9th International IEEE EMBS Conference on Neural Engineering
San Francisco, CA, USA, March 20 - 23 , 2019

Decoding Natural Positive Emotional Behaviors from Human Fronto-Temporal Mesolimbic Structures

Maryam Bijanzadeh, Maansi Desai, Deanna L. Wallace, Nikhita Mummaneni, Nikhita Kunwar, Heather E. Dawes & Edward F. Chang

***Abstract*— Understanding the correlation between neural features and symptoms of mood disorders, such as depression, could provide objective measurements for diagnosis and facilitate clinical treatments. In this paper, we study the correlation of neural features with positive naturalistic emotional displays, e.g., smiling, in human subjects in a normal setup, without presenting any experimental stimuli to the subjects. We employed a data driven approach and utilized Random Forest classifiers to decode positive emotional displays from brain activity. Our results on all of our eight subjects show that neural features from mesolimbic circuits including cingulate, hippocampus, insula, amygdala and orbitofrontal cortex (OFC) can be used for decoding emotions (mean area under the ROC curve = 0.86 +- 0.04). The most important features based on the Random Forest models were mainly clustered in the gamma frequency band (30-100Hz) and low frequencies, with majority of them in theta band (4-8 Hz). These features were distributed across the limbic network, specific to each individual. Remarkably, the gamma cluster was selective to the positive emotions while the low frequency cluster showed selectivity to the neutral state. These results demonstrate that non-task-based emotions can be decoded from brain neuronal activity, and, may inform biomarker identification for objective symptom assessment in the treatment of severe mood disorders.**

I. Introduction

Identification of neural biomarkers correlated with behavioral symptoms of mood disorders, such as major depressive disorder, could provide objective metrics for diagnosis, risk-assessment and recovery tracking. In particular, changes in affective stages, e.g., an increase in presence of negative emotional display such as sadness or decreased presence of positive affective displays, such as laughter, are symptoms of depression [1]. Using behavioral paradigms such as memory recall [2] or image/video streams with specific valence [3], fMRI imaging [3] and EEG (Electroencephalography) [4] studies have reported that brain regions including limbic circuits, frontal cortex and temporal cortex are involved in affective displays of laughter and emotional states, such as happiness. These studies have provided fundamental understanding of neural mechanisms underlying emotion and its affective processing. However, they are all influenced by limitations: (1) the aforementioned paradigms afford experimental control, but they may engage neural circuits which are distinct from those underlying endogenous changes during "natural emotional displays", in which no task/stimulus would be delivered to elicit specific emotions, (2) available neural and behavioral datasets from each participant in both fMRI and EEG studies are limited (hour long sessions), (3) fMRI studies suffer from low temporal resolution (above 0.5 second), while EEG studies lack spatial resolution and do not have access to deep mesolimbic structures such as anterior cingulate region, which has been shown to be a prominent brain hub for inducing laughter [5,6].

Here we study the neural correlates of naturalistic emotional displays, in the absence of any experimental task. Multiple days of continuous intracranial electroencephalography (iEEG) signals were recorded from human subjects undergoing seizure localization. Positive

