Mental Workload Classification via Hierarchical Latent Dictionary Learning: A Functional Near Infrared Spectroscopy Study

Srinidhi Parshi, *Student Member, IEEE*, Md. Rafiul Amin, *Student Member, IEEE*, Hamid Fekri Azgomi, *Student Member, IEEE*, and Rose T. Faghih, *Member, IEEE*

***Abstract*— Variations in the brain's blood oxygenation and deoxygenation reflect neuronal activation patterns, and can be measured using functional near infrared spectroscopy (fNIRS). We aim to utilize fNIRS to obtain insights into the dynamic functional connectivity of the brain as a function of the mental workload. Interpreting connectivity in the brain using noisy fNIRS data with low signal to noise ratio is challenging. To overcome the challenges with fNIRS data, we use a hierarchical latent dictionary learning approach. This approach provides covariance matrices to obtain the dynamic functional connectivity and neuronal activation patterns that change over time. We use features from the dynamic functional connectivity of the brain reflected in fNIRS data collected from the prefrontal cortex to investigate mental workload. In particular, we study three types of mental workload tasks called *n-back* tasks and perform binary classification for each *n-back* task compared to the other *n-back* tasks and rest condition using support vector machines. The results of our binary classification of various *n-back* tasks compared to the rest condition outperforms binary classification results reported previously.**

I. Introduction

Functional Near Infrared Spectroscopy (fNIRS) is a functional brain imaging technique based on measuring hemodynamic responses that reflect neuronal activation patterns of the brain. The near infrared region of the electromagnetic spectrum (620-1000 nm) is scattered by most biological tissue, but absorbed by hemoglobin. fNIRS uses this phenomenon to estimate the levels of oxygenated and deoxygenated hemoglobin (HbO and HbR, respectively) using modified Beer Lambert Law [1]. Of the many other techniques, such as Functional Magnetic Resonance Imaging (fMRI) and Magnetoencephalography (MEG) used in brain imaging, fNIRS is the one with the most compact and portable equipment, thus making it ideal to measure a large number of subjects in different kinds of environments. Studies also show that electroencephalography (EEG) data can be used to classify *n-back* tasks in single trials [2], [3]. EEG, while providing excellent temporal resolution, suffers from poor spatial resolution whereas fNIRS exhibits the opposite characteristics [4]. We use the better spatial resolution of fNIRS to obtain the functional connectivity of the different regions of brain. fNIRS optodes can also be placed over the entire scalp region covering the full 10-5 international system [5]. Once the region of interest is known, it can also be used to extract data from exactly those regions of the cortex. Most fNIRS studies focus on specific functionalities such as mental arithmetic, cognitive functions, and difficulty levels of executing certain tasks [6], [7]. A brain computer interface interprets neuronal activation in the brain as input commands to a computer. Study by *Herff et al* [8] focuses on the *n-back* task and its corresponding neural activation in the prefrontal cortex of the brain that is known to serve an important role in memory related tasks such as the *n-back* experiment. There have been studies conducted to assess the cognitive state of the brain using various biological signals such as skin conductance, neural spiking activity, etc [9], [10], [11], [12], [13], [14]. The low signal to noise ratio (SNR) of the fNIRS data hinders analysis. A novel approach to this problem has been proposed in [15] where repeated trials or scenarios are used to extract data from adjacent sensors and recover the underlying signal. The correlations obtained can be used to gain insights into the dynamic functional connectivity of the brain. The shared covariance between the trials of a single *n-back* experiment reflects the coactivation patterns in the brain that are specific for an experiment, instead of the trial.

