COGNITIVE STATE PREDICTION VIA TWO-DIMENSIONAL FEATURE VECTOR

Sultan Alshehri ⋆*, Fahad Albalawi* ⋆*, Abderrazak Chahid* ⋆ *and Taous-Meriem Laleg-Kirati*

King Abdullah University of Science and Technology
Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE)
Thuwal 23955-6900, KSA

ABSTRACT

Predicting human cognitive tasks from their corresponding functional Magnetic Resonance Imaging (fMRI) data is very challenging. The difficulty of this prediction problem can be summarized in two points first: the size of the dataset is very small due to the small number of subjects (i.e., patients) who can contribute to these research-based experiments, second: the size of feature vector is very large compared to the few number of the samples that can be used to derived a prediction model. One possible way to overcome these obstacles is to develop a feature generation methodology that can result a small-sized and descriptive feature vector that may improve the overall performance of the cognitive state classification problem. Motivated by these considerations, we proposes a novel feature generation methodology termed voxel weight-based (VW) features that can represent the voxels intensity when a subject is performing a certain cognitive task. This feature generation technique can project the high-dimensional feature vector into a two-dimensional feature domain. After generating the VW-based feature set, a logistic regression model (LRM) is utilized to distinguish between two cognitive states that correspond to two distinct tasks (whether a subject is viewing a picture or a sentence). To demonstrate the efficacy of the proposed feature generation scheme, a benchmark fMRI dataset is utilized to assess the performance of the LRM under the proposed features.

***Index Terms*—** Machine Learning, Position Weight Matrix, Quantiziation, Classification, Feature Generation.

1. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) has emerged as a powerful brain imaging modality that measures indirectly the brain activity with good spatial and temporal resolution. It provides three-dimensional scans of the brain per unit of time which allows the detection of the activated regions as a response to a neural activity [1]. Since its discovery [2], fMRI has been used for the detection and monitoring of neurological diseases and disorders such as Schizophrenia [3] and Alzheimers disease [4]. fMRI captures the changes in blood oxygen levels that occur in the activated regions in the brain which provides a signal called Blood Oxygen Level Dependent (BOLD) signal. The BOLD signal is triggered by changes in the cerebral blood flow and reflects the increase in the deoxyhemoglobin content. fMRI modality has been used in many clinical and research conditions allowing the generation of a huge amount of data; as a result, several studies focused on the use of data-driven approaches to analyze, interpret and extract relevant information from fMRI data. A particular interest has been on the development of machine learning methods to classify the cognitive state of a human subject based on fMRI data. For example, in [5], the authors decoded cognitive states that correspond to distinct tasks such as viewing a sentence/picture and reading an ambiguous/non-ambiguous sentence. The main challenge that the authors have addressed when decoding human cognitive states is the big discrepancy between the number of available samples for a given cognitive state and the dimension of the feature vectors. Therefore, optimal search techniques can be developed to choose an optimal feature combination set with a well-suited classifiers [6, 7].

Motivated by this, in [8] a cascade of classifiers that can improve the prediction performance of cognitive state has been proposed. However, the issue of over-fitting associated with the fMRI data has not been considered. Therefore, the authors in [9] have developed a group of predictors named Generalized Sparse Classifiers (GSC) to address the issue of over-fitting, due to the large dimensional data. The derived group of classifiers were applied to a benchmark dataset called star/plus [10], and achieved an average accuracy of 93.7%.In [11], an algorithm called Support Vector Decomposition Machine (SVDM) that combines feature selection and classification learning into one single step was developed. Although the SVDM was able to reduce the high dimensional of the data to 8 features, its prediction performance to the star/plus dataset was unsatisfactory with an average accuracy of 78%. In [12], the authors have proposed a procedure to enable classification between two chosen cognitive tasks, using their respective fMRI image sequences. Different classifica-

Computer Electrical and Mathematical Science and Engineering (CEMSE) division, King Abdullah University of Sciences and Technology (KAUST), Thuwal, KSA e-mail: fahad.albalawi@kaust.edu.sa , abderrazak.chahid@kaust.edu.sa, TaousMeriem.Laleg@kaust.edu.sa

⋆ These authors are shared first co-authors having contributed equally.