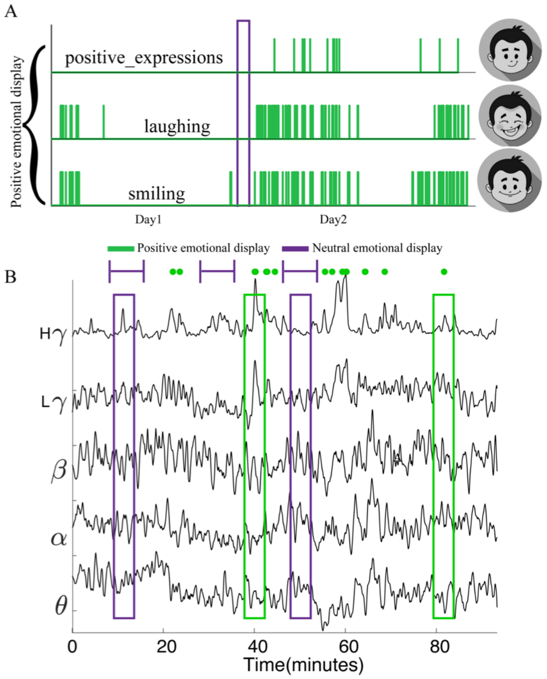

Figure 1. A) Annotated positive emotional displays for an example patient across two days. The green lines and purple boxes are instances of the positive behavior, and the neutral state, respectively. B) z-score of analytic amplitude within each frequency band for an example channel, which are all averaged in 10 sec non-overlapping bins. θ: theta, α: alpha, β: beta, $L\gamma$: low gamma, Hγ: high gamma

* This research was funded by the Defense Advanced Research Projects Agency (DARPA) under Cooperative Agreement Number W911NF-14-2-0043. The views, opinions, and/or findings contained in this material are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government

M.B., N.M, N.K, H.E.D. & E.F.C. are with the department of Neurological Surgery, University of California San Francisco, San Francisco, CA, 94143, USA (corresponding author e-mail: maryam.bijanzadeh@ucsf.edu).

M.D. is with the department of Communication Sciences and Disorders, University of Texas at Austin, Austin, TX 78712 USA.

D.L.W. and M.D. were with the department of Neurological Surgery, University of California San Francisco, San Francisco, CA, 94143, USA.

978-1-5386-7921-0/19/$31.00 ©2019 IEEE

emotional displays, such as smiling, laughing and positive expressions, along with other human natural behaviors, i.e. drinking, eating and etc., were hand annotated from 24-hour audio and video recordings of consented subjects. This unique clinical situation allowed us to evaluate whether there are neural features correlated with natural positive emotional behavior in humans. In particular, we used machine learning and data driven approaches to address two main questions: 1. Do neural features allow distinction between positive emotional displays and neutral state? 2. If yes, what are the neural correlates underlying such positive behaviors?

II. Methods

A. Neural Recordings

Electrophysiological data were collected by Natus EEG clinical recording system at sampling rates at either 512 Hz or 1024 Hz. Based on epileptic pathology and clinical needs, each subject (n=8, 4 females & 4 males, age: 20-36) had specific electrode coverage within the mesolimbic circuit, but some regions were common across all patients such as orbitofrontal cortex (OFC), cingulate and insula. All mesolimbic structures were sampled by either 4 contact strip or 4/10 contact depth electrodes. Electrode locations were validated by visual examinations (co-registered CT and MRI).

B. Behavioral Data Recordings and Annotations

During several days of hospitalization, 24-hour video and audio were recorded for the same human subjects. Then a set of human raters were asked to manually annotate the emotional displays using ELAN [7] software by putting a single mark at each time stamp in the recordings. To increase reliability and precision of annotations, video labels were verified by two other annotators for each subject. The labels include both emotions, e.g., smiling, laughing, positive, and pain expressions, and other natural activities such as drinking, eating and etc. In this paper, we focus on instances of smiling, laughing and positive expressions that were all grouped under positive emotional displays (Figure 1-A). Neutral display is defined from 10-minute long periods of annotated data where there are neither positive nor negative emotional displays. Emotional displays were later aligned with the neural recordings and formed a binary time-domain trace (Figure 1-A).

Comorbid depression and anxiety disorders were quantified by Beck Depression Inventory (BDI) and Beck Anxiety Inventory scores prior to surgery. Both scores ranged between minimal to moderate, 4-28 and 5-28, respectively.

C. iEEG Preprocessing

Raw iEEG recordings that were time aligned with the positive emotional displays, were demeaned, notch filtered (2nd order butterworth filter) at 60 Hz and its harmonics, and decimated (zero-phase 30th order FIR filter) to 512 Hz. Then the preprocessed signals were visualized to remove noisy electrodes and mark time epochs when there was motion or interictal artifacts [8]. After excluding noisy channels, common average referencing was performed on the electrodes that were on the same depth/strip lead.

