Single trial prediction of normal and excessive cognitive load through EEG feature fusion

Pouya Bashivan*
Mohammed Yeasin
Department of Electrical and Computer Engineering
The University of Memphis
Memphis, TN, USA
{pbshivan;myeasin}@memphis.edu

Gavin M. Bidelman
Institute for Intelligent Systems and
School of Communication Sciences & Disorders
The University of Memphis
Memphis, TN, USA
gmbdlman@memphis.edu

***Abstract*—Detection of subtle changes in cognitive states (e.g., cognitive overload) or epistemic state of mind remains a challenge. As they typically lack visible expressions, indirect methods like analysis of facial expressions are ineffective at best. Towards solving such problem, we present a statistical approach to predict cognitive load from single trial electrophysiological recordings of brain activity (i.e., EEG). We evaluated the utility of two commonly used sets of features, namely, wavelet entropy and band-specific power (theta, alpha, and beta) to predict levels of cognitive load. We show that performance of the model (i.e., support vector machine) could be improved by feature fusion (such as wavelet entropy and spectral power features together) and also integrating nonlinear representations learned through deep belief networks. Our results demonstrate predictions of cognitive load across four different levels with an overall accuracy of 92% during execution of a memory task.**

Keywords—EEG; cognitive load;wavelet entropy; support vector machines; deep belief networks; random forest

I. INTRODUCTION

Humans are naturally capable of understanding other people's emotions and feelings from facial and behavioral cues. While computational algorithms have made remarkable progress in predicting emotions from facial expressions [1], [2], their ability in prediction of cognitive states (e.g., confusion) and epistemic states of mind remains a challenge. On the other hand, much progress has been made in estimating cognitive states from physiological markers (e.g., pupil dilation, galvanic skin response, heartbeat rate, and brain activity) [3]. Notably, analyzing brain signals ("brain decoding") for cognitive load prediction has received much attention recently [4]–[6]. Gevins et al, [29] used spectral band features to distinguish between three levels of working memory load (low, medium and high). While they reported an accuracy of 95% when discriminating between low and high load levels, classification performance decreased greatly when comparing adjacent load levels (e.g., med-high ~80%). On the other hand, more recently Zarjam et al. [6], adopted a spectral feature called wavelet entropy (WE) for the EEG signals to classify seven cognitive load states during an arithmetic task achieving an impressive accuracy of 98%. However, no comparison between WE and spectral power features have been conducted in the literature to evaluate the prediction power derived from each feature set and whether or not complementary information exists among them.

Computing or predicting changes in cognitive states using single trial EEG is critical for next generation brain-machine interfaces. Demand for monitoring of mental states (e.g., cognitive overload) is important in various fields including the design and implementation of cognitively informed software applications, and diagnosis of psychological disorders [7], [8]. Based on the theory of cognitive load (CL), mental load is directly related to individual's working memory (WM) capacity [9]. Since human WM has a limited capacity, increasing the cognitive demand beyond this limit will result in cognitive overload, a state of decreased task performance as well as lower learning rates [10].

A plethora of neuroimaging modalities are available to monitor brain activity in response to cognitive tasks (e.g., fMRI, MEG, EEG). Electroencephalogram is preferred in many applications because of its portability, lower cost, and ease of use compared to other modalities. Different temporal and spectral characteristics of EEG are often used as representative features of EEG that co-vary with cognitive load [11]. Among these, band-specific power values from the EEG's spectrum have proved to be robust markers in many classification applications. Selection of multiple frequency bands can quickly lead to very high dimensional feature vectors leading to curse of dimensionality and poor classification performance. Naturally, more abstract signal representations are always desired to reduce the feature space dimension. Wavelet entropy has recently been proposed as an abstract representation of the spectral activity which can be used for a more compact spectral feature set [12]. Studies investigating neural activity within the WM network have largely revealed several cortical regions active in response to WM operations (see [13] for a review). In brief, activity in prefrontal cortex and precuneus are mostly observed during a wide range of cognitive and memory tasks. These regions are part of the fronto-parietal pathway, which arguably forms the core of memory-related processing in the brain [14], [15].