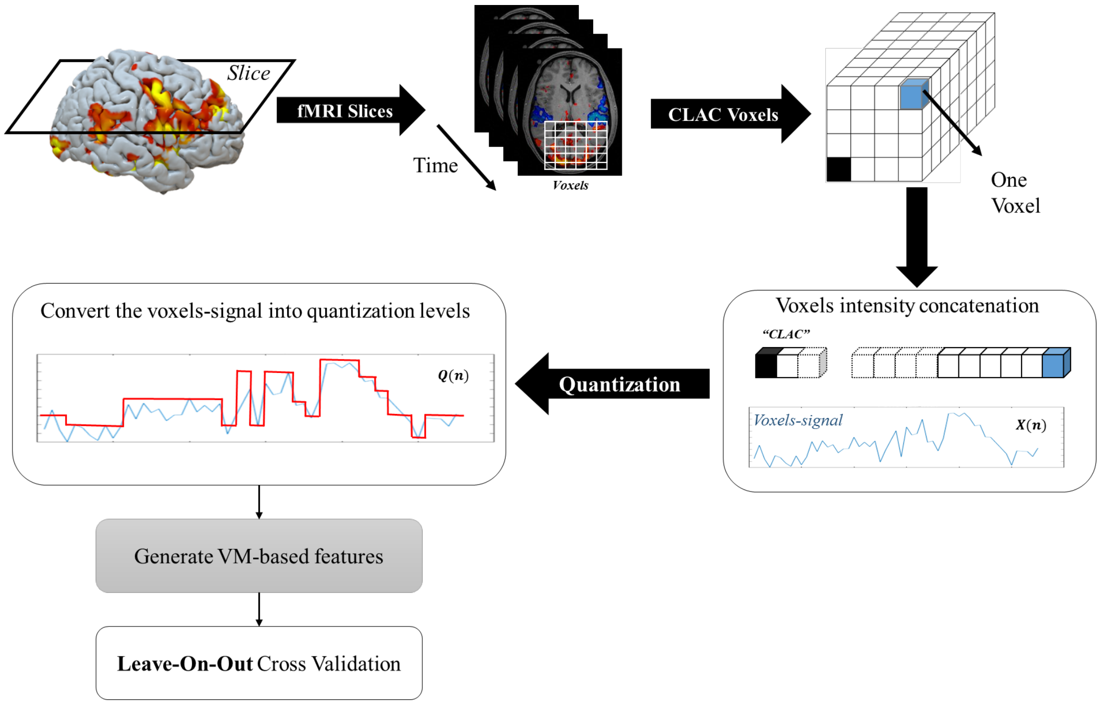

Fig. 1: The proposed method framework.

tion methods with two signal processing-based features were applied to the star/plus dataset where an average accuracy of 99% was obtained using the support vector machine algorithm. Even though the proposed method was able to achieve high prediction accuracy, the dimension of the extracted feature vector was extremely high. For classification purposes, it is highly unsatisfactory to utilize high dimensional feature vector and few number of observations (i.e., samples) to generate a prediction model because this affects negatively the model extraction. Therefore, a new feature generation method with well-suited classifier that can both reduce the high-dimensional fMRI feature vector and improve the overall prediction performance is necessary.

Inspired by the above considerations, we propose a novel methodology that can generate a set of features termed voxel weight-based (VW-based) features. This set of features can represent the voxel activity in the human brain when performing cognitive tasks. The main advantage of this new feature set is the ability to project the high-dimensional voxels features vector into a two-dimensional feature domain. Star/plus dataset has been used to assess the performance of the proposed features for the classification of cognitive states.