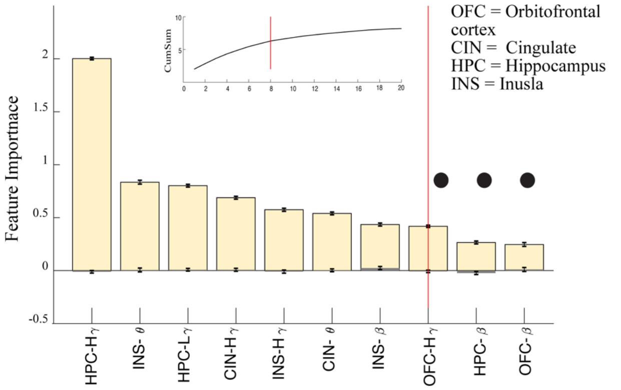

Figure 2. Example of feature selection procedure for Subject 2. Red lines on both inset and the main figure are the knee point of cumulative summation curve(inset). Regions' nomenclatures are as in the legend. θ: theta, α: alpha, β: beta, $L\gamma$: low gamma, Hγ: high gamma

D. Feature Extraction

To extract neural features, we applied the Hilbert transform on band pass filtered signals in five frequency bands: 4-8 Hz (theta), 8-12 Hz (alpha), 12-30 Hz (beta), 30-55 Hz (low gamma), and 70-100Hz (high gamma), by 4th order butterworth filter. Using 1 second non-overlapping bin, we averaged the analytic amplitude within each frequency band. The resulting signals were z-scored within each frequency band for each electrode and were averaged by 10 time points (10 seconds) centered on the occurrence of each emotional display, i.e. positive or neutral (Figure 1-B). Finally, the z-scored analytic amplitudes from all electrodes on the same lead were averaged. Thus, the resulting input for the decoder has dimension of number of regions times frequency for each label (i.e. positive emotional displays). The 1-second moving average was also applied on binary time-domain trace of the emotional displays.

E. Classification

As mentioned in section B, instances of neutral display were extracted from 10-minute long periods of data during which there were no annotated emotions. To make unbiased labels for neutral state, they were chosen from different periods of annotations. We also maximized the number of neutral displays satisfying aforementioned criteria and randomly sampled the same number of labels as in the positive emotion group to make a balanced dataset. We applied this procedure 5 times for each subject to make 5 datasets.

We trained a Random Forest classifier [9] using k-fold cross validation for each subject. The folds were selected such that the training and test divisions do not share adjacent samples in time to avoid any information overlap between training and testing. K (5 or 10) was picked such that we get at least 10 samples in each fold. Number of samples varied between 42 to 164 samples within each class (e.g. positive or neutral emotional display) across subjects. The Random Forest classifiers were trained with 300 trees and were optimized for two hyper parameters: (1) each tree was grown such that the maximum number of samples per leaf was varied in the range of 1 and 20, (2) number of features at each node varied in the range of 1 to maximum number of features minus 1.

F. *Feature Selection*

Random Forest models give the relative importance of features as an output. In the case of classification, it is defined as the mean prediction error for each sample, such that those decision trees including that sample in the aggregated boot strapping, will be removed and the error is computed by the remaining trees. Using this approach the relative importance of tree nodes, i.e. features, is obtained [10]. We refer to the model prediction error for each feature as feature importance (FI).

Subsequently, we ranked the FI and found the knee point of its cumulative summation curve for each subject using an algorithm called "kneedle". This method estimates the knee point based on maximum curvature for a discrete set of points [11]. Those features up to the knee point (Figure 2, red vertical lines) were selected as the important features.

G. *Statistical Analyses*

To assure that the results of main models are significantly above chance (50%), permuted Random Forest models were trained in the same way as explained in Section E, using the shuffled labels within each fold (to keep the balance between positive and neutral labels). Non-parametric Wilcoxon ranksum test was used for decoder statistical tests. To test the separation between features across the two classes, we used student t-test to obtain T-value and the p-value. All analyses were programmed in MATLAB.