Here, we built a statistical model to predict level of cognitive load from spectral features of single trial EEG. In particular, we explored the classification performance of a support vector machine with band-power and wavelet-entropy feature sets in predicting levels of cognitive load and to detect cognitive overload during a WM task. Our key contributions are the selection and fusion of different features to develop a model that can predict the cognitive load based on single trials of a WM task.Critically, our model can distinguish between multiple loads which covered the range within and exceeding the normal range of cognitive demand. In particular, detecting the overload condition is a difficult problem to solve because overload could occur at different load levels depending on the individual's WM capacity. Moreover, in order to find compact representations of EEG data, we adopted a feature selection approach using random forest and furthermore a nonlinear transformation using deep belief network for nonlinear mapping of feature space into lower dimensions.

II. Methods

A. Data recording and preprocessing

Fifteen graduate students (7 male) voluntarily participated in the experiment. Participants completed 240 trials of a visual variant of Sternberg working memory paradigm. The task consisted of memorizing a set of English characters (SET) for a brief time (3 seconds) and respond whether a randomly presented test character (TEST) was among the memorized set or not. To induce various levels of cognitive load, the number of characters presented during each repetition of the task was varied between 2, 4, 6, or 8 items. The load level was randomly selected for each trial with the restriction that each ¼ of trials would be dedicated to each load level (n = 60). The individual's WM capacity K was computed as $K = S(H-F)$, where S is the number of characters in the stimulus, H is the hit rate and F is the false alarm rate [16].

The electroencephalogram (EEG) was recorded throughout the task using a 64 channel electrode cap (Neuroscan Quik-Cap). Electrodes were placed on the standard 10-10 locations. In addition, two bipolar electrodes were also placed on the outer canthi of the eyes and the superior and inferior orbit to detect ocular artifacts. Electrode impedance was kept below 5kΩ throughout the data recording. EEG was recorded in reference to an electrode placed ~1 cm posterior to Cz at a sampling rate of 500 Hz and re-referenced offline to the common average reference.

Portions of the continuous EEG containing ocular artifacts were cleaned using principal component analysis [17]. Data was subsequently down-sampled to 250 Hz for analysis. EEGs were band-pass filtered with a zero-phase FIR filter of order 500. Data was then epoched into segments of 5500 ms spanning the interval from two seconds prior to SET-onset to the TEST onset.

B. EEG Features

1) Band Power

In studies of EEG, the frequency range has conventionally been divided into multiple sub-bands (e.g., delta, alpha, beta). Neural oscillatory responses within each of these sub-bands have found to be correlated with various neural mechanisms [11], [18]. Fast Fourier transform (FFT) was employed to extract the power spectra for each EEG signal. For each electrode and trial, the time span from SET presentation to TEST presentation was selected. This time window contained the complete encoding and maintenance stages of WM operation. Mean spectral power (MSP) within theta (4-7 Hz), alpha (8-13 Hz) and beta (14-30 Hz) bands were extracted by averaging the FFT magnitudes over their corresponding frequency bands. These frequency bands were selected in relation to the numerous evidence of their role in various cognitive functions [27], [28]. to the aggregated feature vector consisted of 192 features (64 electrodes x 3 sub-bands).

2) Wavelet Entropy

Wavelet entropy (WE) provides information about the degree of order existing in a multi-frequency signal [12]. In this framework, periodic signals with a narrow band spectrum are more ordered and therefore would gain lower WE values compared to more complex and unpredictable ones. In the context of cognitive load measurement, this would be of particular interest since higher load levels enhance the power in higher frequency bands while reducing the lower bands power [19]. In order to compute WE, we first find the discrete wavelet transform of the signal with N levels.

$$x(t) = \sum_{j=-N}^{-1} \sum_{k} C_j(k) \psi_{j,k}(t) \tag{1}$$

where $C_j(k)$ is the j-th level wavelet coefficient at sampled time k, and $\psi_{j,k}(t)$ wavelet function with parameters j and k at time t. The energy of the signal at each level (E_j) is then computed by finding the average of wavelet coefficients across time samples (of trial).

$$E_j = \sum_k \left| C_j(k) \right|^2 \tag{2}$$

The relative WE is then computed through normalizing the sub-band energy values by the total energy.

$$E_{tot} = \sum_j E_j \qquad p_j = \frac{E_j}{E_{tot}} \tag{3}$$

Since the resulting normalized energy values are bounded between zero and one they can be treated as a probability distribution. Wavelet entropy is defined as the entropy value of the distribution of normalized energy values.

$$WE = -\sum_j p_j \ln(p_j) \tag{4}$$

We computed the wavelet entropy of each EEG electrode signal for the duration of each trial following the presentation of the stimulus SET.