In this paper, we aim to extract the shared covariance matrix for each of the *n-back* tasks for each subject from the dataset in [8]. We seek to observe the dynamic functionality of the brain and their differences for each of the *n-back* tasks. Some limitations of using latent factor models can be that they sometimes yield irrelevant structure and it is difficult to find meaning for the several latent factors found from the data. Nevertheless, the Hierarchical Latent Dictionary Model (HLDM) still gives the dynamic connectivity with reasonable accuracy. We obtain definite results of differences in the neuronal pathways when various *n-back* were performed. We then make use of machine learning algorithms, more specifically the support vector machines (SVM) method to classify these tasks using the covariance matrix derived by the latent factor model. We successfully obtain a higher classification accuracy compared to [8], thus demonstrating the viability of our method.

II. Methods

In this section, we first discuss the experimental. We then outline the hierarchical latent dictionary model (HLDM). Furthermore, we discuss the details of how the model was applied to the dataset described.

Srinidhi Parshi, Md. Rafiul Amin, Hamid Fekri Azgomi and Rose T. Faghih are with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA (e-mail:{sparshi, mamin, hfekriazgomi, rtfaghih}@uh.edu). This work was supported in part by NSF Grant 1755780. Correspondence should be addressed to senior author Rose T. Faghih.

A. Description of the n-back Experiment

The *n-back* task is a continuous performance task used to measure the working memory and working memory capacity. The subject is presented with a series of stimuli, and is then asked to respond when the given stimulus is the same as the previous n^{th} stimulus. In the dataset, the subjects use the space key to indicate the target . As is evident from the nature of the experiment, the difficulty of the *n-back* task increases with n, since the subject is required to remember more of the preceding stimuli for larger n values. This implies that the covariance matrix of the *1-back* task should be significantly different from that of the *3-back* task. This feature causes it to have better classification accuracy between the two tasks. In this study, we analyze the hemodynamic data from the *n-back* experiment in [8].

Dataset: Focusing mainly on the cognitive tasks and the prefrontal cortex of the brain, the dataset in [8] was collected by performing 3 n-back tasks with 10 trials for each task. Each of the trials started with a 5 seconds instruction about the task followed by 22 numbers, with 0.5 seconds for displaying the number and 1.5 seconds for the subject to respond to the given stimulus resulting in a total of 44 seconds of total stimulus time. A 15 seconds relaxation break was provided after each task. Every trial consisted of 3 ± 1 targets. After 15 trials a break period of 150 seconds was incorporated. An Oxymon Mark III by Artinis Medical Systems was used to measure the hemodynamic responses. Two wavelengths of 765 nm and 856 nm were used to measure concentration changes in HbO and HbR, respectively. Since the region of interest was the prefrontal cortex, dataset [8] makes use of four transmitter and four receiver optodes placed on the forehead such that each detector measures time multiplexed data from two sources thus resulting in a total of 8 channels of HbO and HbR. The sampling frequency was set to 25 Hz.

Table I refers to the information of the dataset used. In the study in [8], responses to 3 *n-back* tasks were recorded for each subject. Each *n-back* task had 10 trials resulting in a total of 30 trials per subject. In the dataset, there were 8 channels, all present in the prefrontal cortex region. Therefore, the observation matrix $y^{tr,n}(t)$ was 8×1. The task-specific covariance matrix has a dimension of 8×8. The mean dimensions are trial specific. Since dataset [8] has 10 trials per subject per *n-back* task, the mean dimensions will be 10×8, i.e, 10 trials per *n-back* task for 8 channels.

TABLE I: Dataset information

Data	dataset in [8]
Subjects	10
Channels	8
$y^{tr,n}(t)$ Dimension	8×1
Covariance Dimension	8×8
Mean Dimension	10×8