III. Results

A. *Decoder Performance*

Random Forest models that were trained and optimized by cross validation methods were able to significantly differentiate between the two classes across all subjects. The area under the receiver operating characteristic (ROC) [12] curve across seven subjects was in the range of 0.82 – 0.97 and for one subject was 0.58 (mean±sem = 0.86 ± 0.04). The median area under the ROC curve (AUC) across all subjects was 0.89. Figure 3-A represents the ROC curve for an example subject, which was averaged across 10 folds of 5 data sets (shading indicates sem). For all subjects, the AUC of the main model was significantly larger than the permuted model (Figure 3-B; mean +- sem = 0.495 +-0.005, green vs black boxplots), using Wilcoxon ranksum test (p<0.001). Furthermore, the mean accuracies across all subjects for the main and permuted models were 0.79 (sem = 0.03) and 0.5 (sem = 0.0048), respectively. These results show that the positive emotional behavior could be distinguished from the neutral state using neuronal features across mesolimbic circuits.

B. *Neural Features Distinguishing Positive Emotional Displays from Neutral State*

As mentioned above (Section F, methods), for each subject, we utilized Random Forest prediction error to extract important features from the main models (Figure 2). In sum, the cumulative sum of the ranked FI (Figure 2-inset), was computed and the top ranked features were selected by using an objective threshold, the knee point. In addition, FI was extracted for permuted models and as shown in Figure 2 (gray bars), these FIs were significantly smaller than the main model. To evaluate whether there is collinearity between selected features, the correlation matrix was computed across labels. Observing collinearity between features (Figure 4-A), we performed hierarchical clustering to objectively group the correlation matrix that resulted into two main groups: 'gamma cluster' (30-55 and 70-100 Hz) and 'low frequency cluster' including theta (4-8 Hz), alpha (8-12 Hz) and beta (12-30 Hz) across subjects. Within each subject there were specificities in both selected mesolimbic regions and frequency band. Table 1 shows the summary of selected features for each subject as well as the limbic coverage used in the feature space. Clusters were named based on the majority of frequency band within each. Pooling clusters across all patients, low and high gamma frequencies formed 80% of the features within the gamma cluster, while just 15% of the low-frequency cluster was composed of gamma range.

C. *Selectivity of Gamma and Low Frequency Clusters*

In the feature domain, gamma and low frequency clusters usually had opposite selectivity for emotional states (Figure 4-B). Specifically, gamma distribution had larger values for positive expressions than the neutral state, while inverse direction was observed in lower frequencies, i.e. theta band.