C. Feature Selection with Random Forest

Performance of a computational model can be improved by choosing compact features that are maximally relevant with minimum redundancy to the desired categories. In many cases, high dimensional features usually lead to degradation of classification performance due to the curse-of-dimensionality. In addition, large number of irrelevant features may lead to poor generalization of the model and incur computational cost. It is always desirable to generate compact representations of data both for storage and computational cost optimization. In order to reduce the size of feature vector, we employed the random forest algorithm to rank the best features for the classification task.

Random forest [20] is an ensemble method, which consists of an array of random trees grown independently. Random forests can handle large number of features without deletion and run efficiently on large datasets. Each tree in the ensemble is grown using data that is randomly sampled from the training set with replacement. In addition, only a subset of features is used to construct each tree that is also randomly selected from the complete set of features. For each tree, one-third of data is kept out of the sample (out-of-bag) and used to obtain an unbiased estimate of classification error of the tree. Random forest uses a permutation process to measure the importance of each variable (feature) in categorizing data samples and subsequently to rank the features. An ensemble of 100 decision trees was constructed to rank the 192 features in this study and to select the top 64 features (equal to the number of electrodes and maximum number of wavelet entropy features). To quantify the variable importance we used the difference between the number of raised margins and the number of lowered margins if the values of that variable were permuted across the out-of-bag observations. This measure was computed for every tree and then averaged over the entire ensemble and divided by its standard deviation.

D. *Classification with support vector machine*

Support vector machine (SVM) [21] is a kernel based classification/regression method widely used in diverse applications. Among various approaches to multivariate classification, SVM has been shown to be a practical and robust method for brain decoding [22]. SVM projects the data into a higher dimensional space through a kernel function and finds a decision boundary with maximal margin.. A common nonlinear kernel used in SVM is the radial basis function (RBF). Because of the inherent nonlinearity existing in the RBF kernel, it can handle the nonlinear relationship between class labels and attributes. SVM maximizes the distance between all data points and the decision boundary in the transformed space (kernel space) and therefore is a large-margin classifier. For a D dimensional feature vector and N number of data points, the solution coefficients are found by solving a standard quadratic program in N+D+1 variables subject to O(N) constraints. SVM provides a sparse solution by considering only a subset of points that are either incorrectly classified or are classified correctly but are on or inside the margin.

All features were z-scores normalized. A SVM with RBF kernel was then used to predict the load level from single-trials of the EEG. SVM hyper parameters consisting of regularization penalty parameter (C) and inverse of RBF kernel's standard deviation ($\gamma = 1/\sigma$) were selected by grid-search through 5 fold cross-validation (C = {0.01, 0.1, 1, 10, 100}, γ = {0.1, 0.2, …, 1, 2, …, 10}). The F-score was used to determine the best hyper-parameters. Finally, results were evaluated using a 10-fold stratified cross validation approach performed on the complete set of trials. The reported number of support vectors was computed by finding the average of all folds.

E. *Deep Belief Network*

To compare performance, we also trained a deep belief network (DBN) to predict the load levels from spectral features. DBN has similar structure to multi-layer perceptron (MLP) with the main difference in the training approach. Deep belief networks find lower dimension representations of the data while iteratively optimize the weights to decrease the prediction error. Parameters of each layer of DBN are pre-trained greedily by treating it as a restricted Boltzmann machine (RBM). This improves learning by setting the initial parameters to more realistic values overcoming the difficulties to train deep neural networks [23]. RBM consists of two layers of stochastic hidden units with only cross-layer connections (i.e., no within layer connections). In spirit, RBM essentially models the distribution of its input and learns this relation by tuning its weights in order to reduce the difference of the true (from training data) and estimated (from model) joint probability between visible (first layer) and hidden (second layer) units. A detailed guide on training RBM can be found in [24].

The network we used here consisted of one Gaussian-Binary RBM and two Binary RBM layers (128, 128, 64) and a final softmax layer which mapped the data representations of the last RBM layer to a multinomial distribution (or binomial for the overload prediction case) for classification. The activation function of hidden layers was selected as sigmoid function. The network was pre-trained using a greedy layer-by-layer unsupervised method [25] on the standardized training data. We used batch stochastic gradient descent (batches of size 10 samples) with L1 regularization to reduce the overfitting during the fine-tuning stage. The network consisted of 49,924 parameters (with 192 MSP features) and was implemented in Theano [30].