2. DATASET

The star/plus dataset experiment proposed in ([10]) was used to demonstrate the proposed feature methodology. In that experiment, fMRI snapshots were obtained every half second (repetition time) when six subjects were performing two distinct cognitive tasks. Particularly, every subject first sees a sentence (semantic stimulus) or a picture (symbol stimulus) for 4 seconds, then a blank screen for 4 seconds is shown to the subject. Every sample is a collection of fMRI 8-seconds period: 4-seconds period of sentence or picture stimulus followed by 4-seconds period of blank screen. Following this strategy, a total of 80 samples are generated from each subject (40 samples for sentence class and 40 samples for picture class). Every sample includes 16 fMRI snapshots, resulting in an input feature vector of size $16N$ where N represents the number of active voxels in a particular Regions Of Interest (ROIs) when the subject sees either a picture or a sentence. Due to the variation of the brain morphology between subjects, the number of active voxels N within the ROIs is different for each subject. The dataset consists of 25 anatomically defined ROIs. Based on the hint mentioned in [10], only seven of these regions were recommended to be used when one plans to train a classifier to distinguish whether the subject is viewing a picture or sentence. A recent research work [13] has examined the classification performance of these seven ROIs, and the authors concluded that fMRI dataset that corresponds to a region of interest called "CALC" achieved the maximum prediction accuracy. Therefore, this work only uses the dataset that corresponds to the "CALC" ROIs.

3. FEATURE GENERATION FOR COGNITIVE STATE PREDICTION

In this section, the proposed features generation technique for the aforementioned fMRI dataset will be presented. The first subsection provides a brief description of the quantization of voxel-signals, as a pre-processing step before generating the VW-based feature vector, which transforms the real-valued voxel-sequences into a discrete sequence of levels. In the second subsection, the VW-based features methodology will be presented. Figure 1 depicts the framework of the VW-based features generation methodology beginning from the raw data (i.e., fMRI snapshots) and resulting in the final VW-based feature vector.

3.1. Voxel intensity quantization

Quantization is a well known technique used for analog-to-digital conversion [14]. In this work, the quantization is utilized to convert the real-valued voxels intensity sequence $X(n)$ to a discrete-valued sequence $Q(n)$. In order to choose a suitable quanatization scheme that can effectively transform the real-valued voxels intensity sequences $X(n)$ into discrete sequence $Q(n)$, we first need to analyze the probability distribution of the real-valued voxels intensity sequence for both classes (i.e. picture and sentence). To do so, we observed the histogram of the voxels intensity values of the six subjects for both classes. Figure 2 shows the fitted Gaussian Probability Distribution Function (PDF) to the histogram plots for both classes where we zoomed in at the region where there is a discrepancy between the two Gaussian distributions.

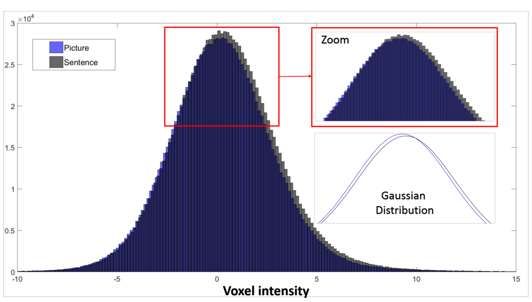

Fig. 2: Voxels intensity histogram for the six subjects.

For an efficient discretization of the voxels intensity values shown in the histogram plot (figure 2), we choose an appropriate quantization scheme depicted in figure 3 that can transform $X(n)$ into $Q(n)$. The main parameters that adjust the quantaization scheme are the number of levels M and the resolution r where the scheme shown in figure 3 sets these values to 8 and one standar deviation σ, respectively. In this study, the k^{th} quantization level denoted as l_k on an interval of length L is defined as follows:

$$l_k = \mu + k\,\mathbf{r}, \tag{1}$$

$$\mathbf{r} = \alpha\,\sigma \tag{2}$$

where $k = \dfrac{-M}{2} + 1, \cdots, \dfrac{M}{2}$ and α is a positive scaling factor. μ and σ are the average mean and standard deviation of the six subjects defined as:

$$\mu = \frac{1}{6}\sum_{n=1}^{6}\mu_n \quad \text{and} \quad \sigma = \frac{1}{6}\sum_{n=1}^{6}\sigma_n\ , \tag{3}$$

where μ_n and σ_n are the mean and the standard deviation of the voxels intensity for the n^{th} subject .