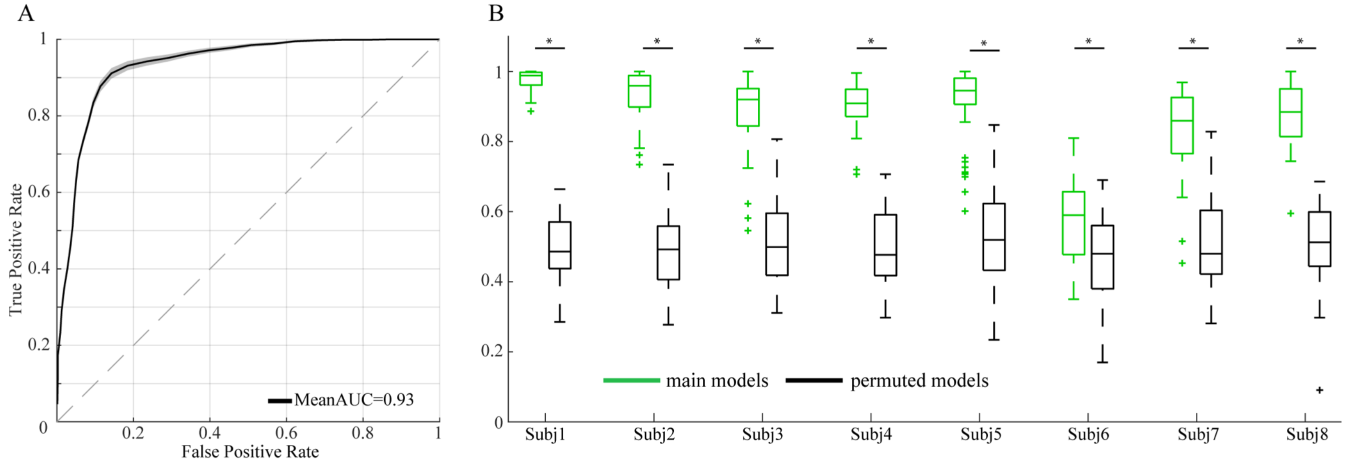

Figure 3. A) Decoder ROC curve for an example patient (Subject 2), B) population results of the area under the ROC curve for each subject. Green and black boxplots show main and permuted models, respectively. Wilcoxon ranksum test is used for statistics here.

This effect can be observed in the cingulate region of the example patient in Figure 4B: left panel shows that the high gamma distribution within cingulate is significantly different between positive and neutral state (t (657) = 16.34, p<0.0001, median of positive – median of neutral = 0.66). While, the theta distribution within cingulate has larger median value for neutral state compared to the positive emotional display (median of positive – median of neutral = - 0.43, t (657) = -10.77, p<0.0001).

The specificity of frequency band to the emotional state, was commonly seen across subjects with high decoder accuracy. Specifically, low- and high- gamma bands that were assigned to the gamma cluster are selective to the positive emotional state in 6/8 patients. While, the low-frequency cluster, including theta, alpha and beta, were generally selective to the neutral state in the same subject pool. In one subject, all frequency bands were selective in both clusters and for one other subject, who was holding the lowest decoder performance (Subject 6), mixed effect was observed in both clusters. These results suggest that the gamma frequency band can serve as a biomarker of positive emotional display within specific mesolimbic regions in each subject.

IV. Discussion

Utilizing machine learning methods along with data driven approaches, we were able to decode positive emotional expressions in human subjects from a unique behavioral and neural dataset. This dataset not only provided continuous multi-day neural recordings from both deep limbic structures, i.e. amygdala, hippocampus and insula, as well as OFC and cingulate cortex, it also contained rich behavioral data that were hand annotated. The aforementioned properties, are not feasible in most imaging and EEG setups, studying mood and emotion related tasks.

First, Random Forest models, reached promising performance, e.g. mean AUC = 0.86, distinguishing positive emotional displays from neutral state. This high-performance result, suggests that model inputs, i.e. the neural features from mesolimbic structures, are selective to the positive emotional expressions. Specifically, analytic amplitudes within traditional EEG frequency bands, including: 4-8 Hz (theta), 8-12 Hz (alpha), 12-30 Hz (beta), 30-55 Hz (low gamma), and 70-100Hz (high gamma), were used across available mesolimbic structures within each subject.

Second, we asked which neural features contributed the most to the decoder. Ranking Random Forest prediction error output, and assigning a threshold to its cumulative summation curve, allowed us to find subject-specific neural features that were selective to positive expressions (Table 1). A recent study by Sani et al. [13], reported that dynamical decoding approaches can be used to identify specific mood biomarkers for each subject. In both this paper and the results presented here, subjects were patients with epilepsy under clinical monitoring for seizure activity, and electrode coverage was based on the clinical needs. The heterogeneous coverage motivated us to search for personalized emotion biomarkers within each subject. Furthermore, if each node of the mesolimbic network has a specific function, then based on experience and brain plasticity each subject might have a dominant node or pair of nodes that are active more than other regions encoding emotions.

Remarkably, selected features were divided into two main clusters including gamma cluster and low frequency cluster which contained theta, alpha and beta, across selected limbic regions (Figure 4 and Table1). Pooling all selected features across subjects, gamma and theta frequency were the most repeatable biomarkers in each cluster. In addition, the gamma frequency band was selective to the positive emotional displays compared to the neutral state within majority of subjects.