III. RESULTS

Performances for different classification approaches are presented in the Table-1. We report the average accuracy and F-score [26] for all classes. Percent of support vectors are tabulated in each case. Accuracy demonstrates the achieved performance in successfully predicting the cognitive load whereas, number of support vectors provides a measure of overlap between the various classes around the separating hyperplane in the transformed space derived from SVM's kernel function. Naturally, having higher accuracy and lower number of support vectors are desired but this will usually remain as a trade-off between the prediction error on training data and the generalization on test data. We predicted WM load levels using mean spectral power (MSP), wavelet entropy (WE), joint MSP and WE (MSPWE), random forest top features of MSP (MSP64), and random forest top features of joint MSP and WE (MSPWE64). The dataset consisted of 2670 trials of the WM task collected from 15 subjects. Table-1 summarizes the classification results for various approaches of features/classifiers for four levels of cognitive load.

TABLE-1 - PERFORMANCE FOR PREDICTING COGNITIVE LOAD VIA DIFFERENT CLASSIFICATION APPROACHES. CHANCE LEVEL PRECISION EQUALS 28%. SV = SUPPORT VECTORS.

Approach	Ratio of SV (%)	Accuracy (%)	F-score
MSP+SVM	31	90.60	0.91
WE+SVM	62	88.17	0.88
MSP64+SVM	**30**	90.26	0.90
MSP+DBN	-	90.35	0.90
MSPWE+SVM	44	**92.13**	**0.92**
MSPWE64+SVM	37	91.16	0.91
MSPWE+DBN	-	90.73	0.91

The best performance across all approaches was observed using the joint WE and MSP feature sets and RBF SVM classifier. While WE features resulted in worst performance across all approaches, considering them jointly with MSP features boosted the accuracy by ~%2. Classification using lower dimension representations (¼ of all features) still performed comparably well, lagging by 1% behind the best performing metrics (MSPWE64+SVM). Interestingly, training SVM with representations extracted from the top layer of DBN gave almost equal performance to MSPWE+SVM (F-score=0.92) proving DBN as an efficient method for finding lower dimension data representations. The confusion matrix for the MSPWE+SVM classifier is shown in Fig. 1. Most errors originated from misclassification between lower load levels while the higher load levels are relatively easier to predict. Specially the lowest load level (set size=2) is mistaken for all three other loads which could be an indication of more signal variation across trials and individuals for the lower loads.

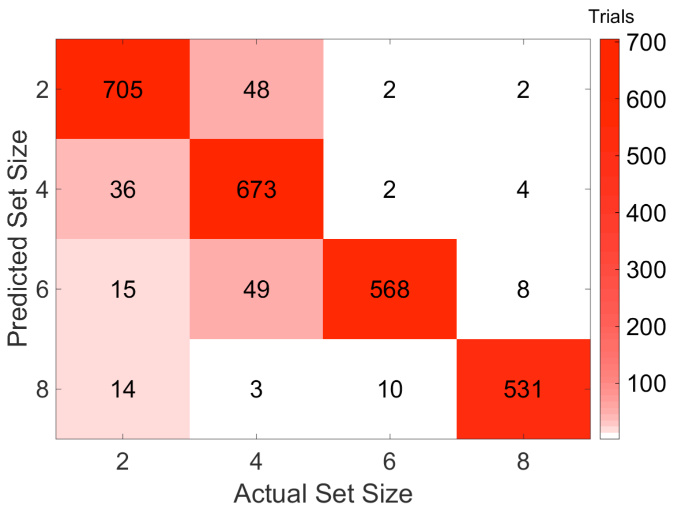

Fig. 1 – Confusion matrix for the best performing classifier (MSPWE+SVM)

Subsequently, we used similar models to detect the WM overload from single trial EEG. Overload condition occurs when the induced load level (i.e. number of characters in SET) exceeds the individual's WM capacity. Here, trials with load level greater than the individual's behavioral WM capacity were considered as overloaded condition. Table-2 summarizes the results derived from various approaches. MSP+DBN and WE+SVM performed best and worst respectively. Interestingly, adding WE features when differentiating between normal load versus overload did very little in improving the results and slightly degraded the performance of DBN.

TABLE-2 – DETECTION PERFORMANCE FOR COGNITIVE OVERLOAD AND DIFFERENT FEATURE SETS. SV = SUPPORT VECTORS.