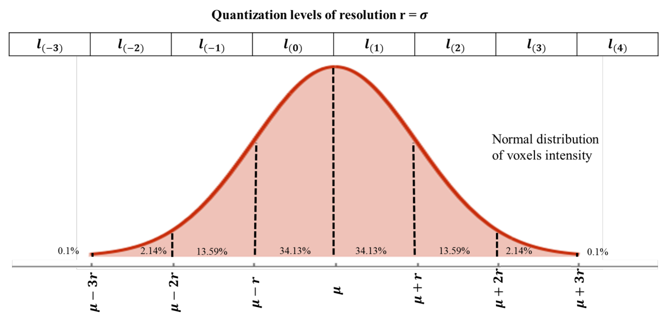

Fig. 3: The quantization of the voxels-sequence with a resolution $r = \sigma$ and $M = 8$.

The implementation strategy of the proposed quantization scheme is defined in Algorithm 1. The output of this quantization scheme is a discrete sequence that can only take finite set of integer values (i.e., $(1, \ldots, M)$. When this scheme is applied to the real-valued voxel intensity samples for both classes, two matrices of size $40 \times 16N$, namely a picture matrix P and a sentence matrix S, will be generated. Thee two integer-valued matrices P and S will be utilized to generate the final VW-based feature vector.

3.2. Voxel-Weight Based (VW-based) Features

Position Weight Matrix (PWM)-based features extraction is a popular method used for motifs representation in DNA/RNA sequences. In this features extraction paradigm, the position weight matrices (PWMs) are often extracted from a set of aligned DNA/RNA sequences that are believed to be functionally related. PWMs have a significant role in many software tools for computational motif discovery [15]. Traditionally, two PWMs are usually derived from two sets of aligned DNA sequences that are thought to be functionally related. Then, these two matrices are utilized to generate number of features that may help improve the classification performance [16]. In order to construct the PWMs, the dataset or the samples have to belong to a finite set of integers or characters. DNA sequences are naturally represented by a set of four nucleotides A, C, G and T, however; this is not the case for the fMRI

Algorithm 1: Uniform Quantizer Algorithm

Input : X: Real-valued voxel sequence
M: Number of Quantization levels
μ: Quantization centroide
$\mathbf{r}$: Quantization resolution

Output: Q: Integer valued voxel sequence

$N_X \leftarrow$ length of X

for $n \leftarrow 1$ **to** N_X **do**
  for $k \leftarrow \frac{2-M}{2}$ **to** $\frac{M}{2}$ **do**
    if $X(n) < \mu + \frac{(2-M)\,\mathbf{r}}{2}$ **then**
      $Q(n) = 1;$
    else if $\mu + (k-1)\,\mathbf{r} \leq X(n) < \mu + k\,\mathbf{r}$ **then**
      $Q(n) = k + \frac{M}{2};$
    else
      $Q(n) = M;$
  end
end

dataset (snapshots). Therefore, the real-valued voxel intensity sequence $X(n)$ needs to be quantized into a finite set of integers (i.e., levels). The quantization scheme that best fits the fMRI dataset was given in the previous subsection.

To extract the VW-based features, we extend, in this paper, the existing PWM technique and propose a methodology that can generate a reduced feature vector of two values. These two values provide a probabilistic representation of the voxel pattern in response to two distinct cognitive tasks. Having such a small feature vector (i.e., 2-D feature vector) can overcome the high dimensionality of the feature vector associated with cognitive state prediction. To generate such a feature vector, the following two steps need to be implemented:

Step 1: When Algorithm 1 is applied to all the picture and sentence voxel intensity sequences (samples), two integer-valued matrices (picture P and sentence S matrices) are constructed. After that, two weighting matrices namely the Picture Voxel Weight Matrix (PVWM) and the Sentence Voxel Weight Matrix (SVWM) can be constructed as exactly the way that the Position Weight Matrices (PWMs) are generated in [15]. A brief description of how the PWMs are generated for two classes of DNA sequences is given to help clarify the proposed concept.