Consistent with the literature [1], we have observed negative correlations between number of positive emotional displays and BDI score (n=17, r=-0.59, p = 0.012, Spearman correlation). Specifically, those patients expressing more positive behaviors had lower BDI scores. Thus, monitoring emotional expressions and their correlated neural features, can be employed to objectively assess symptoms of mood

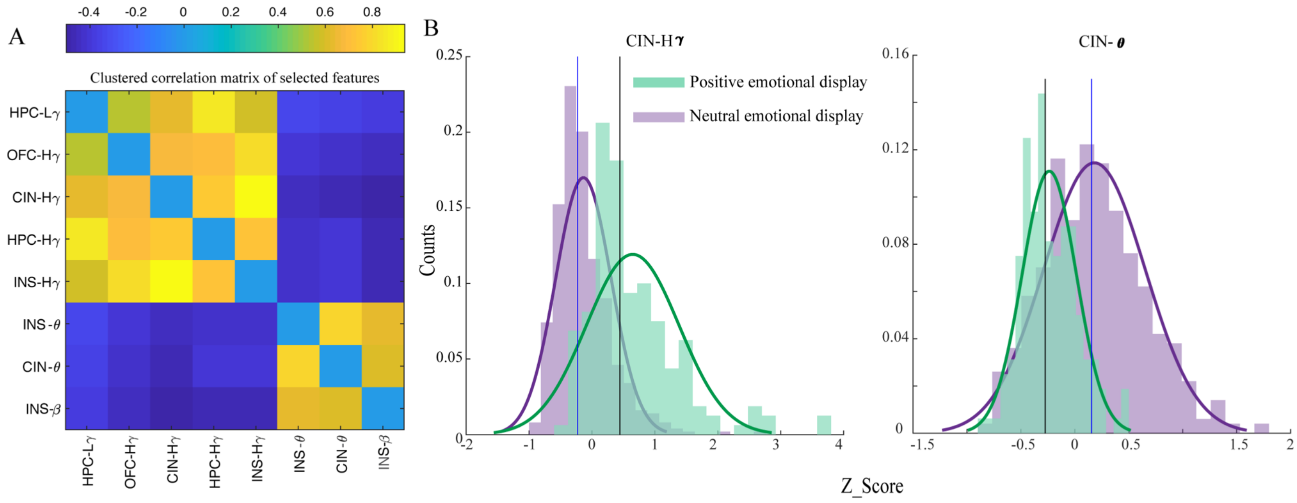

Figure 4. A) Clustered correlation matrix of the important features for subject 2. B) shows distribution of two features (cingulate-theta and cingulate-high gamma) for the positive emotional display and the neutral state in green and purple, respectively. Black and blue vertical lines are distribution's median.

disorders.

Furthermore, these results would introduce a beneficial framework for future studies to utilize experimental paradigms with naturalistic setups to further understand dynamics of the neural biomarkers underlying emotions. These biomarkers can be targeted by therapeutic interventions, including neurofeedback and closed-loop electrical stimulation, as a method to shift to positive states.

V. Conclusion

Undertaking machine learning approaches we have identified personalized biomarkers distinguishing positive emotional displays in human subjects, which could serve as an objective mood tracker in the treatment of severe mood disorders.

Acknowledgment

The authors would like to acknowledge Alia Shafi, Ben Speidel, Dharshan Chandramohan, Kristin Sellers, Lowry Kirkby, Ankit Khambhati, Omid Sani and behavioral data annotators: Nicole Goldberg-Boltz, Lucia Bederson, Muriel Solberg, Claire Eun, Joe Gordon, Dale Tager, Vivian Cheng.

References

[1] E. E. Forbes and R. E. Dahl, “Neural systems of positive affect: relevance to understanding child and adolescent depression?,” *Dev Psychopathol*, vol. 17, no. 3, pp. 827–850, 2005.