B. Latent Factor Analysis

Latent factor model relates a set of observable variables to a set of latent factors that have a lower dimension to that of the observed variables. Modelling the observed data to be expressed as functions of a number of latent or hidden factors helps in reducing the high dimensionality of the fNIRS data and to isolate the unobserved latent variables. The observed data are modelled as linear combinations of the latent factors in addition to an error term. This will result in finding the covariance matrices that reflect the relationships among many variables in terms of a few underlying, but unobserved random quantities. Designing the observation matrix $y^{tr,n}(t)$ as a $p \times 1$ matrix wherein each element represents data from one sensor at a particular point in time, and assuming $y^{tr,n}(t)$ is a Gaussian process with $\mathcal{N}_p(0, \Sigma)$, the latent factor model can be written as

$$y^{tr,n}(t) = C_n(t)x^{tr,n}(t) + \epsilon(t) \tag{1}$$

where $x^{tr,n}$ is the k dimensional latent factor with $k << p$, C_n is the factor loading matrix, and ϵ is the Gaussian noise factor $\epsilon \in \mathcal{N}_p(0, \Sigma_0)$, Σ_0 is a diagonal matrix with correlations as the diagonal elements. $\Sigma = C_n \times C_n^T + \Sigma_0$ is the $p \times p$ covariance matrix of the Gaussian process $y^{tr,n}(t)$ [16], [17], [18].

A Dynamic Latent Factor model assumes markov evolution for the latent factors $x^{tr,n}$, with time invariant parametrization [19], [20]. Hence the stationary process $x^{tr,n}(t)$ follow the equations described in equations (2) where $A^{tr,n}$ is the transition matrix, I_k is a k dimensional identity matrix and $v(t)$ is the Gaussian noise factor ($v(t) \in \mathcal{N}_k(0, I_k)$):

$$x^{tr,n}(t) = A^{tr,n}(t)x^{tr,n}(t-1) + v(t) \tag{2}$$

C. Hierarchical Latent Model

The factors are modelled to evolve non-parametrically, so as to capture the long range dependencies between data points. Thus the p dimensional observation matrix $y^{tr,n}(t)$ for trial number tr, task n at discrete point time t with trial specific mean $\mu^{tr,n}(t)$ and task specific covariance $\Sigma^n(t)$ can be written as,

$$y^{tr,n}(t) \sim \mathcal{N}_p(\mu^{tr,n}(t), \Sigma^n(t)) \tag{3}$$

The latent factors $x^{tr,n}(t)$ are modelled as k latent dictionary functions $\phi^{tr,n}(t)$ each of which is a trial specific random function [21], [16] :

$$x^{tr,n}(t) = \phi^{tr,n}(t) + v(t) \tag{4}$$

Sharing of knowledge within this latent space is done by hierarchically coupling the trial specific dictionary functions around the mean $\phi_0^{tr,n}(t)$:

$$\phi^{tr,n}(t) = [\phi_1^{tr,n}(t), ..., \phi_k^{tr,n}(t)] \tag{5}$$

We define q_0 & q_1 as kernel functions of the Gaussian Process depicting characteristics such as smoothness, continuity, periodicity, etc. Here we consider the squared exponential kernel function defined as:

$$q_i(t, t') = d_i exp(-\kappa \parallel t - t' \parallel_2^2) \quad \text{for } i = 0, 1 \tag{6}$$

where d_i is a scaling parameter and κ is the bandwidth. Equations (7) and (8) describe the hierarchical coupling by equating the mean of the child processes to the parent process $\phi_0^{tr,n}(t)$ [15]:

$$\phi_j^{tr,n}(t) \sim GP(\phi_0^{tr,n}(t), q_1) \tag{7}$$

$$\phi_0^{tr,n}(t) \sim GP(0, q_0) \tag{8}$$

where GP is a Gaussian process. Heteroscedasticity is captured by making the factor loading matrix C_n to be time evolving and weighted combination of smaller dimensional latent dictionary functions:

$$C_n(t) = \Theta^n(t)\zeta^n(t) \tag{9}$$

where $\Theta^n \in \mathbb{R}^{p \times L}$ is distributed according to a conditionally Gaussian shrinkage prior and $\zeta^n(t)$ is a Gaussian processes with zero mean and kernel function q_0.