Feature set	Ratio of SV (%)	Accuracy	F-score
MSP+SVM	26	88.80	0.90
WE+SVM	48	80.80	0.83
MSP64+SVM	32	87.12	0.88
MSP+DBN	-	**89.46**	**0.91**
MSPWE+SVM	**22**	89.14	0.90
MSPWE64+SVM	30	88.16	0.90
MSPWE+DBN	-	88.78	0.90

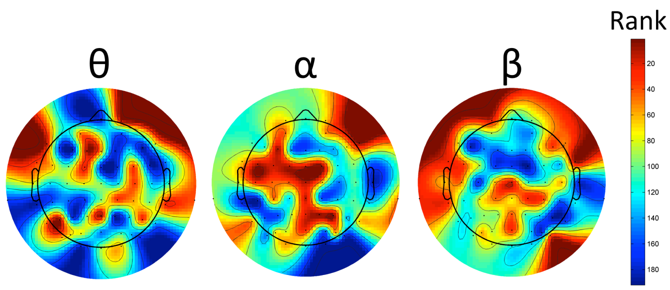

Fig. 2 - Topographic maps of feature importance for theta (left), alpha (middle), beta (right) differentiating between normal and overload conditions. Warmer colors indicate lower rank of the corresponding electrode in the specified frequency band.

Finally, we ranked the spatial and spectral features of EEG based on their prediction power using the random forest method. In order to investigate the spatial and spectral mapping of our features and how they relate to prediction power, we mapped each feature's ranks for the three explored frequency bands over the electrode locations on the scalp. These maps are demonstrated in Fig. 2. Features with lower ranks contain more information regarding the discrimination of various load conditions. Qualitatively, the frontal regions were ranked lower in all three frequency bands; central and parietal electrodes were mostly informative in the alpha frequency band.

IV. DISCUSSION

We adopted a machine learning approach to develop a cognitive load model from EEG data with a high degree of accuracy (> 92 %). The key contributions were to use existing techniques to develop model that can predict the outcome based on a single trial of a WM task. Moreover, we fused the two common feature sets in EEG classification studies (spectral power and wavelet entropy) to improve the prediction power of the model. We evaluated the classification performance of these feature sets on single trial EEG recordings from a challenging working memory experiment. While the SVM classifier generally performed better in both classification problems, nevertheless, we found evidence of improved accuracy when considering both features together. In addition, we performed feature selection using random forest to compute the top features and to identify the underlying cortical networks within each frequency band. In contrast to the usual approach of finding important regions with respect to the task (univariate power tests), random forest ranks the features (cortical regions) in a semi-multivariate approach. This is done through permuting the values of a single variable in out-of-bag set and quantifying its effect on correct classification by all the trees in the forest [20]. Spatial distribution of spectral features over cortical regions including frontal theta and parietal alpha bands were largely in line with findings from other neural spectral studies [27], [28]. The overlapping map between the unsupervised computational approach and experimentally discovered spatial/spectral characteristics of cortical regions suggests the applicability of such multivariable approach to explore the space of potential relationships between brain and behavior. Moreover, our derived model could distinguish between the overload and normal cognitive load states of individuals. This is a critical problem since the cognitive overload condition occurs at different load levels for each individual, which makes this distinction particularly sensitive to individual differences.

Finally, we showed that using nonlinear representation learning methods like DBN to transform the MSP features into a compact set of features, could generate an optimally predictive set of features of desired size. This is a particularly interesting approach which can further be extended to compare with other decomposition methods like principal component analysis, independent component analysis, and non-negative matrix factorization.

V. REFERENCES

[1] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," Pattern Recognit., vol. 36, no. 1, pp. 259–275, 2002.

[2] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," Real-Time Vis. Human-Computer Interact., pp. 181–200, 2005.

[3] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory," Educ. Psychol., vol. 38, no. 1, pp. 63–71, 2003.

[4] C. Berka, D. J. Levendowski, M. N. Lumicao, A. Yau, G. Davis, V. T. Zivkovic, R. E. Olmstead, P. D. Tremoulet, and P. L. Craven, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks.," Aviat. Space. Environ. Med., vol. 78, no. 5 Suppl, pp. B231–B244, 2007.

[5] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," Int. J. Ind. Ergon., vol. 35, no. 11, pp. 991–1009, Nov. 2005.

[6] P. Zarjam, J. Epps, F. Chen, and N. H. Lovell, "Estimating cognitive workload using wavelet entropy-based features during an arithmetic task.," Comput. Biol. Med., vol. 43, no. 12, pp. 2186–95, Dec. 2013.