Traditionally, two groups of DNA sequences that belong to two distinct classes are aligned first, then the probability of occurrence of each of the four DNA nucleotides (A,C,G and T) is calculated at each position of the DNA sequence. The probability of occurrence of each DNA nucleotide in a certain position (A,C,G or T) is equal to the number of occurrences of such nucleotide divided by the total number of sequences. For instance, if the length of each of the DNA sequence is 5, then the PWM for every class will be of size 4×5 and can take the following structure:

$$\text{PWM} = \begin{pmatrix} 0.4 & 0.2 & 0.6 & 0.25 & 0.4 \\ 0.3 & 0.2 & 0.1 & 0.25 & 0.1 \\ 0.1 & 0.2 & 0.1 & 0.25 & 0.4 \\ 0.2 & 0.4 & 0.2 & 0.25 & 0.1 \end{pmatrix}$$

In every column of the above PWM, we have four probabilities that sum up to 1 where every element in that column indicates the probability of occurrence of one of the four DNA nucleotides. Similarly, when the number of quantization levels M of Algorithm 1 is set to 6 and for any resolution $\mathbf{r}$, the size of the two Voxel Weight-based matrices PVWM and SVWM will be $6 \times 16N$. The matrix structure of the Picture and the Sentence Voxel Weight Matrices can take the following form:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \text{\# columns = 16N} \\ \begin{pmatrix} 0.1 & 0.5 & 0.4 & 0.1 & .....W(1,16N) \\ 0.2 & 0.1 & 0.2 & 0.1 & .....W(2,16N) \\ 0.1 & 0.1 & 0.1 & 0.1 & .....W(3,16N) \\ 0.1 & 0.1 & 0.1 & 0.3 & .....W(4,16N) \\ 0.3 & 0.1 & 0.1 & 0.2 & .....W(5,16N) \\ 0.2 & 0.1 & 0.1 & 0.2 & .....W(6,16N) \end{pmatrix} \end{array} \tag{4}$$

where the weights at every column of the above matrix indicate the probability that a certain voxel intensity will be quantized in one of the six quantization levels along the 40 picture trials or the 40 sentence trials.

Step 2: In step 1, the frequencies of any of the M quantized levels along the picture and sentence sequences were used to derive the Picture Voxel Weight Matrix (PVWM) and the Sentence Voxel Weight Matrix (SVWM). In this step, we utilize these two matrices to compute two scores for every integer-valued sequence $Q(n)$. These two scores indicate the likelihood of the sequence to be a picture or a sentence sequence. Following this strategy, two scores are computed for each sequence in the dataset and are calculated as follows:
Let S_j be the j^{th} row of the integer-valued sentence matrix S, and similarly P_j be the j^{th} row vector of P. From step 1, $PVWM_{i,j}$ and $SVWM_{i,j}$ represent the probability of occurrence of the integer number i where $i = 1, \ldots, M$ at a time instant j along the picture and the sentence sequences, respectively. Therefore, the two scores can be calculated for

Table 1: Comparison of the accuracy obtained for each subject and the average accuracy

Star/plus Subject	Method1 ([10])			Method2 ([13])			VW method [$M = 6$, r $= \sigma$]		
	classifier	*feature size*	*Accuracy*	*classifier*	*feature size*	*Accuracy*	*classifier*	*feature size*	*Accuracy*
4799	NB	4080	0.45	SVM	255	0.99	LR	2	1
4820	NB	6528	0.9	SVM	408	0.89	LR	2	1
4847	NB	5088	1	SVM	318	1	LR	2	1
5675	NB	6384	0.95	SVM	399	0.98	LR	2	1
5680	NB	5344	0.95	SVM	334	1	LR	2	0.987
5710	NB	3504	1	SVM	219	0.99	LR	2	1
Average	-	5155	**0.8750**	-	322	**0.9750**	-	2	**0.998**

any integer-valued sequence $Q(n)$ as follows:

$$Score(1) = \sum_{j=1}^{16N} PVWM_{Q(j),j} \tag{5}$$

$$Score(2) = \sum_{j=1}^{16N} SVWM_{Q(j),j} \tag{6}$$

Originally, every picture or sentence sample is represented by a $16N \times 1$ feature vector. However, after applying the proposed feature generation methodology, this high dimensional feature vector will be mapped into a 2-D feature vector. The size of the full feature matrix is 80×2 where half of these samples represent the picture trails and the other half represent the sentence trails. Finally, a prediction model can be derived using this reduced feature vector.