[2] E. Kross, M. Davidson, J. Weber, and K. Ochsner, “Coping with Emotions Past: The Neural Bases of Regulating Affect Associated with Negative Autobiographical Memories,” *Biol. Psychiatry*, vol. 65, no. 5, pp. 361–366, 2009.

[3] S. H. Kim and S. Hamann, “Neural Correlates of Positive and Negative Emotion Regulation,” *J. Cogn. Neurosci.*, vol. 19, no. 5, pp. 776–798, 2007.

[4] J. Slobodskoy-Plusnin, “Behavioral and brain oscillatory correlates of affective processing in subclinical depression,” *J. Clin. Exp. Neuropsychol.*, vol. 40, no. 5, pp. 437–448, 2018.

[5] S. Arroyo, R. P. Lesser, B. Gordon, S. Uematsu, J. Hart, P. Schwerdt, K. Andreasson, and R. S. Fisher, “Mirth, laughter and gelastic seizures,” *Brain*, vol. 116, no. 4, pp. 757–780, 1993.

[6] H. S. Mayberg, A. M. Lozano, V. Voon, H. E. McNeely, D. Seminowicz, C. Hamani, J. M. Schwalb, and S. H. Kennedy, “Deep Brain Stimulation for Treatment-Resistant Depression,” *Neuron*, vol. 45, no. 5, pp. 651–660, 2005.

[7] H. Sloetjes and P. Wittenburg, “Annotation by category - ELAN and ISO DCR,” *Proc. 6th Int. Conf. Lang. Resour. Eval.*, pp. 816–820, 2008.

[8] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” vol. 134, pp. 9–21, 2004.

[9] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, p. 5_32, 2001.

[10] D. W. Gareth James Trevor Hastie, Robert Tibshirani, *An introduction to statistical learning : with applications in R*. New York : Springer, [2013] ©2013.

[11] V. Satopää, J. Albrecht, D. Irwin, and B. Raghavan, “Finding a ‘kneedle’ in a haystack: Detecting knee points in system behavior,” *Proc. - Int. Conf. Distrib. Comput. Syst.*, pp. 166–171, 2011.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[13] O. G. Sani, Y. Yang, M. B. Lee, H. E. Dawes, E. F. Chang, and M. M. Shanechi, “Mood variations decoded from multi-site intracranial human brain activity,” *Nat. Biotechnol.*, vol. 36, no. 10, 2018.

TABLE I. Summary of biomarkers

Subject Identifier	**Summary of limbic regions and selected features**	
	limbic coverage after removing interictal epochs	***Selected features***
Subject 1	OFC, CIN, AMY, INS	CIN (Lγ, Hγ), AMY(Hγ), INS (θ, Hγ)
Subject 2	OFC, CIN, HPC, INS	OFC(Hγ), CIN (θ, Hγ), HPC (Lγ, Hγ), INS (θ, β, Hγ)
Subject 3	OFC, CIN, HPC, AMY	CIN (Lγ, Hγ), HPC(all), AMY(all)
Subject 4	R_OFC, R-CIN, R-HPC, R-INS, L-CIN, L-INS	R-HPC(all), L-CIN (α, Hγ), L-INS (θ, Hγ)
Subject 5	OFC, CIN, HPC, AMY , INS	CIN (β, Lγ, Hγ), HPC (α, Lγ, Hγ), AMY(θ), INS (θ, β)
Subject 6	OFC, CIN, INS	OFC (θ, β, Hγ), CIN(β), INS (Lγ, Hγ)
Subject 7	CIN, HPC, AMY, INS	CIN (θ, Lγ, Hγ), AMY (θ, α, Lγ, Hγ), INS (θ, α, β, Hγ)
Subject 8	OFC, CIN, HPC, INS[1]	OFC (θ, α), HPC (α, β, Lγ, Hγ), INS (θ, Lγ, Hγ)[2]

[1] OFC: Orbitofrontal cortex, CIN: Cingulate cortex, AMY: Amygdala, HPC: Hippocampus & INS: Insula.

[2] θ: theta, α: alpha, β: beta, Lγ: low gamma, Hγ: high gamma.