[7] P. Antonenko, F. Paas, R. Grabner, and T. Gog, "Using Electroencephalography to Measure Cognitive Load," Educ. Psychol. Rev., vol. 22, no. 4, pp. 425–438, Apr. 2010.

[8] P. Barttfeld, B. Wicker, S. Cukier, S. Navarta, S. Lew, and M. Sigman, "A big-world network in ASD: Dynamical connectivity analysis reflects a deficit in long-range connections and an excess of short-range connections," Neuropsychologia, vol. 49, no. 2, pp. 254–263, 2011.

[9] J. Sweller, "Cognitive Load During Problem Solving : Effects on Learning," vol. 285, pp. 257–285, 1988.

[10] J. Sweller, J. J. G. Van Merrienboer, and F. G. W. C. Paas, "Cognitive Architecture and Instructional Design," vol. 10, no. 3, pp. 251–296, 1998.

[11] P. Bashivan, G. M. Bidelman, and M. Yeasin, "Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity," Eur. J. Neurosci., vol. 40, no. 12, pp. 3774–3784, 2014.

[12] O. a Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, E. Başar, M. Schu, and E. Bas, "Wavelet entropy: a new tool for analysis of short duration brain electrical signals.," J. Neurosci. Methods, vol. 105, no. 1, pp. 65–75, Jan. 2001.

[13] C. Rottschy, R. Langner, I. Dogan, K. Reetz, a R. Laird, J. B. Schulz, P. T. Fox, and S. B. Eickhoff, "Modelling neural correlates of working memory: a coordinate-based meta-analysis.," Neuroimage, vol. 60, no. 1, pp. 830–46, Mar. 2012.

[14] P. Bashivan, G. M. Bidelman, and M. Yeasin, "Neural correlates of visual working memory load through unsupervised spatial filtering of EEG," in Proceedings of 3rd workshop on Machine Learning and interpretation in neuroimaging, 2013.

[15] P. Bashivan, G. M. Bidelman, and M. Yeasin, "Modulation of Brain Connectivity by Memory Load in a Working Memory Network," in Proceedings of IEEE Symposium Series on Computational Intelligence (SSCI), 2014.

[16] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behav. Brain Sci., vol. 24, pp. 87–185, 2000.

[17] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods," Int. J. Psychophysiol., vol. 53, no. 2, pp. 105–119, 2004.

[18] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," Brain Res. Rev., vol. 29, no. 2, pp. 169–195, 1999.

[19] L. Michels, K. Bucher, R. Lüchinger, P. Klaver, E. Martin, D. Jeanmonod, and D. Brandeis, "Simultaneous EEG-fMRI during a working memory task: modulations in low and high frequency bands.," PLoS One, vol. 5, no. 4, p. e10298, Jan. 2010.

[20] L. Breiman, "Random forests," Mach. Learn., pp. 5–32, 2001.

[21] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[22] J. R. Sato, A. Fujita, C. E. Thomaz, M. D. G. M. Martin, J. Mourão-Miranda, M. J. Brammer, and E. A. Junior, "Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction," Neuroimage, vol. 46, no. 1, pp. 105–114, 2009.

[23] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," Adv. Neural Inf. Process. Syst., vol. 19, p. 153, 2007.

[24] G. E. Hinton, "A practical guide to training restricted boltzmann machines," in Neural Networks: Tricks of the Trade, Springer, 2012, pp. 599–619.

[25] G. E. Hinton, G. E. Hinton, S. Osindero, S. Osindero, Y. W. Teh, and Y. W. Teh, "A fast learning algorithm for deep belief nets.," Neural Comput., vol. 18, no. 7, pp. 1527–54, 2006.

[26] D. Powers, "EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION," J. Mach. Learn. Technol., vol. 2, no. 1, pp. 37–63, 2011.

[27] O. Jensen and C. D. Tesche, "Frontal theta activity in humans increases with memory load in a working memory task," Eur. J. Neurosci., vol. 15, no. 8, pp. 1395–1399, 2002.

[28] O. Jensen, J. Gelfand, J. Kounios, and J. E. Lisman, "Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task.," Cereb. Cortex, vol. 12, no. 8, pp. 877–82, Aug. 2002.

[29] Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G., 1998. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum. Factors J. Hum. Factors Ergon. Soc. 40, 79–91.

[30] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. Goodfellow, A. Bergeron, N. Bouchard, D. Warde-Farley and Y. Bengio. "Theano: new features and speed improvements". NIPS 2012 deep learning workshop.