The resulting trial specific mean $\mu^{tr,n}(t)$ and task specific covariance $\Sigma^n(t)$ will be as,

$$\mu^{tr,n}(t) = \Theta^n\zeta^n(t)\phi^{tr,n}(t) \tag{10}$$

$$\Sigma^n(t) = \Theta^n\zeta^n(t){\zeta^n}^T(t){\Theta^n}^T + \Sigma_0^n \tag{11}$$

where Σ_0^n is a diagonal matrix with diagonal elements being the correlations.

III. Classification

In this section, we outline how binary classification was performed using the covariance matrices.

We obtain and analyze the covariance matrix for one subject for one task. This matrix represents the dynamic functional changes in the brain with respect to different tasks. The obtained covariance matrix for one task at one time point is a $p \times p$ matrix with diagonal values as 1. For N time points, we can imagine a 3 dimensional covariance matrix shown in Figure 1. Since the correlation between channels 1 & 2 is same as correlation between 2 & 1 the matrix is symmetric in nature, i.e out of the $8 \times 8 = 64$ values obtained from dataset [8], 28 are unique. We take the mean, variance and third order central moment of these 28 values across the N time points. Thus we have a $3 \times 28 = 84$ features that are used for classification. Each *n-back* task results in one covariance matrix. Recall that the covariance $\Sigma^n(t)$ is task specific whereas the mean $\mu^{tr,n}(t)$ is trial specific. The end result of a *1-back* task with 10 trials for one subject will be

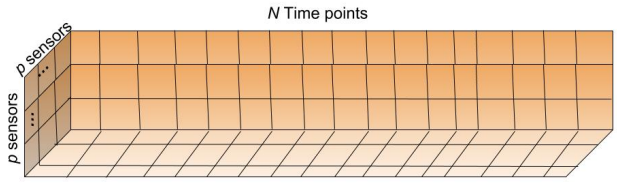

Fig. 1: Visualization of the covariance matrix. At each time point from 1 to N, the $p \times p$ covariance matrix represents correlations between the p sensors. Diagonal elements are 1.

For classification, we use SVM. We use a linear SVM classifier that represents the data as points in space such that separate categories are divided as clusters with wide gaps between them. It creates hyperplanes or set of hyperplanes that can then be used for classification. We perform classification on the covariance matrices obtained thus differentiating between the 3 *n-back* tasks. This differs greatly form the mean feature from each of the windows used by [8]. A straight line is fitted to a window of the data. This line's slope used in [8] to classify the different tasks. The overall slope of the line differs for each *n-back* task. The classification accuracies are presented and discussed in Section IV.

IV. Results

In this section, we state the different results obtained by applying HLDM [15] on the dataset. We highlight the dimensions of the covariance matrices obtained and compare them to isolate the region of interest for memory related tasks like the *n-back*.

Figure 2 shows the 8×8 plot of the covariance matrix for dataset [8]. We see a marked difference in the *1-back* task before and 500 ms after stimulus. The correlations between sensors have increased as is evident from the darker shades of blue and violet colours. Similarly we note the difference in correlations before and 500 ms after stimulus in the *3-back* task. We see a remarkable difference between the *1-back* and *3-back* tasks. While the *1-back* task has more correlations between sensor channels, the *3-back* task's correlations seem to be focused between certain select sensor channels.

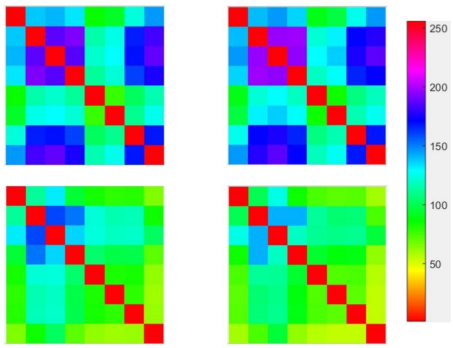

Fig. 2: An example plot of the 8×8 covariance matrix. The top left matrix depicts the result for *1-back* task when the subject is at rest. The top right matrix shows the response 500 ms after the stimulus. The bottom left matrix shows the *3-back* task result when the subject is at rest. The bottom right is the covariance matrix 500 ms after the stimulus. The diagonal elements are self correlated, hence are equal to 1. For better distinguishable plots each of the elements were assigned on the scale of 1 to 256

Table II refers to the classification accuracies obtained for all the 3 tasks in the dataset [8] and the accuracies obtained with our HLDL-SVM approach.

