EEG features showed significantly different distributions between two groups of subjects. This preliminary data-driven study indicates that depression can affect a subject's memory processing considerably. In the future work, more statistically valid neural signatures of depression with regard to its effect to memory processing will be investigated and explored using advance data mining and machine learning techniques. The long-term goal of this study is to facilitate the understanding of neural mechanism of depression, and to develop better data-driven tools for diagnosis and treatment of depression in clinical practice.

References

1. Burt, D., Zembar, M., Niederehe, G.: Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. Psychological Bulletin. 117, 285-305 (1995).
2. Ragland, J., Blumenfeld, R., Ramsay, I., Yonelinas, A., Yoon, J., Solomon, M., Carter, C., Ranganath, C.: Neural correlates of relational and item-specific encoding during working and long-term memory in schizophrenia. NeuroImage. 59, 1719-1726 (2012).
3. Loewenthal, K.: Depression. Encyclopedia of Psychology and Religion. 1-5 (2015).
4. Hosseinifard, B., Moradi, M., Rostami, R.: Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. Computer Methods and Programs in Biomedicine. 109, 339-345 (2013).

5. Li, Y., Fan, F.: Classification of Schizophrenia and Depression by EEG with ANNs. 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. (2005).
6. Yuan, Q., Zhou, W., Li, S., Cai, D.: Epileptic EEG classification based on extreme learning machine and nonlinear features. Epilepsy Research. 96, 29-38 (2011).
7. Baddeley, A., Eysenck, M., Anderson, M.: Memory. Psychology Press, Hove [England] (2009).
8. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods. 134, 9-21 (2004).
9. Gandy, Kellen C.: An EEG investigation of memory in depression: The effect of cognitive processing. Master Thesis, Department of Psychology, University of Texas at Arlington (2015).
10. Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M.: ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology. 48, 229-240 (2011).
11. Wang, S., Gwizdka, J., Chaovalitwongse, W.: Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task. IEEE Transactions on Human-Machine Systems. 1-12 (2015).
12. Wang, S., Lin, C., Wu, C., Chaovalitwongse, W.: Early Detection of Numerical Typing Errors Using Data Mining Techniques. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 41, 1199-1212 (2011).
13. Addison, P.: The illustrated wavelet transform handbook. Institute of Physics Pub., Bristol, UK (2002).
14. Rosso, O., Martin, M., Figliola, A., Keller, K., Plastino, A.: EEG analysis using wavelet-based information tools. Journal of Neuroscience Methods. 153, 163-182 (2006).
15. Subasi, A.: EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications. 32, 1084-1093 (2007).
16. Grimes, D., Tan, D., Hudson, S., Shenoy, P., Rao, R.: Feasibility and pragmatics of classifying working memory load with an electroencephalograph. Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08. (2008).
17. Oh, S., Lee, Y., Kim, H.: A Novel EEG Feature Extraction Method Using Hjorth Parameter. IJEEE. 106-110 (2014).
18. Qian, B., Rasheed, K.: Hurst Exponent and Financial Market Predictability. IASTED conference on Financial Engineering and Applications (2004).
19. Drummond, J., Brann, C., Perkins, D., Wolfe, D.: A comparison of median frequency, spectral edge frequency, a frequency band power ratio, total power, and dominance shift in the determination of depth of anesthesia. Acta Anaesthesiologica Scandinavica. 35, 693-699 (1991).
20. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence. 27, 1226-1238 (2005).
21. Yuvaraj, R., Murugappan, M., Mohamed Ibrahim, N., Iqbal, M., Sundaraj, K., Mohamad, K., Palaniappan, R., Mesquita, E., Satiyan, M.: On the analysis of EEG power, frequency and asymmetry in Parkinsons disease during emotion processing. Behavioral and Brain Functions. 10, 12 (2014).
22. Golinkoff, M., Sweeney, J.: Cognitive impairments in depression. Journal of Affective Disorders. 17, 105-112 (1989).
23. Bruzzone, L.Prieto, D.: Automatic analysis of the difference image for unsupervised change detection. IEEE Trans. Geosci. Remote Sensing. 38, 1171-1182 (2000).
24. Wang, S., Lin, C., Wu, C., Chaovalitwongse, W.: Early Detection of Numerical Typing Errors Using Data Mining Techniques. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 41, 1199-1212 (2011).
25. Wong, S., Baltuch, G., Jaggi, J., Danish, S.: Functional localization and visualization of the subthalamic nucleus from microelectrode recordings acquired during DBS surgery with unsupervised machine learning. J. Neural Eng. 6, 026006 (2009).
26. Kaiser, J.: On a simple algorithm to calculate the 'energy' of a signal. International Conference on Acoustics, Speech, and Signal Processing. 1, 381-384 (1990).
27. Rosso, O., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schrmann, M., Baar, E.: Wavelet entropy: a new tool for analysis of short duration brain electrical signals. Journal of Neuroscience Methods. 105, 65-75 (2001).
28. Shannon, C.: A Mathematical Theory of Communication. Bell System Technical Journal. 27, 379-423 (1948).