4. RESULTS

4.1. Machine Learning Model

Due to the high dimensionality of the feature vector compared with the small number of samples, the Leave-One-Out (LOO) cross-validation scheme was used to avoid an unbiased measure of test accuracy. The generated features were training using Logistic Regression (LR) classifier due to its simplicity and the satisfying results obtained with this model. Unlike most of the features used for classification, the VW-based features cannot be generated independently. This is because the VW-based features are correlated in the sense that the features cannot be extracted without the knowledge of the other training dataset (i.e., the whole training datset needs to be processed together in order to generate the PVWM and SVWM). Therefore, one should be careful when implementing such features. To do so, the PVWM and SVWM are reconstructed for every training dataset (79 samples) that correspond to every fold of the 80 leave-one-out cross validation folds. The reason of doing this is that in any classification problem the testing set should not be included in the learning stage. Hence, the PVWM and the SVWM have to be recalculated for every leave-one-out fold.

4.2. Comparison with existing methods

For every subject of the star/plus dataset, the voxel intensity sequence $X(n)$ was first quantized using 6 levels and a resolution of one standard deviation (i.e., $M = 6$ and $r = \sigma$), and then the VW-based features were generated as shown in section 3. To illustrate the effectiveness of the proposed VW-based features, we compare the performance of the LR classifier with two different classifiers ([10, 13]). For a fair comparison, the performance of the proposed method was only compared with prediction models that utilize feature vectors derived only from the ”CALC” Region of Interest. In [10, 13], the authors applied the Naive Bayesian (NB) and the Support Vector Machine (SVM) classifiers to feature vectors that were derived from 7 ROI, respectively. However, in Table 1 we report only the results obtained by using the feature vectors derived from the ”CALC” ROI.

The two metrics used to compare these classifiers are the size (i.e., dimension) of the feature vector and the prediction accuracy. To illustrate the overall performance of each method, we compute also the average feature size and the average accuracy. From Table 1, the average size of the VW-based feature vector is significantly small compared to those used for the other two methods. Using the VW-based prediction method, the average accuracy of the cognitive state prediction problem was improved by about 2.3% and 14% compared to that obtained by Method 2 and Method 1, respectively. For subject 5680, Method 2 outperformed the proposed VW-based method and Method 1.

5. DISCUSSION

Table 2 depicts the sensitivity analysis that we performed in which the average accuracy across the six subjects was computed for every scenarios. The purpose of this analysis is to investigate the impact of the relationship between the statistical measures (μ and σ) of the voxel intensity sequence $X(n)$ and the quantization parameters (M and r) on the average accuracy of the prediction model. From Table 2, one can achieve a consistent (i.e., stable) prediction model with 100% average accuracy when the number of qunatizaion level M is greater than or equal to 6 and the resolution r is between 0.8σ

and 1.2σ.

Generally, when the resolution r is smaller than or equal to 0.6σ, the average accuracy for every subject fluctuates and goes down for most of the scenarios. One possible reason for this phenomena is the overfiiting issue that may happen when the resolution gets very small and the number of levels increases. On the other hand, when the resolution r gets wider (i.e., $r >= 1.4\sigma$), the two VW-based matrices namely PVWM and SVWM will be sparse. As a result, the generated features under this choice of quantization parameters for both classes will not be different from each other. Consequently, the average accuracy of the LR classifier decreases as shown in Table 2.