TABLE II: Classification accuracies for *1-back, 2-back, 3-back* tasks using mean, variance and third order moment of $\Sigma^n(t)$.

Task	Accuracy in [8]	Accuracy using our HLDL-SVM
1-back, 2-back	58.5%	65.6%
2-back, 3-back	61%	67.5%
1-back, 3-back	78.0%	75.0%

Table III refers to the classification accuracies and their respective accuracies for *1-back and 2-back* tasks against RELAX (i.e. rest period) in the dataset [8] and the accuracies obtained with our HLDL-SVM approach.

TABLE III: Classification accuracies for *1-back, 2-back* tasks against RELAX task using mean, variance and third order moment of $\Sigma^n(t)$.

Task	Accuracy in [8]	Accuracy using our HLDL-SVM
1-back-RELAX	71.5%	75%
2-back-RELAX	80.3%	87.5%
3-back-RELAX	80.5%	87.5%

V. Discussion and Conclusion

Table II clearly depicts the successful classification between the three *n-back* tasks for a subject. The *1-back* task and the *3-back* task have the highest accuracy, as is expected due to the differences in their difficulty levels. The same reasoning can be applied to *2-back* and *3-back* accuracy of 67.5%. In Table III we classify the 3 *n-back* tasks against the RELAX signal where the rest period was taken as baseline. In this case, we are able to achieve better accuracy while distinguishing *1-back*, *2-back*, and *3-back* tasks against RELAX compared with results reported in [8].

In this paper we apply heirarchical latent dictionary functions on fNIRS data. We obtain and analyze the covariance matrices which depict the neuronal connections in the brain. The correlations obtained can be used to gain insights into the dynamic functional connectivity of the brain. The shared covariance between the trials of a single *n-back* experiment reflects the coactivation patterns in the brain that are specific for an experiment. The heteroscedasticity built into the model further helps to capture time varying covariances between sensors. One of the applications of functional connectivity is to detect variations in different cognitive states of the brain. We then use the covariance matrices obtained to classify the tasks using the SVM algorithm. Studies like [22] show the application of fNIRS for cognitive tasks. Our model can be used to better obtain and classify these cognitive processes. Final classification results verify that our proposed method obtains better classification accuracies between a variety of binary classifications than [8].

References

[1] A. Sassaroli and S. Fantini, "Comment on the modified beerlambert law for scattering media," *Physics in Medicine Biology*, vol. 49, no. 14, p. N255, 2004.

[2] 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, "Research article eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks."

[3] C. L. Baldwin and B. Penaranda, "Adaptive training using an artificial neural network and eeg metrics for within- and cross-task workload classification," *NeuroImage*, vol. 59, no. 1, pp. 48 – 56, 2012. Neuroergonomics: The human brain in action and at work.

[4] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K.-R. Mller, and B. Blankertz, "Enhanced performance by a hybrid nirseeg brain computer interface," *NeuroImage*, vol. 59, no. 1, pp. 519 – 529, 2012. Neuroergonomics: The human brain in action and at work.

[5] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713 – 719, 2001.

[6] K. K. Ang, C. Guan, K. Lee, J. Q. Lee, S. Nioka, and B. Chance, "A brain-computer interface for mental arithmetic task from single-trial near-infrared spectroscopy brain signals," in *2010 20th International Conference on Pattern Recognition*, pp. 3764–3767, Aug 2010.

[7] S. D. Power, A. Kushki, and T. Chau, "Intersession consistency of single-trial classification of the prefrontal response to mental arithmetic and the no-control state by nirs," *PLOS ONE*, vol. 7, pp. 1–12, 07 2012.

[8] C. Herff, D. Heger, O. Fortmann, J. Hennrich, F. Putze, and T. Schultz, "Mental workload during n-back task quantified in the prefrontal cortex using fnirs," *Frontiers in Human Neuroscience*, vol. 7, p. 935, 2014.

[9] R. T. Faghih, P. A. Stokes, M.-F. Marin, R. G. Zsido, S. Zorowitz, B. L. Rosenbaum, H. Song, M. R. Milad, D. D. Dougherty, E. N. Eskandar, *et al.*, "Characterization of fear conditioning and fear extinction by analysis of electrodermal activity," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 7814–7818, IEEE, 2015.

[10] D. Wickramasuriya, C. Qi, and R. Faghih, "A state-space approach for detecting stress from electrodermal activity.," in *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2018, pp. 3562–3567, 2018.

[11] D. S. Wickramasuriya and R. T. Faghih, "Online and offline anger detection via electromyography analysis," in *Healthcare Innovations and Point of Care Technologies (HI-POCT), 2017 IEEE*, pp. 52–55, IEEE, 2017.

[12] X. Deng, R. T. Faghih, R. Barbieri, A. C. Paulk, W. F. Asaad, E. N. Brown, D. D. Dougherty, A. S. Widge, E. N. Eskandar, and U. T. Eden, "Estimating a dynamic state to relate neural spiking activity to behavioral signals during cognitive tasks," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 7808–7813, IEEE, 2015.

[13] R. T. Faghih, "From physiological signals to pulsatile dynamics: a sparse system identification approach," in *Dynamic Neuroscience*, pp. 239–265, Springer, 2018.

[14] M. R. Amin and R. T. Faghih, "Inferring autonomic nervous system stimulation from hand and foot skin conductance measurements," in *Asilomar Conference on Signals, Systems, and Computers*, 2018.

[15] A. Fyshe, E. Fox, D. Dunson, and T. Mitchell, "Hierarchical latent dictionaries for models of brain activation," in *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics* (N. D. Lawrence and M. Girolami, eds.), vol. 22 of *Proceedings of Machine Learning Research*, (La Palma, Canary Islands), pp. 409–421, PMLR, 21–23 Apr 2012.

[16] A. Bhattacharya and D. B. Dunson., "Sparse bayesian infinite factor models," *Biometrika vol. 98,2 (2011): 291-306.*

[17] C. M. Carvalho, J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang, and M. West, "High-dimensional sparse factor modeling: Applications in gene expression genomics," *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1438–1456, 2008. PMID: 21218139.

[18] D. Knowles and Z. Ghahramani, "Infinite sparse factor analysis and infinite independent components analysis," in *Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation*, ICA'07, (Berlin, Heidelberg), pp. 381–388, Springer-Verlag, 2007.

[19] H. F. Lopes, E. Salazar, and D. Gamerman, "Spatial dynamic factor analysis," *Bayesian Anal.*, vol. 3, pp. 759–792, 12 2008.

[20] M. West, "Bayesian factor regression models in the "large p, small n" paradigm," in *Bayesian Statistics*, pp. 723–732, Oxford University Press, 2003.

[21] E. B. Fox and D. B. Dunson, "Bayesian nonparametric covariance regression," *Journal of Machine Learning Research*, vol. 16, pp. 2501–2542, 2015.

[22] P. M. Arenth, J. H. Ricker, and M. T. Schultheis, "Applications of functional near-infrared spectroscopy (fnirs) to neurorehabilitation of cognitive disabilities," *The Clinical Neuropsychologist*, vol. 21, no. 1, pp. 38–57, 2007.