Table 2: The sensitivity of the average accuracy (Acc_{avg}) w.r.t the quantization Levels M of and the resolution **r**

M \ r	**0.2σ**	**0.4σ**	**0.6σ**	**0.8σ**	σ	**1.2σ**	**1.4σ**	**1.6σ**	**1.8σ**	**2σ**
4	0.91	0.92	0.84	0.84	**1**	0.98	0.92	0.84	0.83	0.83
6	0.90	0.6625	0.67	**1**	0.99	**1**	0.92	0.84	0.83	0.83
8	0.75	0.59	0.69	0.99	**1**	**1**	0.92	0.84	0.83	0.83
10	0.78	0.42	0.70	**1**	**1**	**1**	0.92	0.84	0.83	0.83
12	0.56	0.29	0.76	**1**	**1**	**1**	0.92	0.84	0.83	0.83
14	0.66	0.35	0.75	**1**	**1**	**1**	0.92	0.84	0.83	0.83

6. CONCLUSION

In this work, we developed a novel feature generation methodology that can improve the performance of cognitive state prediction models. We also present a quantification scheme required for the implementation strategy of the proposed feature generation. The novelty of this feature generation method is its ability to reduce the high dimensionality associated with the original feature vector into a two-dimensional feature vector while attaining significant prediction accuracy. To demonstrate that, we compare the performance of the LR classifier under the VW-based feature with a NB and SVM classifiers. Finally, a sensitivity analysis of the quantization parameters was performed to help choose the optimal quantization parameters for similar classification problems.

7. AUTHOR'S CONTRIBUTIONS

FA, AC and TMLK conceived and designed the experiments. SA performed the VM-based experiments. FA performed the comparison analysis while AC performed the sensitivity analysis. FA, AC wrote the paper. TMLK and FA read and approved the final manuscript.

8. ACKNOWLEDGMENT

Research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST) Base Research Fund (BAS/1/1606-01-01).

9. REFERENCES

[1] Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. Neurophysiological investigation of the basis of the fMRI signal. Nature. 2001;412:150.

[2] Ogawa S, Lee TM, Nayak AS, Glynn P. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magnetic resonance in medicine. 1990;14:68–78.

[3] Shenton ME, Dickey CC, Frumin M, McCarley RW. A review of MRI findings in schizophrenia. Schizophrenia research. 2001;49.

[4] Rombouts SA, Barkhof F, Goekoop R, Stam CJ, Scheltens P. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Human brain mapping. 2005;26:231–239.

[5] Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, et al. Learning to decode cognitive states from brain images. Machine learning. 2004;57:145–175.

[6] Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 2009;45.

[7] Ryali S, Supekar K, Abrams DA, Menon V. Sparse logistic regression for whole-brain classification of fMRI data. NeuroImage. 2010;51:752–764.

[8] Cabral C, Silveira M, Figueiredo P. Decoding visual brain states from fMRI using an ensemble of classifiers. Pattern Recognition. 2012;45:2064–2074.

[9] Ng B, Vahdat A, Hamarneh G, Abugharbieh R. Generalized sparse classifiers for decoding cognitive states in fMRI. In: International Workshop on Machine Learning in Medical Imaging; 2010. p. 108–115.

[10] Wang X, Mitchell T. Detecting cognitive states using machine learning. 2002;.

[11] Pereira F, Gordon G. The support vector decomposition machine. In: Proceedings of the 23rd international conference on Machine learning; 2006. p. 689–696.

[12] Ramasangu H, Sinha N. Cognitive state classification using transformed fMRI data. In: Signal Processing and Communications (SPCOM); 2014. p. 1–5.

[13] Ramakrishna JS, Ramasangu H. Classification of cognitive state using statistics of split time series. In: India Conference (INDICON), 2016 IEEE Annual; 2016. p. 1–5.

[14] Gray RM, Neuhoff DL. Quantization. IEEE transactions on information theory. 1998;44(6):2325–2383.

[15] Bucher P. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. Journal of molecular biology. 1990;212:563–578.

[16] Albalawi F, Chahid A, Guo X, Albaradei S, Magana-Mora1 A, Jankovic BR, et al. Hybrid model for efficient prediction of Poly(A) signals in human genomic DNA. Methods. Submitted;.